Australian Mathematical Sciences Institute
and Statistical Society of Australia

DATA SCIENCE REVIEW

NOVEMBER 2023

AMSI
AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Statistical Society of Australia

Contents

Summary and Recommendations

With the ever-expanding collection of information for a wide range of purposes across all areas of life, the work of handling, analysing and using data has truly taken flight in the last decade. So has the professional field of data science, which encompasses the wide range of efforts to extract useful and valid knowledge from data. Data science involves general or specialised activity in fields such as data infrastructure, the collation and management of data, the development and application of models and algorithms for data analysis, and visualisation and quantitative communication. The knowledge and skill underpinning these activities are sourced from a range of core academic disciplines including mathematics, statistics, information science, computer science, artificial intelligence, as well as many other disciplines.

This review into the role of the mathematical sciences in data science within Australia was initiated by the Australian Mathematical Sciences Institute (AMSI) and the Statistical Society of Australia (SSA). The Review Panel included representatives from AMSI and SSA as well as universities across Australia. The Panel obtained information through four key activities: a scan of published government and market reports on data science in Australia and overseas, a review of data science courses in Australian universities, an online survey, and focus group meetings with academics, early career professionals, students and industry. In addition, the panel received online submissions via its review website.

This report represents a comprehensive review of data science in Australia. It encompasses university education in data science as well as industry and government requirements for data science professionals.

Main findings of the review

Employment in data science is growing quickly, and is now an officially recognised occupation in Australia. Industry respondents confirm that data science is a growth profession and will remain so for the foreseeable future. There is growing and unmet demand for well-trained data scientists in academia, research and industry.

Data science education is growing strongly in response to this demand. There is a wealth of data science undergraduate and postgraduate courses currently offered by Australian universities. However, tracking the growth and impact of these degrees is difficult, due to the absence of a separate Field of Education for data science and the scattering of degree offerings across multiple Faculties.

Most data science courses in universities are a blend of computer science and mathematical sciences (which include both mathematics and statistics). Our findings indicate that data science degrees can have different foci; some with the emphasis on IT while others have an emphasis on statistics. A strong focus of this review is on degrees with strong mathematics and statistics content, in order to promote skills capability in analysis and the discovery of meaning in data.

Despite the growth in university courses, there is still an unmet demand for training in data science, be it through traditional academic courses, micro-credentialling, in-house professional training, school education or community up-skilling. In particular, there is a strong appetite to upskill employees already in the workforce.

Industry and government feedback clearly shows a demand for data scientists with sufficient training in university-level statistics subjects. Given the importance of understanding core statistical and mathematical concepts for the data science professional, the mathematical sciences, especially statistics, need to be well represented in the teaching of data science degrees. At the same time, data science encompasses a variety of conceptual, technical and soft skills and the level of necessary statistical training might vary depending on the specialisation of the individual. In addition to technical knowledge and understanding, employers place significant value on communication skills and other “soft” skills as part of the data scientists’ toolkit.

Senior secondary school mathematics is important preparation for entry into data science courses, and increasing student engagement with intermediate and higher-level mathematics subjects (called Mathematical Methods and Specialist Mathematics in most jurisdictions) in senior secondary school is seen as vital to grow the future data science workforce.

Among academics, there is mixed satisfaction with current university data science courses and their management. There is uncertainty about whether the growth of data science enrolments has positively impacted employment of statisticians or their research priorities in universities. Given the competing industry demand and commensurate remuneration in data science jobs, it is challenging to retain students into advanced and research degrees in data science, and to fill positions for academic teaching and research staff.

There are currently no commonly agreed standards for data science, the professional accreditation of data scientists or data science education in Australia. This complicates employer recruitment of staff with the required skillset and best fit for the job, especially if the employer is unsure what data science knowledge and skills would best suit their needs. The accreditation infrastructure is split between the computer science and statistics domains; the Australian Computer Society (ACS) provides accreditation of professional data science specialisations within ICT degrees. From the mathematical sciences side, the SSA provides accreditation of (undergraduate and postgraduate) university degrees in statistics.

There is a general wish for better connections between data science, mathematics and statistics within the broader community. The review feedback confirms that professional organisations such as AMSI and SSA can play valued roles in advancing the role of mathematical sciences as a discipline at the heart of data science in Australia. Key roles include advanced research training events in data science for university students and industry professionals, promotion of internships and work-integrated learning, advocacy, career education, facilitation of a community of professionals, and establishment of a contact point between industry and students.

General recommendations

1. To satisfy the growing industry demand for data scientists in Australia, Australian universities should offer both undergraduate and postgraduate degree qualifications in data science.
2. Data science courses should have their own separate Field of Education to track offerings and student load. This would also help resolve any demarcation issues between faculties and disciplines, giving data science a distinct identity.
3. Universities offering a data science degree should include an option, such as a major, that has a minimum of 50% mathematical and statistical content to meet industry demand for high-level data science expertise.
4. Every data science degree or major should include a minimum of 25% of statistical content, taught by qualified statisticians.
5. Given the significance of communication and other “soft” skills, these should be incorporated in data science courses, including the teaching of effective data presentation and interpretation, as well as opportunities for industry experience and work-integrated learning where possible and practical.
6. Given the significant demand for data science upskilling within the current Australian workforce, university training in data science should be offered in learning modes which suit students in a variety of life circumstances and career stages. Full degree offerings should be complemented by nested qualifications such as certificates and diplomas and teaching should include hybrid and online delivery options.
7. A concerted effort should be made to increase the proportion of secondary school students undertaking intermediate mathematics subjects in Year 12 (entitled Mathematical Methods in most States, Mathematics Advanced in NSW), as these are essential knowledge for students entering undergraduate data science degrees.

8. Australia should take full advantage of the international student cohorts studying data science as well as degrees in the mathematical sciences to help meet the current demand, by ensuring graduates of these degrees qualify for the extended post-study work rights and by actively promoting the resulting opportunities to industry and to international students.
9. The learned societies should enhance the accreditation of data science degrees in a coordinated fashion, ensuring that professional accreditation awarded by different professional societies is recognised as having equal status.
10. Annual graduate schools and other forms of shared university teaching should offer advanced research training subjects in data science which are open to both current university students and industry professionals meeting prerequisite requirements.
11. Public funding should be made available to promote data science and careers to the broader public and raise its profile.

Suggested actions include:

AMSI and SSA provide support to universities that wish to upgrade their qualifications in data science with statistics-heavy majors and various other components including work-integrated learning.

AMSI and SSA lobby government agencies to review and upgrade the Australian Standard Classification of Education (ASCED) to include Data Science as a separate field.

Members of the data science discipline including AMSI and SSA continue to advocate for increased participation in intermediary mathematics in senior secondary schools. AMSI will leverage its schools program to provide careers information and lobby its partners in the data science discipline to actively cooperate to promote data science to secondary schools.

AMSI and SSA continue to lobby the government to ensure the widest possible eligibility for data science and related degrees in the mathematical sciences, for extended post-study work rights for international students.

SSA approach ACS for discussions to coordinate accreditation of data science degrees, to ensure the widest possible accreditation options for data science graduates. The SSA should enhance its accreditation of data science degrees that meet its requirements for mathematical and statistical content. Actions for consideration would include investigating separate accreditation of data science degrees with statistical content alongside the current accreditation of statistics degrees and increasing the status and profile of SSA accreditation to ensure equal footing with accreditation by other professional societies.

AMSI continues to routinely offer subjects at its annual graduate schools and online teaching suitable for data science students and professionals and promote the offerings to those groups.

1. Background and introduction

The Review on "The Role of Mathematical Sciences in Data Science in Australia" was initiated in 2022 by the Australian Mathematical Sciences Institute (AMSI) and the Statistical Society of Australia (SSA) as national representatives of the mathematical sciences (mathematics and statistics) communities in Australia.

A targeted review of the role of mathematical sciences in data science in Australia had not been conducted previously. The Review was commissioned in response to discussions among AMSI and SSA members about the growth of data science in education and the workplace, and the subsequent impact on the mathematical sciences profession. As members and representatives of the mathematical sciences discipline, the impact on university departments of mathematical sciences, in terms of degrees offered and research undertaken, was of particular interest to the review panel.

1.1 Overview of AMSI and SSA

AMSI is Australia's national voice and champion for mathematics and statistics. A not-for-profit, the Institute works with schools, universities, industry, government and the community to help shape policy and skill Australia for the future. Building engagement and capability, AMSI drives programs to strengthen the mathematical sciences and enhance the discipline's impact and role in Australian education, research, innovation and industry. AMSI's membership network includes over 40 universities, professional societies, government agencies and industry members.

Founded in 1962, SSA represents Australian and overseas statisticians and is the home for professionals working in statistics. The overall objective of the Society is to further the study, application and good practice of statistical theory and methods in all branches of learning and enterprise. SSA is closely linked to the national and international statistics community. Its affiliated organisations include the International Statistical Institute, the Australian Statistical Advisory Council, the Committee QR/4-Statistical Quality Procedures of the Standards of Australia and the Australian Data Science Network (ADSN).

Both AMSI and SSA are members of Science & Technology Australia (STA) and are represented in the National Committee for Mathematics of the Australian Academy of Science (AAS).

1.2 Review Panel and Reference Group

With Distinguished Professor Kerrie Mengersen as Chair, a review panel consisting of 12 members from AMSI, SSA, and academia from nine universities located in five Australian states and one Territory conducted the review.

A Review Reference Group provided advice to the Panel. The Group consisted of volunteers from key stakeholder groups, including business and industry, government and academia. A list of the Members of the Review Panel and Reference Group, and their affiliations can be found in Appendix 8.

1.3 Terms of Reference

For the purposes of this review, the Panel adopted a broad definition of data science:

"Data science seeks to extract knowledge from data. It is interdisciplinary in nature and encompasses a range of core disciplines including mathematics, statistics, information science, computer science and AI, as well as many applied disciplines. Data Scientists can have general or specialised expertise in fields such as data infrastructure, the collation and management of data, the development and application of models and algorithms for data analysis, and visualisation and quantitative communication." (AMSI/SSA Data Science Review & Survey - AMSI)

The Review Panel agreed to the following terms of reference:

- Report on the current statistics and data science degrees that AMSI member universities offer, both at the undergraduate and postgraduate level, together with the level of engagement in data science programs by member departments in the mathematical sciences at universities as well as research institutes with a data science focus.
- Report on any recent changes to member department profiles, degree or subject offerings, which have occurred as a result of data science initiatives.
- Develop recommendations for practical ways in which service courses can be delivered by professional statisticians. In particular, evaluate the feasibility of a joint delivery model (course delivered jointly by a professional statistician and a discipline expert).
- Investigate options for accreditation of data science professionals by SSA.
- Consider the role that AMSI higher education programs can play in developing the data science skills of advanced undergraduate students and how to engage data science academics in those programs.
- Consider how the mathematical sciences discipline can effectively benefit from the opportunities offered by the growth of data science in Australia.

2. Conduct and findings of the Review

The Review panel employed the following activities:

- The Review Panel held frequent collaborative Zoom meetings, with all members contributing to the key decisions. These meetings were chaired by Professor Kerrie Mengersen or Professor Tim Marchant.
- A scan of relevant government, professional and commercial reports on data science in Australia and internationally was undertaken.
- A review of Australian university courses related to data science and mathematical sciences was undertaken. The review was conducted by AMSI and Queensland University of Technology.
- An open, online survey was created and administered by the Review Panel. The survey was promoted to AMSI and SSA members, the panel members' networks, and social media channels.
- Focus groups were held with key stakeholder groups: students, industry and government representatives and academics. Each focus group was chaired by a member of the Review Panel.
- An online form was published on the Data Science Review website to allow for feedback and comments.

Further details of the context scan, survey and focus groups are provided in the following sections and in the Appendices.

AMSI staff provided substantive background work, including collection of relevant reports, scan of university courses, organisation of the focus groups, coding of the group discussions, and implementation of the survey and initial analysis of the survey results.

2.1 Context Scan

Relevant national and international reports that were included in the context scan are listed in Appendix 1.

2.1.1 Data science is now officially recognised as an occupation and scientific discipline in Australia

Data science is now officially recognised as a job classification in Australia. In its 2022 review of emerging occupations and data science, the Australian Bureau of Statistics (ABS) created new classifications for a 'data scientist' within the mathematical science occupations (group 2241), which also includes mathematicians, statisticians and actuaries. The occupation of a 'data analyst' was added as a specialisation of the data scientist occupation. Two other occupations, 'data engineer' and 'data architect', were included in the information and communication technology (ICT) unit group.

In 2020, data science was included as a Field of Research at the 4-digit level within Computing and Information Sciences, as well as "statistical data science" at the 6-digit level within Mathematical Sciences. The 2021 Report by the Australian Academy of Science (AAS) on "Advancing Data Intensive Research in Australia" also recognises data science as a distinct scientific discipline.

2.1.2 Data science is one of the fastest growing occupations in Australia

Despite data science only recently being awarded official status in Australia, the demand for data science and related skills has been growing apace in the last decade. The context scan provided the following profile of data science in Australia in 2022. (See Appendix 1 for full references.)

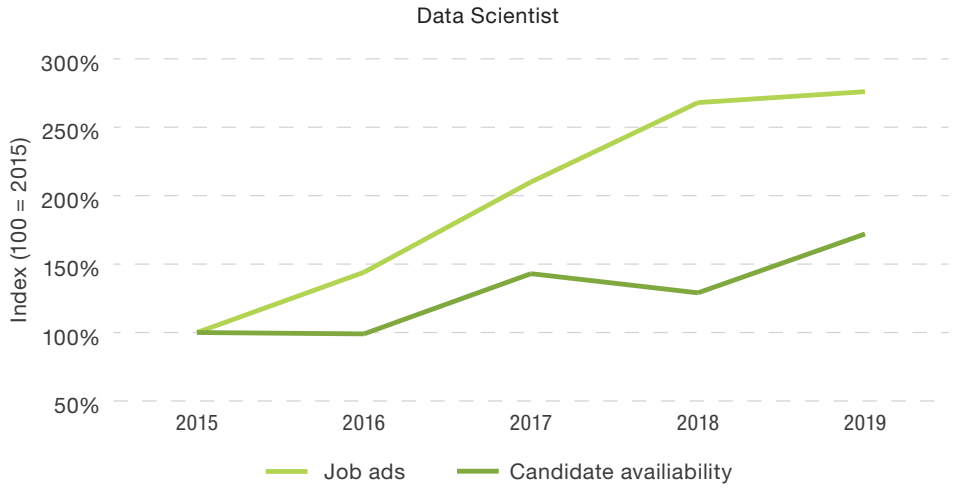
Figure 1. Summary of employment outlook for data scientists

Data Science is in demand	Data Science pays well	Data science will continue to grow
• Data science is ranked in the top 5 most in-demand and highest paying jobs in Australia. • Over 5000 data-science related roles were on offer in a given month.	• The average base salary for data science professionals is generous, around $116,000 in September 2023.	• Predicted annual average growth of the data science workforce is 2.4%, 1% higher than average. • The global market will grow significantly, increasing almost 10-fold in the next 5 years.

Sources: SimpliLearn 2022; Seek 2023; Indeed 2023; Deloitte 2018

The following figure, extracted from the CEDA 2021 Report on "A good match: Optimising Australia's permanent skilled migration" (Figure 2), highlights the increased demand for data scientists and the increasing shortage of available candidates to fill these jobs in the five years from 2015 to 2020.

Figure 2. Data scientist job ads versus candidate availability 2015-2020



Source: CEDA 2021

2.1.3 Data science education is growing strongly, but needs to grow faster

The CSIRO Report on 'Artificial Intelligence for Science' (November 2022) identifies 'Education, training and capability uplift' as one of the six key future development pathways in AI (p.ii), and acknowledges and quantifies the explosion of degrees and courses in AI (pp. 39-41, details in Appendix 2). The report points to a number of developments in AI and the implications for science and research organisations, which, as the findings of this Review show, are in alignment with the broader field which encompasses the closely related areas of data science and AI. Some of the findings include:

- The expansion of education and training courses in AI. These courses differ in content and depth, including professional flash courses, micro-credentialled courses, and longer university undergraduate and postgraduate courses (p. 39).
- The opportunity to encourage professional staff in organisations to consider a career transition into AI focused roles (p. 41).
- The direct competition between science and research sectors and industry for AI skills (p. 40).
- The need for a longer-term view of the AI pipeline, investing in foundational skills – including mathematics – in early childhood learning, primary school and high school (p. 40).

The CSIRO findings are supported by the McKinsey Global Survey (December 2022) which lists sources of AI talent from universities, internal reskilling of existing employees, industry organisations and training academies (e.g. boot camps).

There is also an urgent need to build data skills in Australia's research sector (Australian Academy of Science, 2022). "Data skills are needed through a specialised data science workforce available for research, and also more generally across all researchers. Appropriate funding, attractive employment conditions and recognition of data science expertise are all critical for attracting and retaining people with data skills in the research sector" (p. 18). This is supported by the National Research Infrastructure (NRI) Workforce Strategy and National Digital Research Infrastructure Strategy, which advocates for consideration of "investment and a long-term plan to develop the specialised workforce and support researchers to use, reuse and manage data, develop software, and utilise digital research infrastructure." (p. 18).

2.1.4 There are no commonly agreed standards for data scientists or data science qualifications in Australia

The term 'data science' is acknowledged to encompass a range of professions and occupations and most attempts at definition or description reflect this.

The Australian Computer Society uses a broad definition: "a cluster of academic and occupational streams including Data Science, Data Analytics, Data Architecture, Data Engineering and Data Wrangling often in tandem with Business Analysis, Database Management, Business Intelligence, Visualisation, Statistics, and other related areas." (ACS, 2022, p. 20).

For the purposes of this Review, the Panel adopted an equally broad working definition: "Data science seeks to extract knowledge from data. It is interdisciplinary in nature and encompasses a range of core disciplines including mathematics, statistics, information science, computer science and AI, as well as many applied disciplines. Data Scientists can have general or specialised expertise in fields such as data infrastructure, the collation and management of data, the development and application of models and algorithms for data analysis, and visualisation and quantitative communication." (AMSI/SSA Data Science Review & Survey - AMSI)

While the broad definition of data science is enabling, it also leads to a lack of clarity about the background skills required of data scientists and professional roles of data scientists. There is currently no commonly agreed set of standards for the data science profession in Australia, with the accreditation infrastructure split between the computer science and statistics domains.

The Australian Computer Society (ACS) provides accreditation of professional data science specialisations within ICT degrees. ACS accreditation of professional and advanced professional data science specialisations relies on internationally defined criteria for ICT education and the Skills Framework for the Information Age (SFIA).

From the mathematical sciences side, the SSA provides accreditation of university degrees in statistics. Although there is no explicit accreditation of data science degrees or data science professionals by the SSA, it has accredited data science degrees according to its usual criteria. More details of ACS and SSA accreditation can be found in Appendix 2.

By contrast, the approach that has been taken in the United Kingdom is a collaborative one. The Alliance for Data Science Professionals encompasses leading organisations in mathematics, statistics and IT to create a common set of standards for data scientists, data engineers, data analysts and data stewards. These standards have been used as criteria to accredit data science degrees, data science professionals and data science modules of associated degrees. More details of this approach are also provided in Appendix 2.

2.2 Review of university data science courses

Australian university websites were searched for bachelors, graduate certificates, diplomas, masters by coursework and research masters degree courses that:

a. were aligned to mathematical sciences and cognate disciplines, including mathematics, statistics, actuarial science, commerce degrees in business analytics, science, computing, artificial intelligence, cybersecurity, information technology, and

b. among those, the courses that specifically included data science, analytics or artificial intelligence in their titles (full title, major or specialisation).

c. with “course” defined as consisting of multiple subjects culminating in a degree, major or defined specialisation.

The courses identified under (b) are detailed in Appendix 3. Note that the identification of data science degrees was complicated by the fact that they are not listed under a separate Field of Education. Instead, data science is offered most often, but not exclusively, as a science, mathematical sciences, computer science or business degree. While for the purposes of this Review the search was mainly focused on degrees offered in mathematical sciences, efforts were made to include as many data science degrees located in other faculties as well. However, some data science degrees may have been missed and the below is intended as an indication only of the types and breadth of degree offerings available. The search results were up to date as of the end of 2022 – in this fast-evolving field it is likely that the degree offerings have increased further at the time of publication of this report.

2.2.1. There is a wealth of data science courses currently offered by Australian universities.

- Among 217 identified undergraduate degree courses in the identified disciplines, 57 (26%) were courses in data science, analytics or artificial intelligence, offered by 33 universities. Around half of these are offered by way of a major in a Bachelor of Science or other degree.
- Among 225 identified coursework masters degree courses in the identified disciplines, 94 (42%) were courses in data science, analytics or artificial intelligence, offered by 35 universities.
- Another 44 graduate certificate and diploma courses were identified in data science, analytics and related areas, offered by 22 universities.

2.2.2 The role of mathematical sciences departments in teaching data science is often unclear

- In 16 of the identified undergraduate data science courses, the course is taught or co-taught by departments of mathematical sciences or similar. For the remaining courses the role of mathematical sciences departments could not be identified from the available online information.
- Almost 30% of coursework masters courses are taught or co-taught by science faculties or similar that are presumed to include mathematical sciences. However, only 10 courses are listed as offered by schools or departments that explicitly include mathematics.
- Note that coursework masters degrees can encompass degrees based on an undergraduate degree within the same discipline as well as degrees focused on professional upskilling based on certificates and graduate diplomas.
- About 60% of graduate certificate courses are taught or co-taught by faculties which were presumed to include mathematical sciences (e.g. faculties of science). However, only two universities specifically identify mathematical science departments or schools as involved in the course delivery. The actual involvement of mathematical sciences in the remaining courses was not clear.

While the review successfully identified many data science courses and programs, the investigation also revealed that the involvement of mathematical sciences in the offering of data science courses is not always clear from the information provided on the university websites. Indeed, many of the websites were difficult to navigate and even locate the courses and associated details. Mathematical sciences input is likely to be underrepresented in these listings, as the custody of data science degrees often resides with more than one academic discipline; degree delivery might involve teaching by staff from mathematical sciences departments even if another faculty or department oversees the data science degree.

2.3 Survey Results

A total of 311 survey responses were submitted from September to October 2022. Respondents comprised academics in universities (n=118, 38%), data scientists working in industry or government (n=99, 32%), early career professionals (n=51, 16%), postgraduate students (n=47, 15%), managers of data scientists (n=47, 15%) and others (n=20, 6%), with multiple responses allowed.

The surveys conducted for this Review were not probability samples from defined sample frames; it was not possible with the resources and time available to conduct surveys of this nature. In several cases, the target populations are relatively small, and an attempt was made to reach all members of these populations. To this end, the survey was sent around through AMSI and SSA newsletters and social media channels, and recipients were encouraged to send on the survey to others. However, without sample frames response rates cannot be determined, which should be considered when the quantitative results are examined.

The survey results show that people who are more strongly engaged with the subject matter of the review were more likely to respond. The respondent cohort is predominantly made up of mathematical scientists and cognate disciplines, with 138 citing statistics as their main domain of expertise, and 86 mathematics. The self-reported level of data science knowledge among the respondents is predominantly high (38%) or moderate (37%). Out of the 311 respondents 151 (49%) indicated they have a PhD, 47 (15%) a masters degree, and 45 (14%) a bachelors degree.

Table 1. Domain expertise as self-reported by all survey respondents (n=311)

Domain expertise	#	%
Statistics	138	44.3
Mathematics	86	27.7
Machine Learning	61	19.6
Computer Science	44	14.1
Health Science	14	4.5
Applied Mathematics	13	4.2
Business	13	4.2
Applied Statistics	10	3.2
Biostatistics	10	3.2
Physics	10	3.2

The survey questions developed by the Review Panel are listed in Appendix 4. They covered topics such as the relative content of mathematics and statistics in data science degrees and their management; the implications of the growth of data science for academic professionals; the demand for data science professionals in industry and government, and the skills and knowledge required of data scientists.

In addition to general questions, the survey included specific sets of questions for university staff (n=121), students (n=44), early career professionals (n=63) and industry/government employees (n=147). Respondents could choose to answer more than one specific set of questions—an opportunity taken by 60 survey respondents. The results of the survey are detailed in Appendix 5.

The main outcomes of the survey are as follows:

2.3.1 Most data science courses are a blend of computer science and mathematical sciences.

When asked about the balance between mathematical, statistical and computer science content in their university's data science program, 72 out of 121 university respondents (60%) stated that it was 50% computer science and 50% mathematical science, a balance that no university respondents were dissatisfied with. The next largest cohort (23%) said it was 75% computer science and 25% mathematical science (a combination that dissatisfied 11 out of the 28 university staff involved).

Students and early career professionals (n=107 combined) estimated that their (ongoing or recently completed) degree contained on average around 50% mathematics and statistics content (an average of around 30% statistics and 20% mathematics).

2.3.2 There is mixed satisfaction with current university courses and their management.

There was mixed satisfaction with current degrees and their management. While only 16% of university respondents reported to be unsatisfied, less than half of university respondents (46%) reported to be satisfied, while 38% remained neutral.

On the positive side, 66% of university staff report that a high to very high proportion of statistics content in their data science degrees is taught by academics with postgraduate training in statistics.

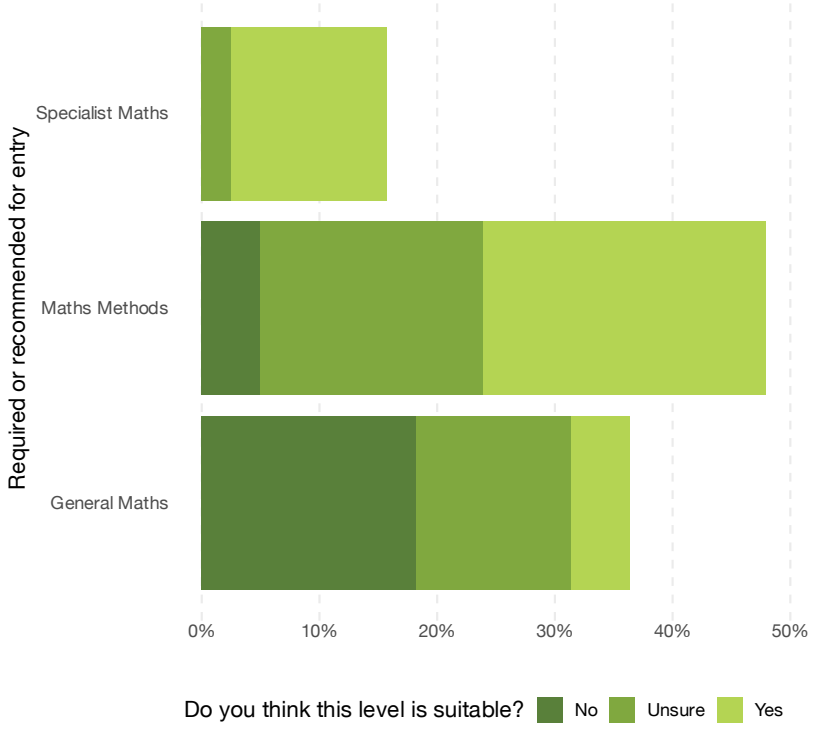
2.3.3 Mathematics is important for entry into data science courses.

Of the 121 university staff, 48% indicated that Mathematics Methods[1] was the recommended knowledge level of high school mathematics to start a data science degree at their university, with 36% stating General Mathematics[2] and 16% that Specialist Mathematics[3] was recommended for entry.

Where Specialist Mathematics (Extension Mathematics in NSW) is recommended prior knowledge, more than 80% agreed, as is to be expected, that this provides enough preparation for a data science degree. Where Maths Methods or equivalent is recommended, the perceived level of suitability is less clear-cut, with only half of the university respondents agreeing that this level is suitable, and around a third unsure.

Most respondents believed that General Mathematics does not provide sufficient background for entry to a data science undergraduate degree, although a relatively high number are unsure as well.

Figure 3. Recommended high school mathematics level for entry into data science degrees; and perceived suitability of the recommended high school mathematics level (n=121)



2.3.4 Training in mathematical sciences is important for data scientists

All survey respondents were asked to list up to five skills they see as essential to data scientists. The question was open-ended, and respondents also added knowledge areas and personal attributes they thought were important. The resulting list therefore includes skills, knowledge and attributes (the top 15 are listed below) that form part of the data science professional's "package". Mathematical sciences rank highly, with statistics topping the list and mathematics ranking fifth. In addition to the technical skills linked to the data science profession – programming, data wrangling, computing, machine learning and AI - the list also emphasises the importance of "soft" communication skills. As some focus group members emphasised, critical thinking and problem-solving skills, while not strictly mathematical by themselves, are crucial parts of mathematical sciences training.

1 Advanced Mathematics in New South Wales.
2 Further Mathematics (Victoria and Australian Capital Territory), Standard Mathematics (New South Wales), Mathematics Applications (Western Australia).
3 Extension Mathematics in New South Wales.

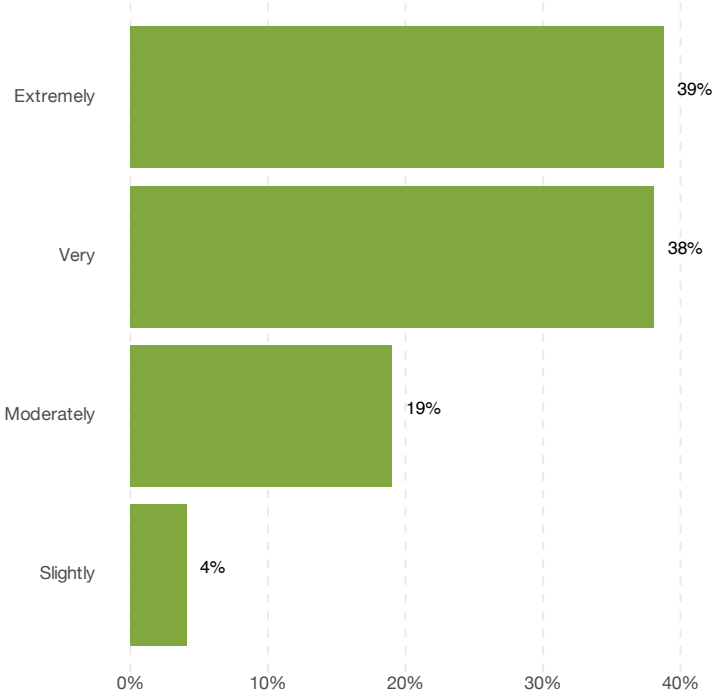
Table 2. Top 15 essential skills, knowledge and attributes for data scientists, identified by all survey respondents (n=311)

Skill Description	#
Statistics	218
Communication	170
Programming	115
Data Wrangling	110
Mathematics	74
Computing	73
Machine Learning & AI	58
Critical Thinking & Problem-Solving	52
Data Visualisation	51
Data Analysis & Data Literacy	49
Psychological Traits (Open-Mindedness, Curiosity, etc)	43
Modelling	34
Database	30
Domain Knowledge	22

The emphasis on the importance of mathematical and statistical training as part of the overall skillset of a data scientist is echoed in other survey responses. Most of the student respondents stated that the amount of mathematics and statistics content in a data science degree should be moderate (45%) or (very) high (52%).

Industry and government professionals place an even higher emphasis on the importance of mathematics and statistics, with 77% indicating they are either “very” or “extremely” important for data science graduates. When asked why (in an open question), industry respondents most often replied that mathematics and statistics provide theoretical understanding in general, as the foundation of data science. In their replies, industry and government professionals indicate that it provides the necessary understanding of the models and tools being used, and supports the appropriate analysis and interpretation of data so that results are valid.

Figure 4. Importance of mathematical sciences training for data scientists according to industry and government respondents (n=147)



2.3.5 The impact of the growth of data science on employment of statisticians or research priorities in universities is unclear

Responses were mixed regarding this issue. While 31% agreed that the growth of data science had positively influenced the employment of statisticians at their university, 26% stated that there had been no such effect and 42% were unsure. The same mix of responses was found for whether the growth of data science had influenced research priorities within their academic unit.

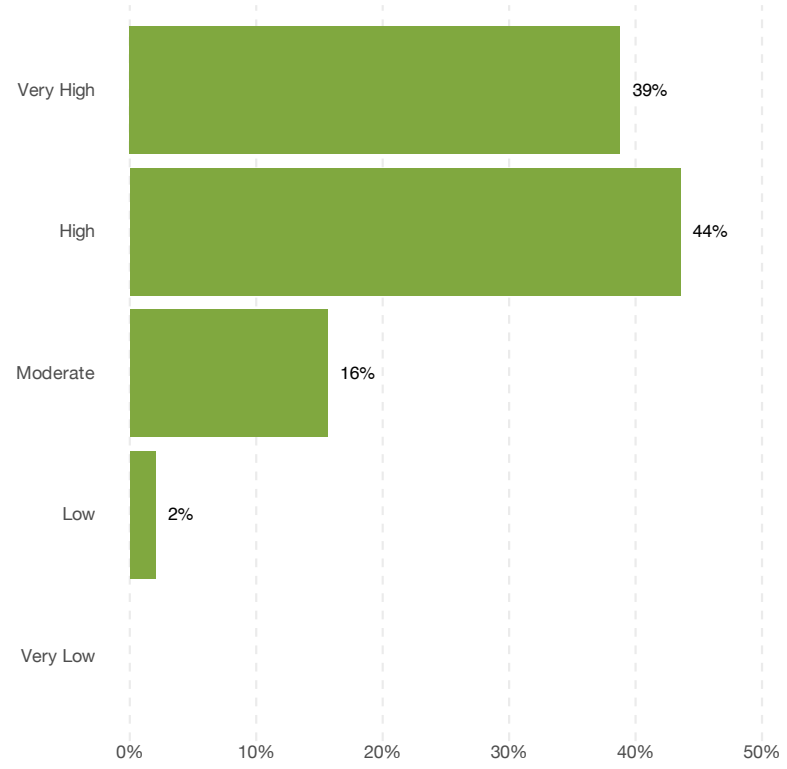
Feedback from the panel members suggested that this could be related to the impact of COVID-19 on academic employment, and also noted that in some areas, favour is shown for employing computer scientists to teach data science over statisticians.

2.3.6 Industry and government respondents confirm that data science is a growth profession.

Over 60% of industry and government respondents (n=147) confirmed the number of data scientists had grown in their organisations in the past five years. Moreover, 52% of respondents expect that their organisation will employ more data scientists in the next five years. Only 5% (n=8) do not expect additional data science employment – the remaining 42% were unsure.

Figure 5 below shows the high level of demand in industry and government for data scientists with the requisite skill sets. When asked what role data science plays in their organisation's operations and strategies, most industry and government respondents replied that they see it as a core or key competency in their organisation, without necessarily specifying its role. Some of the specific ways in which respondents did report the use of data science in their organisation were decision making; product/process design and improvement; policy development, analysis and evaluation; reporting and communicating; consulting; marketing and customer relations. This confirms not only the level of demand but also the wide variety of organisational activities that use data science as a key strategic competency.

Figure 5. The level of demand for data scientists with the requisite skills according to industry and government respondents (n=147)



2.3.7 More training is required.

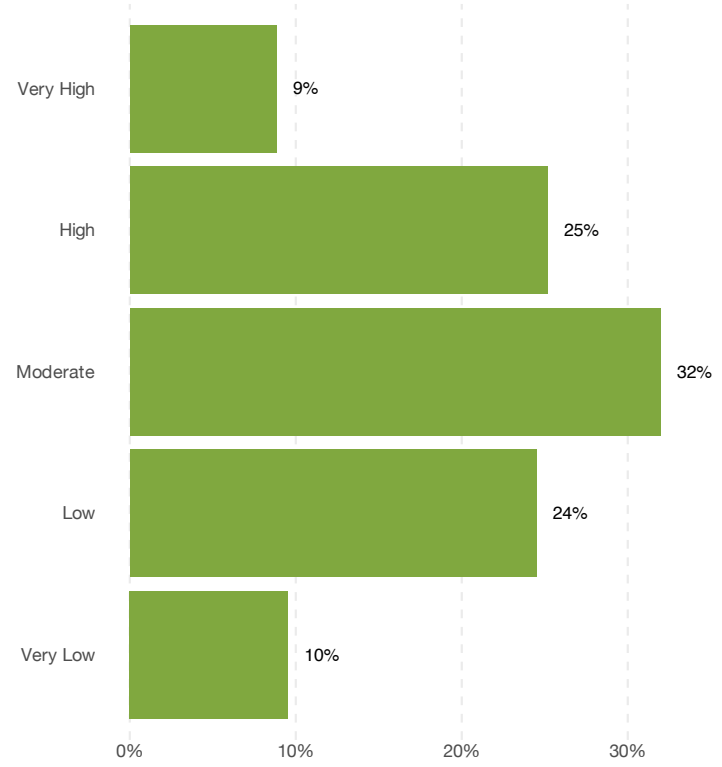
Students and early career professionals in data science and related fields are reasonably satisfied with their preparation for a career as data scientist. About 61% of students (n=44) felt that their degree is teaching them the necessary skills to a high or very high extent. More than 80% of early career professionals (n=63) indicated that the mathematical and statistical training they received was adequate for their current role.

At the same time, over 70% of early career professionals indicated that they have already undertaken or intend to undertake post-degree professional training by way of short courses, a postgraduate degree or self-directed online learning. Some areas for further study, as reported by the early career professionals, included statistics, specific software (such as Python, R, PowerBI, SAS), machine learning and programming.

Almost 40% of industry and government respondents stated that data science graduates do not have the requisite skill sets for their workplace. When asked what skills they were lacking, some of the more frequent responses include lack of statistical skills; lack of experience with "real world" multidimensional data; communication skills; coding/programming; and domain specific knowledge.

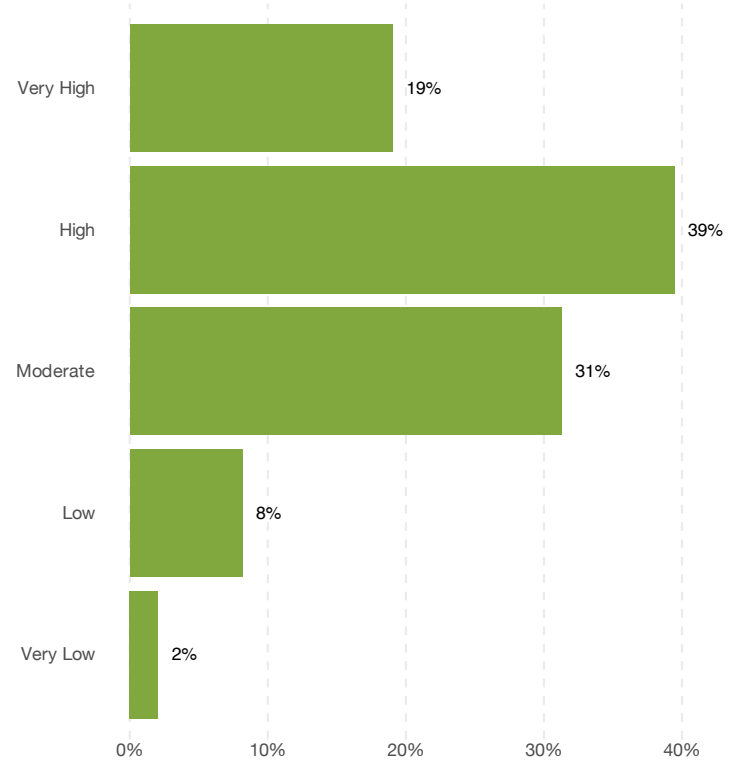
Moreover, only one third of the industry respondents reported a high or very high level of data science knowledge among their current employees. A third reported low or very low levels of knowledge, and the remaining third reported only moderate knowledge levels.

Figure 6. The perceived level of knowledge of data science among staff, according to industry and government respondents (n=147)



Almost 60% of industry and government respondents report that the appetite among their workforce to upskill is high to very high.

Figure 7. Appetite among workforce for upskilling in data science, according to industry and government respondents (n=147)



2.3.8 AMSI and SSA can support the growth of data science

The survey included some open questions on the possible role of professional societies — in particular, SSA and AMSI – in supporting data scientists and the profession in general. The provision of training, advocacy and accreditation were among the most frequent responses. mentioned in the responses. A summary of responses to both questions is in the tables below.

Table 3. Roles professional societies such as SSA can play in supporting data science training, data scientist researchers, professionals and students, according to all survey respondents (n=311)

Suggested activities for professional societies such as SSA	#
Upskilling and training	139
Networking	80
Accreditation	73
Conferences/workshops	72
Advocacy	56
Raising awareness of DS and its definition	12
Careers, recruitment and job seeker support	11
Mentoring	5
Other/don't know	99

Table 4. How AMSI can support advanced training of data science students, according to university staff (n=121)

AMSI support for data science training	#
Summer/Winter Schools	33
ACE Network offerings	19
Don't know	18
Bridging/gap/short courses	9
Separate DS event/conference	8
Advocacy Federal Government	7
Data science standards/accreditation/quality control	7
Australian Postgraduate Research Internships/other internships	5
Industry involvement	5
Career awareness	4
Other	25

2.4 Focus Groups

Four focus groups were held with identified stakeholder groups:

- industry and government;
- academics;
- students, recent graduates and early career professionals;
- members of the Reference Panel of this Review, comprising selected eminent data science professionals from industry, academia and government.

Participants for the first three groups were selected via a combination of purposive and convenience sampling. The Review Panel collated lists of possible participants with an active interest and involvement in data science across a range of industry, government and universities who were invited to participate in a forum group discussion. Industry and government professionals included representatives from large corporate organisations in retail and hospitality, banking and finance, digital infrastructure, online services, and data analytics consultancy, as well as prominent government agencies such as ABS, BoM, and CSIRO - all organisations with significant knowledge and interest in growing data science capability in Australia.

The focus group discussions were facilitated by members of the Review Panel with the assistance of AMSI staff. The facilitators used a list of suggested questions as a guideline for the discussions, encouraging other themes and topics to be discussed as they arose. The recorded discussions were transcribed verbatim, then analysed and coded by AMSI staff with reference to the Terms of Reference of the Review.

The focus group questions are listed in Appendix 6. Details of the focus group responses are presented in Appendix 7. Common themes emerging from the discussions are summarised below.

2.4.1 The lack of a coherent definition of data science is a mixed blessing

As a new field of economic activity, research and education, data science is multi-disciplinary in nature. Focus group participants emphasised the multidisciplinary and multi-faceted nature of data science – as well as the potential strengths that the different perspectives can offer. The diversity in background and disciplines poses both challenges and opportunities.

A common theme among all focus group discussion was the lack of a coherent working definition of what data science actually is, and the ramifications of that lack of coherence for the way data science is perceived at school, taught at universities and practiced in the workplace. Some focus group participants pointed out that the lack of a coherent definition of data science requires the employer to be very specific about the actual skills required when recruiting data science professionals.

2.4.2 There is no one-size-fits-all set of skills required of a data scientist

Many participants emphasised that not everyone involved in data science-related work will need the same skills or be able to operate at the same level.

In terms of the skills people need working in data science related jobs, the focus group participants' views broadly aligned with the survey results. Participants underscored that it is necessary for data science professionals to have skills such as coding and programming, data management and data wrangling.

Given the terms of reference, the discussions included a particular focus on the relative importance of mathematical and statistical knowledge as part of the data scientists' skillset. Most focus group participants agreed that sufficient grounding in the mathematical sciences is essential to aid data scientists in their ability to problem-solve, to think deeply and reason well.

Some other areas of skill and ability were also mentioned – one such area is "soft" skills and attributes, particularly around communication, teamwork, adaptability and self-directedness. There was broad agreement that communication is a vital part of being a data scientist, both to work with others as well as to present results to stakeholders in clear and effective ways.

Another area of discussion was domain-specific knowledge and contextual understanding. Data analysis does not occur in a vacuum but is deeply entrenched with many "real-life" fields and domains which require the data science professional to have a working understanding of. The real-life context of the data, and the domain specific knowledge required, elicited some comments within the focus groups that indicated that this is also of central importance to data science.

Other desirable attributes and skills mentioned in the focus group discussions related to the rapidly evolving tools and technology, which requires data science professionals to be flexible, self-directed and able to learn and adapt to new systems quickly.

2.4.3 Mathematics and statistics skills are essential for data scientists

Coming up consistently among all focus group discussions was that data scientists need to be able to understand what is going on "under the hood" rather than applying the technological tools available without critical reflection. Among other things, this requires a deep understanding of the tools that are used, how they are used correctly and why – which is all based on a solid statistical and mathematical foundation. As one focus group member explained:

"I would just say that mathematics is essential to data science to impart rigour and to understand the foundations of the methods being used in order to obtain correct analyses and predictions." (Reference Group)

2.4.4 Mathematics and statistics are key components of data science degrees

Both in the Industry and Government group and in the Reference Group, current data science degrees received criticism for delivering graduates without enough mathematical and statistical background.

Participants in the Industry and Government focus group expressed some concerns about whether (undergraduate) general data science degrees were in fact at all useful, given that these degrees need to cover a lot of ground from diverse academic disciplines. Some suggested that it might be possible to add data science to existing degrees in science and engineering as an alternative.

In terms of what a general data science degree should contain, there was broad agreement that acquiring computer science, statistics, and mathematics knowledge at an academic level should form most of the degree, with platform-specific technical certifications useful "in an ideal world".

The debate on the relative weighting of mathematics, statistics and computer science related subject matter was nuanced, but with very broad agreement that the mathematical sciences (particularly statistics) contribution to the degree should be roughly the same as the computer science component.

University academics involved in offering data science courses indicated that data science degrees offer flexibility to meet the varying demands, with the content of the actual courses building on the strengths of the expertise that is in-house. It was mentioned that part of the design of data science degrees should involve connecting with the industry sector and their skills needs.

2.4.5 Internships and work-integrated learning are valuable

Students and early career professionals agreed on the usefulness of internships and work integrated learning, particualrly to learn some of the technical skills that are not always on offer within university courses and to gain experience within a “real-world” context.

2.4.6 There is broad concern about the delivery of data science degrees

Despite the emergence of data science degrees at many universities, the degree delivery is often housed in different departments, schools and faculties. The university academic focus group discussed how best to bring together the different areas of expertise and if convergence towards one new discipline of data science is a realistic or desirable prospect.

The discussion among the academics included the topic of discipline expertise, the importance of subjects being delivered by discipline experts (this includes delivery of mathematics and statistics subjects by mathematical scientists) as well as the challenges of finding people with the right kind of expertise.

The challenge for universities is to find the synergy between the interdisciplinary nature of data science while honouring the core disciplines contributing to data science degrees.

2.4.7 Data Science graduates and professionals are in strong demand

The focus group participants cited anecdotal evidence that there is a great demand for data science professionals.

In terms of the level of study, Industry and Government representatives participating in the focus group preferred to employ masters or PhD level graduates.

There is broad agreement about the importance of making data science more visible to potential students, and motivating high school students and undergraduate university students to consider data science as a career through coordinated outreach programs.

Some offered migration and focusing on mature age students as alternative ways to boost supply.

2.4.8 There is support for data science degree accreditation

Several participants pointed to the merits of accreditation and advocated for it for various reasons. For example:

a. accreditation may raise the profile of data science as a sound career prospect;

b. accreditation can provide professional credibility to data scientists;

c. accreditation can provide clarity and certainty around the content of qualifications and the required knowledge and skills of data scientists.

2.4.9 AMSI and SSA can play key roles in advancing Data Science in Australia

Participants highlighted the following roles for AMSI and SSA:

- Advocacy;
- Career information to students, parents and teachers;
- Building a community of professionals;
- A contact between industry and students.

2.5 Online Feedback

The Data Science Review invited feedback from the academic community (academics and students) through an online form on the Review website (https://amsi.org.au/amsi-ssa-data-science-review/).

The online consultation received eight submissions, six from university academics, and two from (former) students.

The comments in the online form provide insights and opinions on the teaching of data science degrees at some of Australia's universities and give further context to the discussion points raised in the survey and the focus groups. One area of discussion was the shared delivery and governance of data science degrees.

2.5.1. Examples of successful collaborative models for data science degree delivery and shared governance

Three submissions contained particulars of the setup of data science and related degrees – all the described degrees include multiple faculties and schools. As one submission noted, shared teaching between faculties and disciplines avoids having to "re-invent the wheel" and builds on existing discipline expertise.

One submission described how the multi-faculty collaboration to facilitate data science degrees was set up and is managed at their university. The Data Science major at this university is managed through "*a shared governance model in the form of a Steering Committee, with two people from each faculty offering core units (the Associate Dean Education and a faculty expert), the major coordinator, one person from faculties offering electives and a rep from the Office of the DVC Education.*"

The online feedback confirmed that the increase in data science degrees has implications for staffing, with additional teaching resources needed but not always easy to find.

2.5.2. Working with industry has benefits for students and universities

The online feedback also confirmed the value of working with industry, or at least for working with "authentic applications" and case studies, for example, as part of the Data Science degree content. The submissions identified a number of advantages of close collaboration with industry:

- Industry connections are useful for new graduates when job seeking;
- Insight in "industry best practice";
- It can create new conversations and partnerships between academia and industry and open up new avenues for research and teaching;
- It provides additional teaching resources to universities;
- It motivates students to work with "real data".

References

Alliance for Data Science Professionals (AfDSP), UK, 2022. https://alliancefordatascienceprofessionals.com/

Anaconda, 2023. *2022 State of Data Science*. https://www.anaconda.com/state-of-data-science-report-2022

Artificial Intelligence Index Report 2023, 2023. Stanford University. https://aiindex.stanford.edu/report/

Australian Academy of Science, 2021. *Advancing Data Intensive Research in Australia*. https://www.science.org.au/files/userfiles/support/documents/advancing-data-intensive-research-in-australia-11-11-21.pdf

Australian Academy of Science, 2022. *Australia's Data-Enabled Research Future: Science. A collaboration between ARDC, ACOLA and Australia's Five Learned Academies*. https://www.science.org.au/files/userfiles/about/documents/Australia's%20data%20enabled%20research%20future%20Science-23JUN22.pdf

Australian Academy of Science, 2022. *Australia's data-enabled research future: Science*. https://www.science.org.au/supporting-science/science-policy-and-analysis/reports-and-publications/australias-data-enabled-research-future-science

Australian Computer Society (ACS), 2021. *The ICT Profession, Professionals and the ACS Body of ICT Knowledge (CBoK)*, Version 2.3. https://www.acs.org.au/content/dam/acs/acs-accreditation/CBoK%20V3.2.pdf

Australian Computer Society (ACS), 2022. *Accreditation Manual 5.3, vol. 2. Accreditation Criteria*. https://www.acs.org.au/content/dam/acs/acs-accreditation/ACS%20Accreditation%205.3%20Volume%202%20Criteria%20Dec22.pdf

Committee for Economic Development of Australia (CEDA), 2021. *A good match: Optimising Australia's permanent skilled migration*. https://www.ceda.com.au/Admin/getmedia/150315bf-cceb-4536-862d-1a3054197cd7/CEDA-Migration-report-26-March-2021-final.pdf

CSIRO, 2022. *Artificial Intelligence for Science report*. https://www.csiro.au/en/research/technology-space/ai/Artificial-Intelligence-for-Science-report

Deloitte Access Economics, 2018. *The future of work: Occupational and education trends in data science in Australia*. https://www.deloitte.com/content/dam/assets-zone1/au/en/docs/services/economics/deloitte-au-economics-future-of-work-occupational-education-trends-data-science-170418.pdf

Department of the Premier and Cabinet, Government of South Australia, 2023. *Information sharing and data analytics*. https://www.dpc.sa.gov.au/responsibilities/data-sharing

Government of Canada, 2018. *Report to the Clerk of the Privy Council: A Data Strategy Roadmap for the Federal Public Service*. https://www.canada.ca/en/privy-council/corporate/clerk/publications/data-strategy.html#inline_content

Hope, A., Yemm, N., Nguyen, G., Ali Raza, M., Sparke, E., Buhagier, C. & Neumann, R., 2022. *Digital Skills in the Australian and International economies*, National Skills Commission. https://www.nationalskillscommission.gov.au/sites/default/files/2022-03/ABS%20Paper%20-%20Digital%20Skills.pdf

Indeed, 2023. *Data Science Salary in Australia*. https://au.indeed.com/career/data-scientist/salaries

James Cook University, 2022. *The Future of Data Science*. https://online.jcu.edu.au/blog/future-data-science

Jee, K., 2021. *Is Data Science Dead in 10 Years? True dilemma or alarmist discourse?* Medium. https://towardsdatascience.com/is-data-science-dead-in-10-years-3cde3963552

Jyoti, R., 2022. *Scaling AI/ML Initiatives: The Critical Role of Data* [White Paper]. International Data Corporation. https://www.snowflake.com/wp-content/uploads/2022/03/Scaling-AI-ML-The-Critical-Role-of-Data-.pdf

LinkedIn, 2020. *2020 Emerging Jobs Report Australia*. https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/emerging-jobs-report/AUS-TOP-EMERGING-JOBS_compressedRevised.pdf

McKinsey & Company, 2022. *The state of AI in 2022 - and a half decade in review* [McKinsey Global Report]. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review?cid=eml-web

National Skills Commission, 2020. *Emerging occupations: How new skills are changing Australian jobs*. https://www.nationalskillscommission.gov.au/publications/emerging-occupations

Polaris Market Research, 2022. *Data Science Platform Market Size Global Report 2022-2030*. https://www.polarismarketresearch.com/industry-analysis/data-science-platform-market

SEEK, 2023. *Data Science Jobs Australia*. https://www.seek.com.au/data-scientist-jobs

Simplilearn, 2022. *The Most In-demand and Highest Paying Jobs in Australia 2023*. https://www.simplilearn.com/highest-paying-jobs-in-australia-article

Statistical Society of Australia (SSA), 2019. *Accreditation for University Courses. Retrieved September 2023*. https://statsoc.org.au/resources/Documents/Accreditation/Accreditation%20of%20degree%20programs-%20SSA%20Rules.pdf

Statistics Canada, 2022. *Statistics Canada Data Strategy 2019 to 2022*. https://www.statcan.gc.ca/en/about/datastrategy

Stitch, 2015. *The State of Data Science Benchmark Report*. https://www.stitchdata.com/resources/the-state-of-data-science/

Appendix 1: Context Scan

The results of context scan are included in the list of references. The statements made in Section 2.1.2 of the Report on the profile of data science in Australia in 2022 are based on a blog about the future of data science (James Cook University, 2022). The evidence quoted in the blog is based on SEEK and LinkedIn reports, US Bureau of Statistics data as well as other reports listed in the references.

The statements made in Section 2.1.3 of the Report on education and capability uplift in data science are based on the CSIRO Report on Artificial Intelligence for Science (2022, pp. 39–51). Some other relevant statements in this Report include:

- An analysis of 18 universities across 9 countries found that the number of undergraduate courses teaching students skills necessary to build and deploy AI models doubled from 2016 to 2020, and increased by 42% for postgraduate courses. Enrolments in introductory courses for AI or machine learning have grown by close to 60% over the same time.
- Data from the OECD AI Policy Observatory shows the number of AI courses (delivered in English) worldwide increased by 80.1% between 2018–2021.
- In Australia the number of AI courses offered by universities has grown 1.2 times over the past 4 years, with 235 courses on offer in 2021. Tertiary educational and vocational training institutes in Australia are offering a growing range of studying opportunities for people seeking to gain AI skills.
- In Australia, the fastest growth in AI course offerings was observed for masters degrees.
- In Australia, 37,587 AI-related job advertisements were posted in 2015–2019 by Adzuna – a leading job advertisement search engine.
- An analysis conducted by the OECD examined the prevalence of AI skills across occupations using LinkedIn member profiles in 2015–2020. This analysis found Australia ranked in the middle of the list (13th place out of 26 OECD countries).

Appendix 2: Standards for data science professionals and degrees

There is currently no commonly agreed set of standards for data science professions in Australia.

The Australian Computer Society (ACS) provides accreditation of professional data science specialisations within ICT degrees – any degrees with "a title identifying it as in the Data Science domain" and the aim "to produce graduates prepared for a role in Data Science" (ACS, 2022; ACS, 2021). ACS accreditation of professional and advanced professional data science specialisations refers to the international Skills Framework for the Information Age (SFIA) and requires at least 1 EFTSL (8 subjects) of content from an appropriate internationally recognised body of knowledge. As an example, the Data Science Body of Knowledge (DS-BoK) developed by EDISON (IABAC-EDSF-DSBOK-R2.pdf) presents five areas of knowledge – with only one area of knowledge, Data Analytics, specifically outlining required mathematical sciences content, encompassing machine learning, statistical methods and business analytics. The list of ACS accredited courses can be found here: Accredited Courses | ACS. The list includes a number of accredited data science degrees.

On the mathematical sciences side, the SSA provides professional accreditation to individuals as Accredited or Graduate Statistician. In addition, university degree programs can receive accreditation if they meet certain criteria (SSA Accreditation Rules). An accredited degree would require a minimum of 25% of a year's study in statistics at second year level, and a minimum of 50% of a year's study in statistics at third year level. The degree should include statistical inference, data analysis, statistical communication skills and the use of a statistical package. To qualify, students should also be required to take several units covering probability and distribution theory, linear models, design of experiments, sampling methods, multivariate analysis, analysis of categorical data, time series, survival analysis, statistical consulting, statistical graphics, and databases. Although there is no separate path for accreditation of data science degrees or data science professionals by the SSA, it has accredited a number of data science degrees according to its usual criteria. The list of SSA accredited courses can be found here: https://www.statsoc.org.au/Accredited-courses.

One approach that has been taken in the United Kingdom is a collaborative one. The Alliance for Data Science Professionals (2022) was established in 2021 through an MOU between seven prestigious entities in the UK. The Alliance for Data Science Professionals encompasses leading organisations in mathematics, statistics and IT.

The member entities comprise the Royal Statistical Society, the Operational Research Society, the Royal Academy of Engineering, National Physical Laboratory (NPL), the Royal Society, the Institute of Mathematics and its Applications, the Alan Turing Institute and the BCS Chartered Institute for IT. The remit of the Alliance was to create a common set of standards for data scientists, data engineers, data analysts and data stewards. These standards have been used as criteria to accredit data science degrees and data science modules of associated degrees. Their work includes:

- Defining the standards of professional competence and behaviour expected of people who work with data which impacts lives and livelihoods. These currently include data scientists, data engineers, data analysts and data stewards.
- Delivering these standards as data science certifications offered by the Alliance members to their professional members, with processes to hold certified members accountable for their professional status in this area.
- Using these standards as criteria for Alliance members to accredit data science degrees, and data science modules of associated degrees, as contributing to certification.

The Certification Guide provides details for certification as a Data Science Professional or Advanced Data Science Professional. The Data Science Standard has five main Skill Areas:

a. Data privacy and stewardship
b. Definition, acquisition, engineering, architecture, storage and curation
c. Problem definition
d. Problem solving
e. Evaluation and reflection.

Evidence of consideration of ethics and efficacy is embedded in skill areas.

Two standards are described for these areas: General and Deep.

The Advanced Data Science Professional requires an applicant to provide deep understanding and knowledge in Skill Area E and at least two other Skill Areas.

The distinction between Data Science Professional and Advanced Data Science Professional level is related to the following: level of responsibility and accountability, level of authority for decision making, technical and organisational complexity, understanding and reach of business impact.

Appendix 3: Review of university data science courses

Australian university websites were searched for bachelors, graduate certificates, diplomas, masters by coursework and research masters degree courses that:

a. were aligned to mathematical sciences and cognate disciplines, including mathematics, statistics, actuarial science, commerce degrees in business analytics, science, computing, artificial intelligence, cybersecurity, information technology, and

b. among those, the courses that specifically included data science, analytics or artificial intelligence in their titles (full title or specialisation).

c. with "course" defined as consisting of multiple subjects culminating in a degree, major or defined specialisation.

Note that the identification of data science degrees was complicated by the fact that they are not listed under a separate Field of Education. Instead, data science is offered most often, but not exclusively, a science, mathematical sciences, computer science or business degree. While for the purposes of this Review the search was mainly focused on degrees offered in mathematical sciences, efforts were made to include as many data science degrees located in other faculties as well. However, some data science degrees may have been missed and the below is intended as an indication only of the types and breadth of degree offerings available. The search results were up to date as of the end of 2022 – in this fast-evolving field it is likely that the degree offerings have increased further at the time of publicationof this report.

Bachelors degree courses

Universities	Degree name	Major name	Offered by (if known)
Australian Catholic University	Bachelor of Commerce (Data Analytics)		Faculty of Law & Business, Peter Faber Business School
	Bachelor of Applied Data Analytics		ANU College of Engineering & Computer Science
Bond University	Bachelor of Business Data Analytics		Bond BusinessSchool
Curtin University	Bachelor of Science/Advanced Science	Data Science	School of Elec Eng, Comp & Math Sci (EECMS)
Curtin University	N/A	Statistical Data Science Specialisation	School of Elec Eng, Comp & Math Sci (EECMS)
Deakin University	Bachelor of Data Science		Faculty of Science, Engineering & Built Environment
	Bachelor of Artificial Intelligence		Faculty of Science, Engineering & Built Environment
Edith Cowan University	Bachelor of Science	Data Science	
Griffith University	Bachelor of Science	Data Science	Faculty of Science & Environment
James Cook University	Bachelor of Science	Data Science	Science & Engineering
	Bachelor of Advanced Science	Data Science	Science & Engineering
La Trobe University	Bachelor of Science	Data Science	Department of Mathematical & Physical Sciences
	Bachelor of Information Technology	Data Sciencel Artificial Intelligence	School of Information Technology
	Bachelor of Business Analytics		Business School department of Accounting, Data Analytics, Economics & Finance
	Bachelor of Cybersecurity	Data Sciencel Artificial Intelligence	School of Information Technology
	Bachelor of Computer Science	Data Science	School of Information Technology
Macquarie University	Bachelor of Information Technology	Data Science	Faculty of Science & Engineering, School of Computing
	Bachelor of Information Technology	Artificial Intelligence	Faculty of Science & Engineering, School of Computing

Universities	Degree name	Major name	Offered by (if known)
Monash University	Bachelor of Applied Data Science/ Advanced		School of Mathematics
	Bachelor of Computer Science in Data Science		School of IT
Murdoch University	Bachelor of Data Analytics		
Queensland University of Technology	Bachelor of Data Science		School of Mathematical Sciences/School of Information Technology
	Bachelor of Data Science/double degrees		
RMIT University	Bachelor of Data Science		School of Computing Technologies
	Bachelor of Computer Science	Cyber Security, minor in Data Science	School of Computing Technologies
Southern Cross University	Bachelor of Information Technology	Big Data Technology & Networks & Cyber security	
Swinburne University of Technology	Bachelor of Computer Science - Artificial Intelligence		Department of Computing Technologies
	Bachelor of Computer Science - Data Science		Department of Computing Technologies
	Bachelor of Data Science		Department of Computing Technologies/ Department of Information Technology
	Bachelor of Business Analytics & Analysis		Department of Information Technology/ School of Business. Law & Entrepeneurship
	Bachelor of Business Analytics & Analysis/ Bachelor of Business		School of Business, Law & Entrepreneurship
Torrens University Australia	Bachelor of Software Engineering (Artificial Intelligence)		Business|Technology
University of Adelaide	Bachelor of Applied Data Analytics		
University of Canberra	Bachelor of Business Informatics/ Information Technology/Software Engineering	Specialist Major in Data Science	Faculty of Science & technology
University of Melbourne	Bachelor of Science	Data Science	School of Mathematics & Statistics/ School of Computer Science
University of New England	Bachelor of Computer Science	Data Science	
	Bachelor of Business	Analytics & Informatics, Business Analytics	UNE Business School
University of New South Wales	Bachelor of Data Science & Decisions	Business Data Science	School of Mathematics & Statistics, with Computer Science & Engineering, & Economics
	Bachelor of Data Science & Decisions	Computational Data Science	School of Mathematics & Statistics with Computer Science & UNSW Business School
	Bachelor of Data Science & Decisions	Quantitative Data Science	School of Mathematics & Statistics with Computer Science & UNSW Business School
University of Newcastle	Bachelor of Data Science		
	Bachelor of Data Science/double degree with Mathematics or Computer Science		Information & Physical Sciences
University of Queensland	Bachelor of Mathematics (hons)	Data Analytics & Operations Research	School of Mathematics & Physics
	Bachelor of Commerce	Business Analytics	Faculty of Business, Economics & Law
University of Southern Queensland	Bachelor of Information Technology (Artificial Intelligence & Data Science)		
University of Sydney	Bachelor of Science	Data Science	School of Mathematics & Statistics
	Bachelor of Commerce	Business Analytics	
University of South Australia	Bachelor of Mathematics (Data Science)		
University of Tasmania	Bachelor of Science	Data Science	School of Mathematics & Physics
University of Technology, Sydney	Bachelor of Artificial Intelligence		Faculty of Engineering & Information Technology

Universities	Degree name	Major name	Offered by (if known)
University of Western Australia	Bachelor of Science	Data Science	School of Physics, Mathematics & Computing/ Department of Computer Science & Software Engineering
	Bachelor of Advanced Computer Science (Data Science)		School of Physics, Mathematics & Computing/ Department of Computer Science & Software Engineering
University of Wollongong	Bachelor of Data Science & Analytics		School of Mathematics & Applied Statistics
Victoria University	Bachelor of Data Science		College of Engineering & Science
	Bachelor Of Business	Business Analytics	Victoria University Business Scool
Western Sydney University	Bachelor of Data Science		School of Computer, Data & Mathematical Sciences
	Bachelor of Applied Data Science/ Advanced		

Graduate certificate courses

University	Degree name	Offered by (if known)
Australian Catholic University	Grad Certificate in Data Science	Online, North Sydney campus
	Grad Certificate in Data Science	Online, multi state
Charles Darwin University	Grad Cert in Data Science	Online, Casaurina campus
	Foundations of Data Science (Non award short course)	Online, Casuarina campus
Charles Sturt University	Graduate Certificate in Applied Data Science	
Deakin University	Graduate Certificate of Artificial Intelligence	Online
Flinders University	Graduate Diploma in Data Science	College of Science & Engineering
Griffith University	Graduate Certificate in Data Analytics & Cyber Security	Faculty of Engineering IT & Aviation
	Diploma of Computing & Data Analytics	Faculty of Engineering IT & Aviation
James Cook University	Graduate Certificate in Data Science	Science & Engineering
	Graduate Diploma of Data Science	Science & Engineering
	Graduate Diploma of Data Science (Internet of Things)	Science & Engineering
La Trobe University	Graduate Certificate in Business analytics	Business School department of Accounting, Data Analytics, Economics & Finance
Macquarie University	Graduate Certificate of Applied Business Analytics	Macquarie Business School
Monash University	Graduate Diploma in Data Science	Information Technology
	Graduate Certificate of Data Science	Faculty of Information Technology
Murdoch University	Graduate Diploma in Artificial Intelligence & Data Science	
	Graduate Certificate in Business Analytics	Murdoch Business School
Queensland University of Technology	Graduate Certificate in Data Analytics	Faculty of Science
	Graduate Certificate in Business Analysis	Faculty of Business
RMIT	Graduate Certificate in Data Science	School of Computing Technologies
	Graduate Certifcate in Data Science (Online)	School of Computing Technologies
	Graduate Certificate in Analytics	School of Science
Swinburne	Graduate Certificate of Professional Data Analytics - Bootcamp	School of Business, Law & Entrepreneurship
	Graduate Certificate of Professional Data Analytics - Bootcamp (online only)	School of Business, Law & Entrepreneurship
	Graduate Certificate of Data Science	School of Business, Law & Entrepreneurship
University of New England	Graduate Certificate in Data Science	Faculty of Science, Agriculture, Business & Law
	Graduate Diploma in Data Science	Faculty of Science, Agriculture, Business & Law
University of Newcastle	Graduate certificate in Data Analytics	Information & Physical Sciences

University	Degree name	Offered by (if known)
University of NSW	Graduate Certificate in Data Science & Decisions	Faculty of Science, School of Mathematics & Statistics
	Graduate Certifcate in Data Science (Online)	Faculty of Science, School of Mathematics & Statistics
	Graduate Diploma in Data Science (Online)	Faculty of Science, School of Mathematics & Statistics
	Graduate Diploma in Data Science & Decisions	Faculty of Science, School of Mathematics & Statistics
	Graduate Certificate in Analytics	
	Graduate Diploma in Analytics	
University of Queensland	Graduate Certificate in Business Analytics	Faculty of Business, Economics & Lawdata science
University of Sunshine Coast	Graduate Certificate of Applied Data Science	Faculty of Health, Engineering & Sciences
	Graduate Diploma of Science (Applied Data Science)	Faculty of Health, Engineering & Sciences
University of Technology, Sydney	Graduate Certificate in Data Science & Innovation	School of Mathematical & Physical Sciences
	Graduate Certificate in Data Science in Quantitative Finance	Faculty of Science
Victoria University	Graduate Certificate in Business Analytics	Victoria University Business School
Western Sydney University	Graduate Certificate in Data Science	
	Graduate Certificate in Big Data	
	Graduate Diploma in Data Science	

Coursework masters degrees

University	Degree name	Offered by (if known)
Australian National University	Master of Applied Data Analytics	ANU College of Engineering & Computer Science
Bond University	Master of Actuarial Practice (Specialisation in Data Analytics)	Bond Business School
	Master of Business Data Analytics	Bond Business School
	Master of Business Data Analytics (Professional)	Bond Business School
Charles Darwin University	Master of Data Science	College of Engineering, IT & Environment
	Master of Information Technology (Information Systems & Data Science)	College of Engineering, IT & Environment
Charles Sturt University	Master of Information Technology (Specialisations in Business Analysis, Data Science)	Faculty of Business, Justice & Behavioral Sciences
Curtin University	Master of Predictive Analytics - major in Data Science	School of Electrical Engineering, Computing & Mathematical Sciences
	Master of Artificial Intelligence	School of Electrical Engineering, Computing & Mathematical Sciences
	Master of Information Systems & Technology (Specialisation in Data Analytics & Visualisation)	School of Electrical Engineering, Computing & Mathematical Sciences
	Master of Computer Science (Specialisation in Artificial Intelligence & Cyber Security)	School of Electrical Engineering, Computing & Mathematical Sciences
Deakin University	Master of Business Analytics (Major in Data Science)	Faculty of Business & Law
	Master of Data Science	Faculty of Science, Engineering & Built Environment
	Master of Data Science (Professional)	Faculty of Science, Engineering & Built Environment
	Master of Data Science (Global) (online)	Delivered by Great Learning (India) wholly online
	Master of Applied Artificial Intelligence	Faculty of Science, Engineering & Built Environment
	Master of Applied Artificial Intelligence (Professional)	Faculty of Science, Engineering & Built Environment
Edith Cowan University	Master of Data Science	School of Science

University	Degree name	Offered by (if known)
Federation University Australia	Master of Data Science	Institute of Innovation, Science & Sustainability, Science & Mathematics
Flinders University	Master of Business Administration (Business Analytics)	College of Business, Government & Law
	Master of Computer Science (Artificial Intelligence)	College of Science & Engineering
	Master of Data Science	College of Science & Engineering
James Cook University	Master of Data Science (online)	Science & Engineering
	Master of Data Science (Professional)	Science & Engineering
	Master of Data Science (Professional) (online)	Science & Engineering
La Trobe University	Master of Artificial Intelligence	Computing, Engineering & Mathematical Sciences
	Master of Data Science	Computing, Engineering & Mathematical Sciences
	Master of Business Analytics	La Trobe Business School
	Master of Professional Accounting (Business Analytics)	La Trobe Business School
Macquarie University	Master of Business Analytics	Macquarie Business School
	Master of Data Science	Faculty of Science & Engineering, School of Computing
Monash University	Master of Analytics	Faculty of Business & Economics
	Master of Business Analytics	Faculty of Business & Economics
	Master of Applied Data Science	Faculty of Information Technology
	Master of Artificial Intelligence	Faculty of Information Technology
	Master of Data Science	Faculty of Information Technology
	Master of Health Data Analytics	School of Public Health & Preventive Medicine
Murdoch University	Master of Information Technology (Artificial Intelligence & Data Science)	College of Arts, Business, Law & Social Sciences
	Master of Science in Information Technology (Artificial Intelligence & Data Science)	College of Arts, Business, Law & Social Sciences
	Master of Science in Information Technology (Data Science)	College of Arts, Business, Law & Social Sciences
Queensland University of Technology	Master of Data Analytics	Faculty of Science
	Master of Data Analytics (Biomedical Data Science)	Faculty of Science
	Master of Data Analytics (Computational Data Science)	Faculty of Science
	Master of Data Analytics (Statistical Data Science)	Faculty of Science
RMIT University	Master of Analytics	School of Science
	Master of Data Science	School of Computing Technologies
	Master of Data Science Strategy & Leadership (online)	RMIT Online
Swinburne University of Technology	Master of Data Science	School of Science, Computing & Engineering Technologies
	Master of Business Information Systems (Business Analytics specialisation)	School of Business, Law & Entrepreneurship
	Master of Financial Technologies (Business Analytics specialisation)	School of Business, Law & Entrepreneurship
The University of Adelaide	Master of Business Analytics	Faculty of Arts, Business, Law & Economics
	Master of Data Science	Faculty of Science, Engineering & Technology
	Master of Data Science (Applied) (online)	Faculty of Science, Engineering & Technology
	Master of Artificial Intelligence & Machine Learning	Faculty of Science, Engineering & Technology
The University of Melbourne	Master of Applied Data Analytics (specialisations in education, urban, health & psychology analytics)	Melbourne School of Professional & Continuing Education.
	Master of Business Analytics	Melbourne Business School
	Master of Data Science	Faculty of Science
The University of New England	Master of Business Administration (Data & Cyber Management)	UNE Business School
	Master of Data Science	School of Science & Technology
The University of Newcastle	Master of Data Science	School of Information & Physical Sciences
	Master Of Business Administration(Global)/ Master of Science (Data Analytics)	Information & Physical Sciences
The University of Queensland	Master of Data Science	Engineering, Architecture & Information Technology
	Master of Business Analytics	Business, Economics & Law

University	Degree name	Offered by (if known)
The University of Sydney	Master of Data Science	School of Computer Science
	Master of Digital Health & Data Science	School of Computer Science
The University of Western Australia	Master of Data Science	Physics, Mathematics & Computing
	Master of Business Analytics	UWA Business School
Torrens University Australia	Master of Software Engineering (Artificial Intelligence/ Advanced)	BusinessITechnology
University of Canberra	Master of Data Science	Faculty of Science & Technology
	Master of Business Informatics (Specialisation in Data Science)	Faculty of Science & Technology
	Master of Information Technology & Systems (Specialisations in Data Science, AI & Machine Learning)	School of Information Technology & Systems
University of New South Wales	Master of Data Science (online)	3 faculties: Mathematics, Economics & Econometrics, Computer Science & Information Systems
	Master of Data Science & Decisions	Faculty of Science
	Master of Science in Health Data Science	Faculty of Medicine & Health
	Master of Analytics	UNSW Business School
University of South Australia	Master of Data Science	UniSA STEM
	Master of Management (Business Analytics)	UniSA Business
University of Southern Queensland	Master of Data Science	Faculty of Health, Engineering & Sciences
University of Tasmania	Master of Information Technology & Systems (Artificial Intelligence & Data Intelligence)	College of Sciences & Engineering
	Master of Health Information Management (Data Analytics)	Tasmanian School of Business & Economics
University of Technology Sydney	Master of Data Science & Innovation	
	Master of Business Analytics	UTS Business School
	Master of Business Analytics (Extension)	UTS Business School
	Master of Artificial Intelligence	Faculty of Engineering & Information Technology
	Master of Data Science in Quantitative Finance	Faculty of Science
University of Western Australia	Master of Data Science	
University of Wollongong	Master of Business Analytics	Faculty of Business & Law
	Master of Computer Science (Machine Learning & Big Data)	Faculty of Engineering & Information Sciences
	Master of Mathematical Sciences (Data Science)	Mathematics & Statistics
Victoria University	Master of Business Analytics	VU Business School
Western Sydney University	Master of Artificial Intelligence	
	Master of Data Science	
	Master of Artificial Intelligence	
	Master of Information & Communications Technology/ Master of Data Science	

Research masters degrees

University	Degree name	Offered by (if known)
Queensland University of Technology	Master of Philosophy in Data Science	Faculty of Science, Depts of Mathematical Science & Computer Science

Appendix 4: Survey Questions

Full survey questionnaire

GENERAL INTRODUCTORY QUESTIONS FOR ALL SURVEY RESPONDENTS

Q1: What best describes your current relationship with data science (tick multiple options if relevant):

Undergraduate Student
Postgraduate Student
Early Career Professional
Academic in a university
Other university professional
Data scientist in industry or government
Manager of data scientists
Job seeker
Retired
Other:

If you are currently employed:

Q2.1: What is your current job title?

Q2.2: Does your job involve "data science"? If so, how?

If you are a job seeker:

Q3: What type of job are you seeking?

Q4: What are your main domains of professional expertise (e.g., mathematics, statistics, machine learning, computer science, health science, business, engineering, etc)?

Q5.1: What is your level of Data Science knowledge?

Very low
Low
Moderate
High
Very High

Q5.2: Please tell us a bit more about your knowledge and training (e.g. your degree):

Q6. Please list up to five skills that you believe are essential for a Data Scientist to have. Possibilities could include aspects of statistics, mathematics, computer science, data wrangling, communication skills, and machine learning.

Skill 1:
Skill 2:
Skill 3:
Skill 4:
Skill 5:

Q7: We have four sets of questions appropriate to different groups of respondents. Please tick which sets of questions you would like to respond to:

A. Do you work in a university?
B. Are you a student?
C. Are you an early career professional?
D. Do you work in industry/government?

QUESTIONS FOR UNIVERSITY STAFF

QA1.1: What data science undergraduate and postgraduate degrees are offered, or are being planned, by your university?

QA1.2: Which department(s) or School(s) manages these data science programs?

QA1.3: How satisfied are you with these degrees and their management?

Very unsatisfied
Unsatisfied
Neutral
Satisfied
Very satisfied

QA2.1: What is the balance between mathematical, statistical and computer science content in your university's data science program(s)?

100% computer science
75%-25%
50%-50%
25%-75%
100% maths and stats

QA2.2: How satisfied are you with this balance?

Very unsatisfied
Unsatisfied
Neutral
Satisfied
Very satisfied

QA3: To what level is the statistical content in the data science program(s) taught by academics who have postgraduate training in statistics?

Very low
Low
Moderate
High
Very High

QA4.1: For any undergraduate degrees in data science offered by your university, what level of high school mathematics is required or recommended for entry?

General Maths (or equivalent)
Maths Methods (or equivalent)
Specialist Maths (or equivalent)

QA4.2: Do you think this level is suitable?

Yes
No
Unsure

QA5: How is your University/Faculty responding to the increase in demand for people who have undergraduate or postgraduate training in data science?

QA6.1: Has the growth of data science affected the employment of statisticians at your university?

Yes
No
Unsure

QA6.2: If so, how?

QA7.1: Has the growth of data science led to (or is leading to) a change in research priorities within your Faculty/Department/School?

Yes

No

Unsure

QA7.2: If yes, how?

QA8: How could or should the Australian Mathematical Sciences Institute support advanced training of data science students, e.g. via its summer/winter schools and/or via its Advanced Collaborative Environment (ACE) network?

QUESTIONS FOR STUDENTS

QB1: Which best describes your current degree?

Undergraduate

Masters by coursework

Masters by research

PhD

Other:

QB2: Which best describes the discipline of your degree?

Data science

Data analytics

Statistics

Mathematics

Computer science

Other:

QB3: What % mix of core data science disciplines does your degree contain?

Mathematics %

Statistics %

Computer Science/Programming %

Machine Learning/AI %

Other %

QB4: The amount of mathematics and statistics content in a degree in data science should be:

Very low

Low

Moderate

High

Very High

QB5.1: To what extent are you learning the skills that you will need for your future career as a data scientist?

Very low

Low

Moderate

High

Very High

QB5.2: What additional skills would you like to learn?

QUESTIONS FOR EARLY CAREER PROFESSIONALS

QC1: Which best describes your most recently completed degree?

Undergraduate
Masters by coursework
Masters by research
PhD
Other:

QC2: Which best describes the discipline of that degree?

Data science
Data analytics
Statistics
Mathematics
Computer science
Other:

QC3: What % mix of core data science disciplines did your degree contain?

Mathematics %
Statistics %
Computer Science/Programming %
Machine Learning/AI %
Other %

QC4: Please describe your current role (e.g. data scientist, data analyst, data engineer)

QC5: Was the mathematics and statistics training you received at university adequate for your current role?

Yes
No

QC6.1: Have you undertaken any post-degree professional training or do you want to undertake any (e.g. formal university courses, bespoke short courses, general online courses)?

Yes
No

QC6.2: If yes, can you please provide more information?

QC7: What broader professional support or community of practice in data science or data analytics do you access or would like to access (e.g. professional societies)?

QUESTIONS FOR INDUSTRY OR GOVERNMENT STAFF

QD1: What role does Data Science play in your operations and strategies?

QD2: How would you rate the level of Data Science knowledge among employees in your organisation?

Very low
Low
Moderate
High
Very High

QD3.1: Has the number of Data Scientists grown in in your organisation in the past 5 years?
Yes
No
Unsure

QD3.2: If yes, what primary tasks are they doing?

QD4.1: Does your organisation intend to employ more Data Scientists in the next 5 years?
Yes
No
Unsure

QD4.2: If yes, what primary tasks will they be doing?

QD5: What skills do you look for in Data Science graduates?

QD6.1: Do Data Science graduates have the requisite skill sets needed for your workplace?
Yes
No

QD6.2: If no, what are they lacking?

QD7.1: How important is training in mathematics and statistics for Data Science graduates?
Not at all important
Slightly important
Moderately important
very important
Extremely important

QD7.2: Why?

QD8: In your opinion, what is the level of demand in industry and government for Data Scientists with the requisite skill sets?
Very low
Low
Moderate
High
Very High

QD9: What is the appetite for upskilling in Data Science among your workforce?
Very low
Low
Moderate
High
Very High

Final questions for all respondents

Q8: What roles can Professional Societies such as the Statistical Society of Australia play in supporting data science training, and data scientist researchers, professionals and students? Examples include professional networks, upskilling and training, accreditation, advocacy, conferences, etc.

Q9: The intent of this survey is to gather information about the role of mathematical sciences in data science in Australia. Is there anything else you would like to say about this topic that hasn't been captured in the questionnaire above?

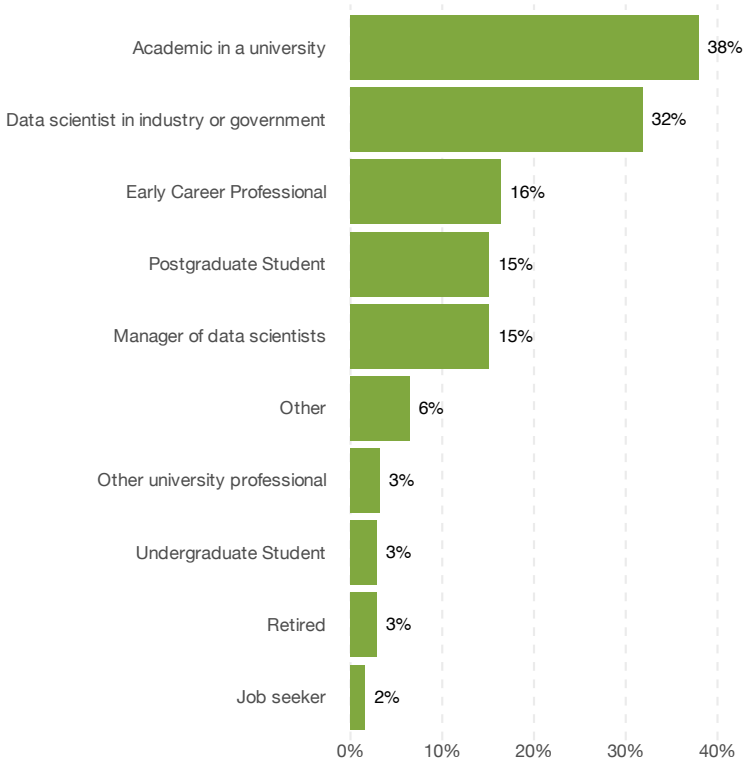
Appendix 5: Detailed results from the survey

Respondent profile

The surveys conducted for this Review were not probability samples from defined sample frames; it was not possible with the resources and time available to conduct surveys of this nature. In several cases, the target populations are relatively small, and an attempt was made to reach all members of these populations. To this end, the survey was distributed via AMSI and SSA newsletters and social media channels, and recipients were encouraged to send on the survey to others. However, without sample frames response rates cannot be determined, which should be considered when the quantitative results are examined. The respondents' self-reported expertise and education show that people who are more strongly engaged with the subject matter – the role of mathematics and statistics in university teaching of data science - were more likely to respond.

A total of 311 survey responses were submitted from September to October 2022. Respondents comprised academics in universities (n=118, 38%), data scientists working in industry or government (n=99, 32%), early career professionals (n=51, 16%), postgraduate students (n=47, 15%), managers of data scientists (n=47, 15%) and others (n=20, 6%), with multiple responses allowed. Of the 311 respondents, 85 chose more than one option.

All respondents: Relationship with data science (multiple choices allowed) (n=311)



All respondents: if you are currently employed, what is your current job title? (n=311)

Job Title	#	%
Analyst	28	9.0
Professor	28	9.0
NA	26	8.4
Associate Professor	19	6.1
Data Scientist	15	4.8
Leadership Position	14	4.5
Lecturer	14	4.5
Senior Lecturer	13	4.2
Director	11	3.5
Consultant	9	2.9

All respondents: Main domain(s) of professional expertise (multiple responses allowed) (n=311)

Domain expertise	#	%
Statistics	138	44.3
Mathematics	86	27.7
Machine Learning	61	19.6
Computer Science	44	14.1
Health Science	14	4.5
Applied Mathematics	13	4.2
Business	13	4.2
Applied Statistics	10	3.2
Biostatistics	10	3.2
Physics	10	3.2

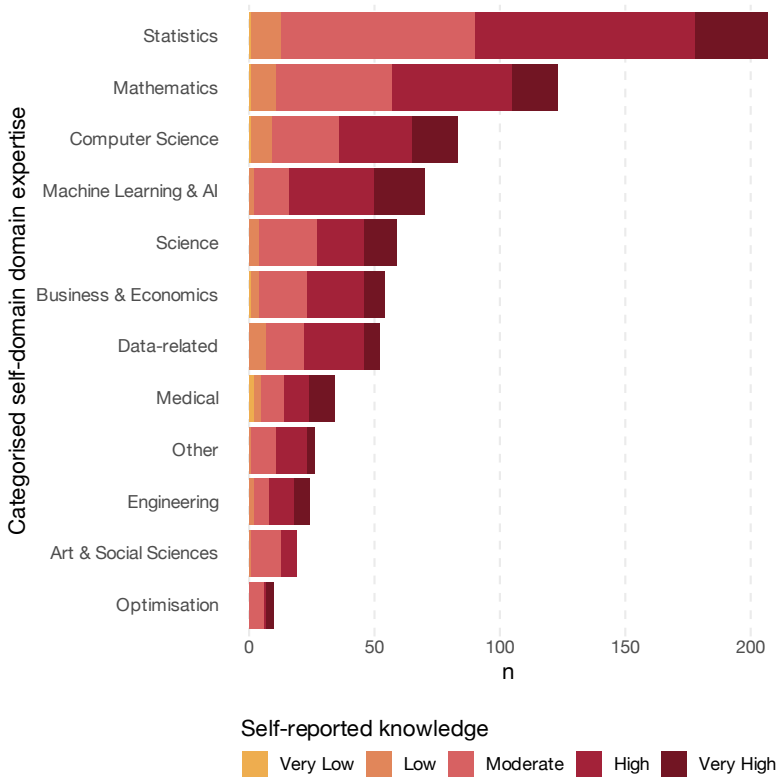
Out of the 311 respondents 151 (49%) indicated they have a PhD, 47 (15%) a masters degree, and 45 (14%) a bachelors degree. The respondent cohort is predominantly made up of mathematical scientists and cognate disciplines, with 138 respondents coming from a statistical background (with another 20 in applied statistics and biostatistics) and 86 from a mathematical background (with another 13 in applied mathematics).

All respondents: Self-reported degrees (only most advanced reported degree included) (n=311)

Self-reported degrees	#	%
PhD	151	49%
Masters	47	15%
BA	45	14%
Degree not specified	25	8%
General knowledge/self-taught/experience	17	5%
Other	13	4%
Honours	12	4%

The self-reported level of data science knowledge among the respondents was predominantly high (38%) or moderate (37%). The reported level of data science knowledge was highest among those who reported machine learning as their domain of expertise.

Self-reported level of knowledge of data science, by self-reported domain expertise



General questions for all respondents

All respondents were invited to list up to 5 skills that they think are essential for a data scientists. As this was an open question, the responses were analysed and skills were grouped together into a consolidated list, see below.

All respondents: Essential skills for a data scientist (n=311)

Skill Description	Count
Statistics	218
Communication	170
Programming	115
Data wrangling	110
Mathematics	74
Computing	73
Machine learning & AI	58
Critical thinking & problem-solving	52
Data visualisation	51
Data analysis & data literacy	49
Psychological traits (open-mindedness, curiosity, etc)	43
Modelling	34
Database	30
Domain knowledge	22
Software development	13
Scientific research skills	12
Linear algebra	11
Understand data	11
Interpersonal skills	10
Optimisation	9
Business acumen	8
Python/R	8

Skill Description	Count
Understand statistical concepts	8
Other	7
Study design	7
Ethics	6
Probability	6
Project management	6
Quantitative & computational literacy	6
Reproducibility & version control	6
Big data & hpc	5
Computer tools	5
Form question/hypothesis	4
Data engineering	3
Computer hardware	2
Computer science	1

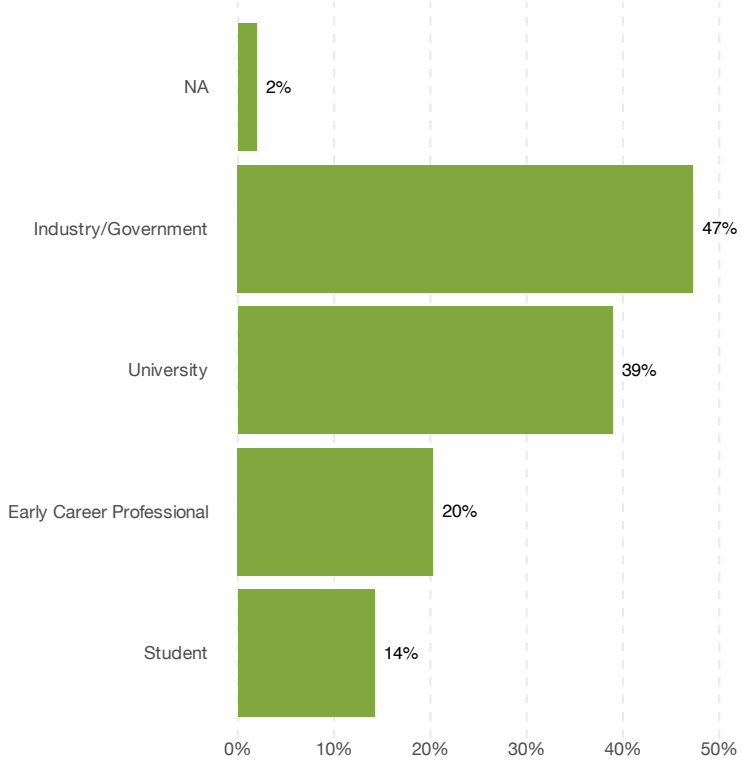
All respondents: Potential roles of professional societies such as SSA in supporting data science training, and data scientist researchers, professionals and students (open textbox) (n=311)

	Count	%
Upskilling and training	139	45%
Networking	80	26%
Accreditation	73	23%
Conferences/workshops	72	23%
Advocacy	56	18%
Raising awareness of DS and its definition	12	4%
Careers, recruitment and job seeker support	11	4%
Mentoring	5	2%
Other/don't know	99	32%

Targeted survey questions

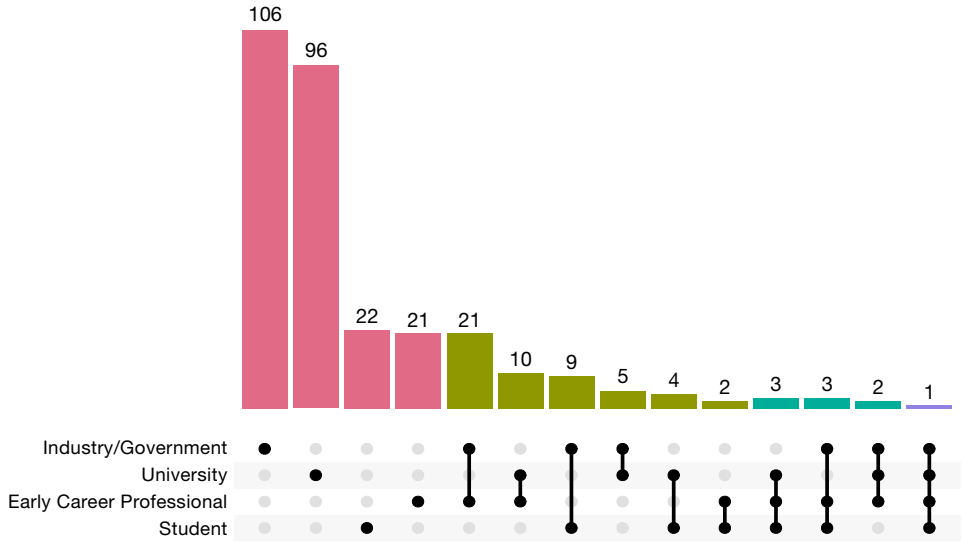
In addition to general questions, the survey included specific sets of questions for industry and government employees (n=147), university staff (n=121), early career professionals (n=63), and students (n=44). Respondents were able to answer more than one specific set of questions - an opportunity taken by 60 survey respondents.

All respondents: We have four sets of questions appropriate to different groups of respondents. Please tick which sets of questions you would like to respond to (multiple responses allowed) (n=311)*



*Note: As respondents could choose to respond to multiple sets of questions, the total of percentages is not equal to 100.

All respondents: Single and multiple choices to respond to different sets of questions for different groups of respondents (n=311)

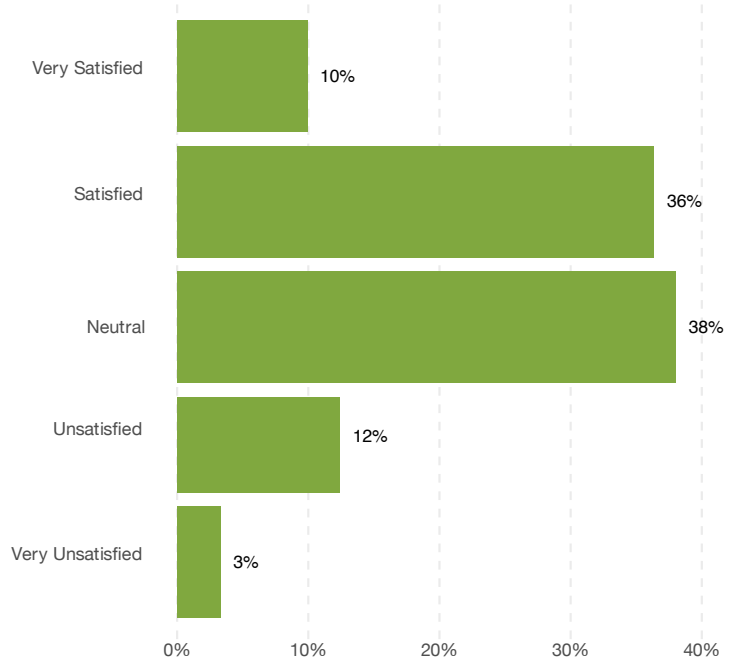


Section A: University staff

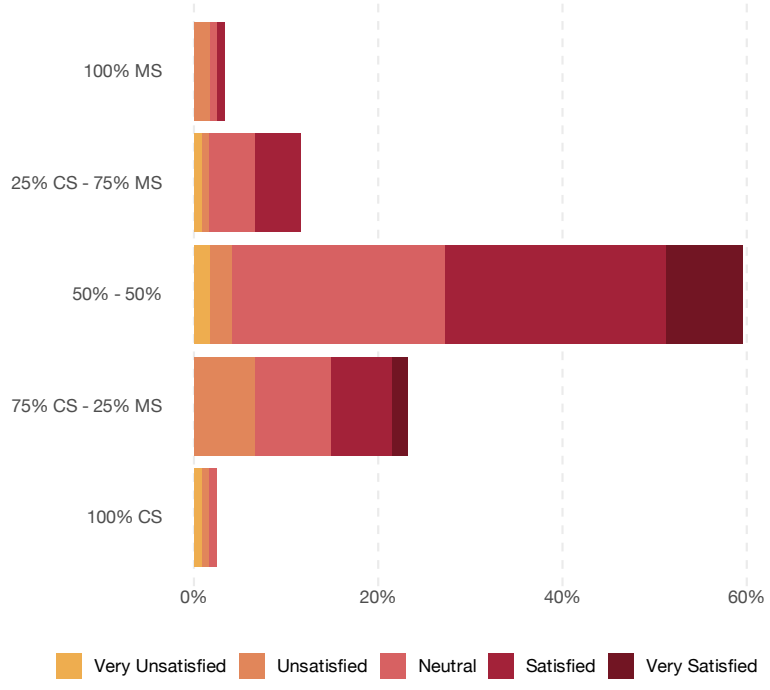
A total of 121 respondents identified themselves as university staff and completed this section of the survey.

There was mixed satisfaction with current data science degrees and their management. While only 19 university staff (19 respondents, 15%) reported to be unsatisfied or very unsatisfied, less than half of university respondents (56 respondents, 46%) reported to be satisfied or very satisfied, while 46 (38%) remained neutral.

University staff: Satisfaction with current data science degrees and their management (n=121)



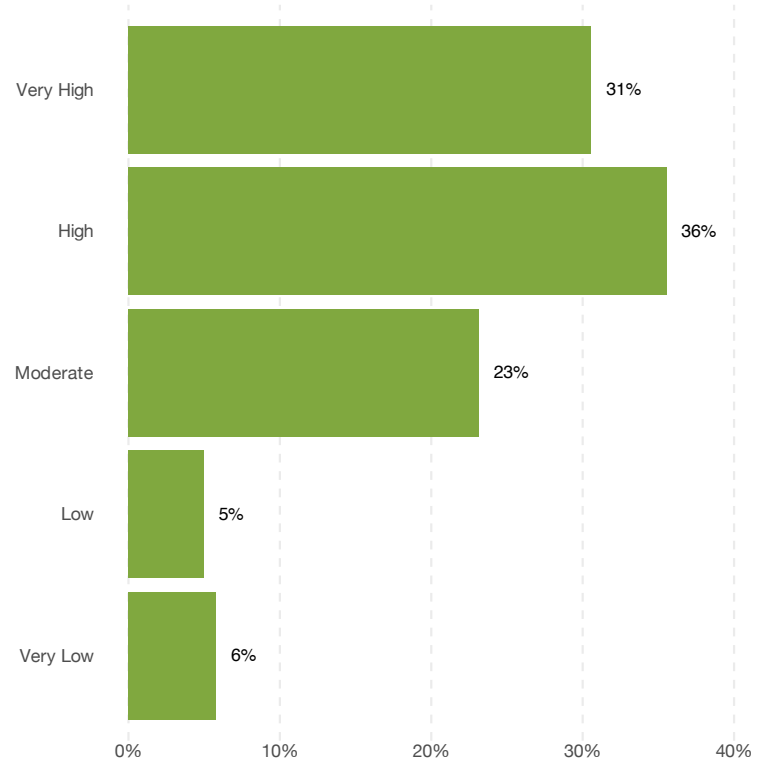
University staff: Balance between mathematical, statistical and computer science content in university's data science program(s); and satisfaction with the current balance (n=121)



When asked about the balance between mathematical, statistical and computer science content in their university's data science program, 72 out of 121 university respondents (60%) stated that it was 50% computer science and 50% mathematical science, a balance that no university respondents were dissatisfied with. The next largest cohort (23%) stated it was 75% computer science and 25% mathematical science (a combination that 11 out of the 28 university respondents were dissatified with).

Of university staff, 80 (66%) report that a high to very high proportion of statistics content in their data science degrees is taught by academics with postgraduate training in statistics.

University staff: Level to which statistical content in the data science program(s) is taught by academics who have postgraduate training in statistics (n=121)



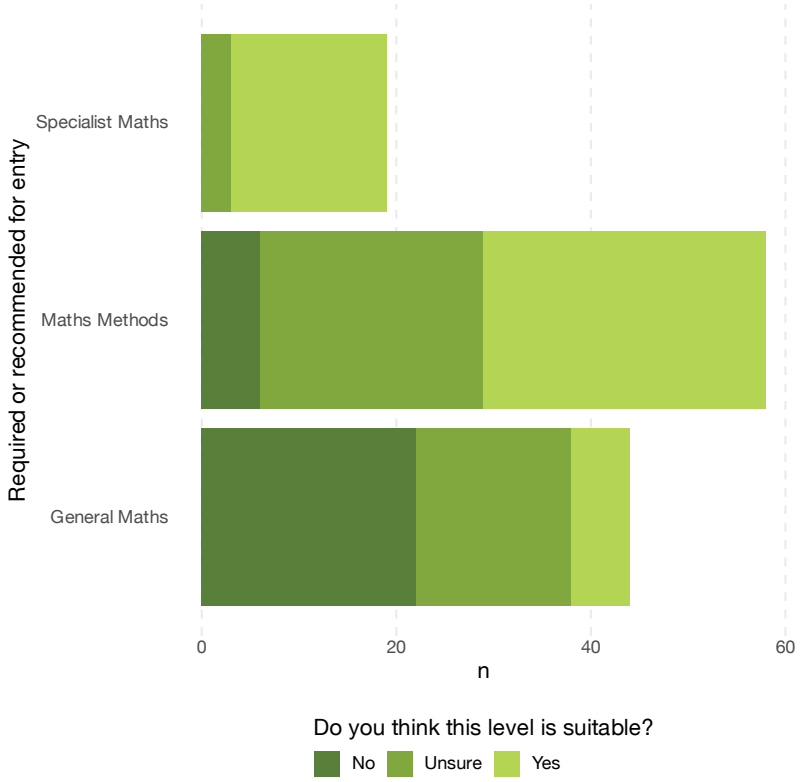
Of the 121 university staff, 48% indicated that Mathematics Methods (or Advanced Mathematics in NSW) was the recommended knowledge level of high school mathematics to start a data science degree at their university, with 36% stating that General Mathematics (Standard Mathematics in NSW) and 16% that Specialist Mathematics (Extension Mathematics in NSW) was recommended for entry.

Where Specialist Mathematics (Extension Mathematics in NSW) is recommended prior knowledge, more than 80% agreed that this provides enough preparation for a data science degree; where Maths Methods or equivalent is recommended as the perceived level of suitability, only half of the university respondents agree that this level is suitable, and around a third is unsure.

Most respondents believed that General Mathematics does not provide sufficient background for entry into a data science undergraduate degree, although a relatively high number reported that they were unsure.

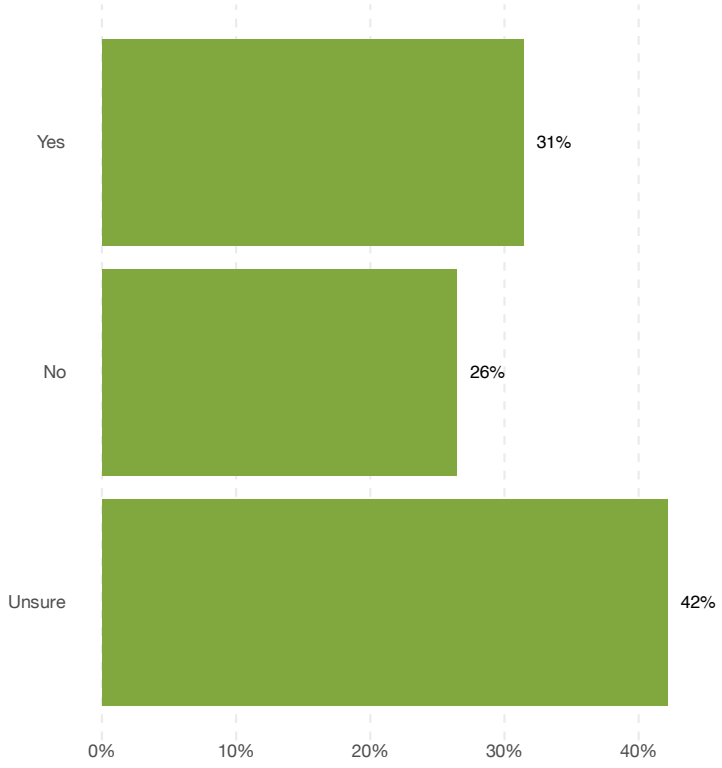
University staff: Recommended high school mathematics level for entry into data science degrees; and perceived suitability of the recommended high school

mathematics level (n=121)



While 31% of university staff respondents agreed that the growth of data science had affected the employment of statisticians at their university, 26% stated that there had been no such effect and 42% were unsure. The same mix of responses was found for whether the growth of data science had affected research priorities within their faculty, department or school.

University staff: Has the growth of data science affected the employment of statisticians at your university (n=121)



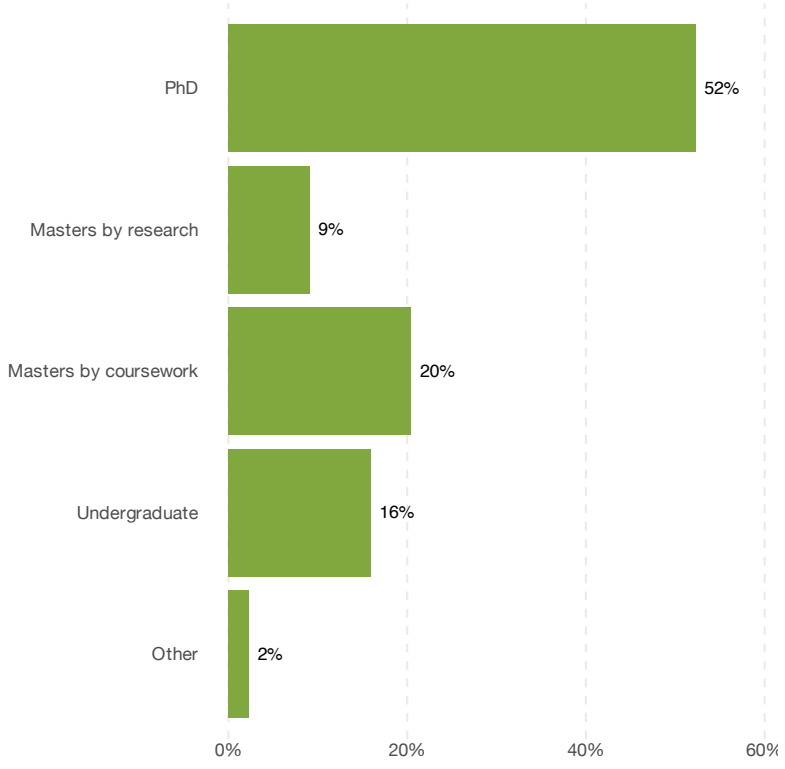
University staff: How AMSI can support advanced training of data science students, according to university staff (n=121)

Summer/Winter Schools	33
Other	23
ACE Network offerings	19
Don't know	18
Bridging/gap/short courses	9
Separate DS event/conference	8
Advocacy Federal Government	7
Data science standards/accreditation/quality control	7
APR Intern/other internships	5
Industry involvement	5
Career awareness	4
Information sharing among universities	1
Partnering with computer science	1

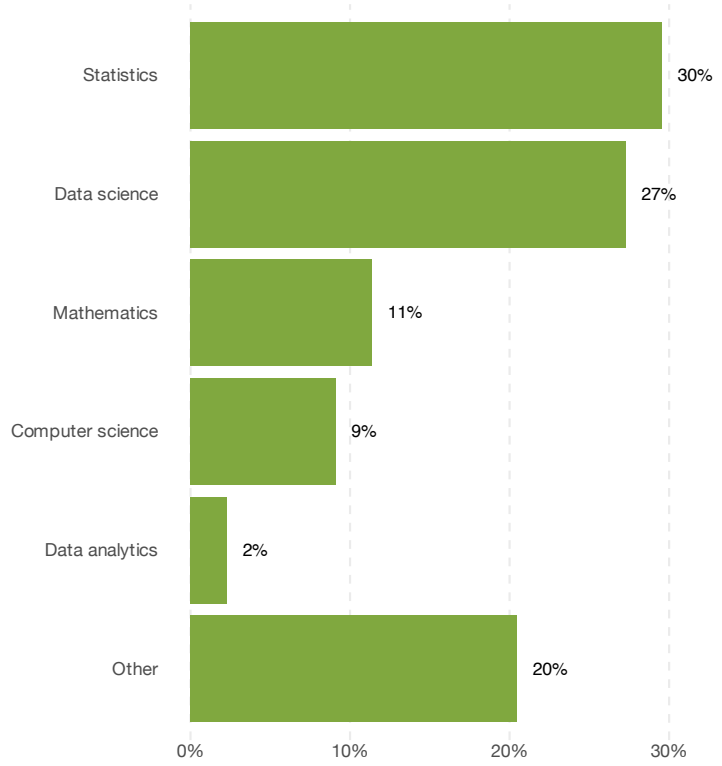
Section B: Students

A total of 44 respondents identified as students and completed this section of the survey. More than half are PhD students. The Students: Discipline content distribution of current degree (n=44) graph below lists individuals on the y-axis. The individuals are ordered by discipline with Statistics on the top.

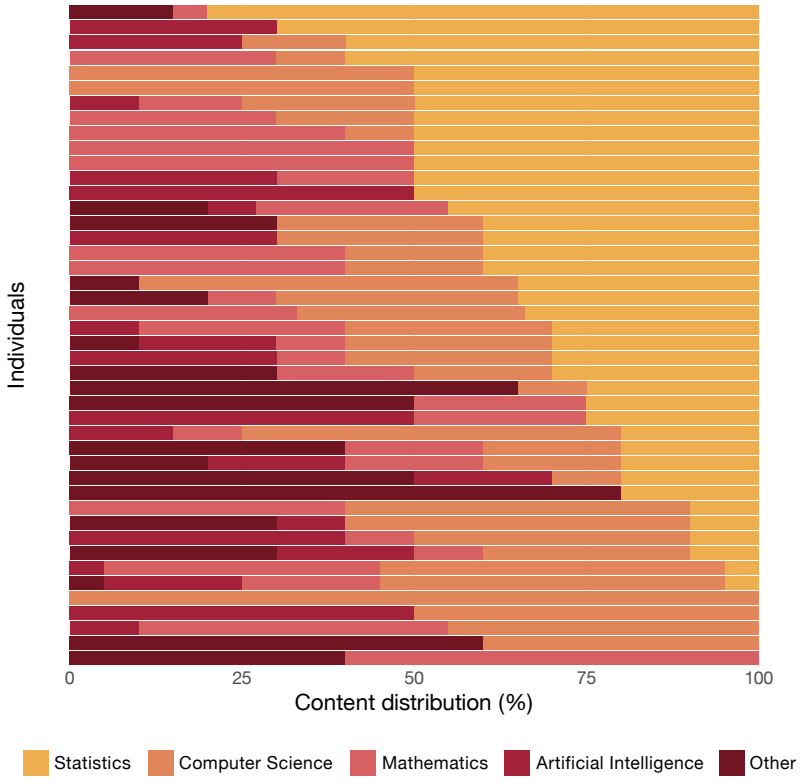
Students: Current degree (n=44)



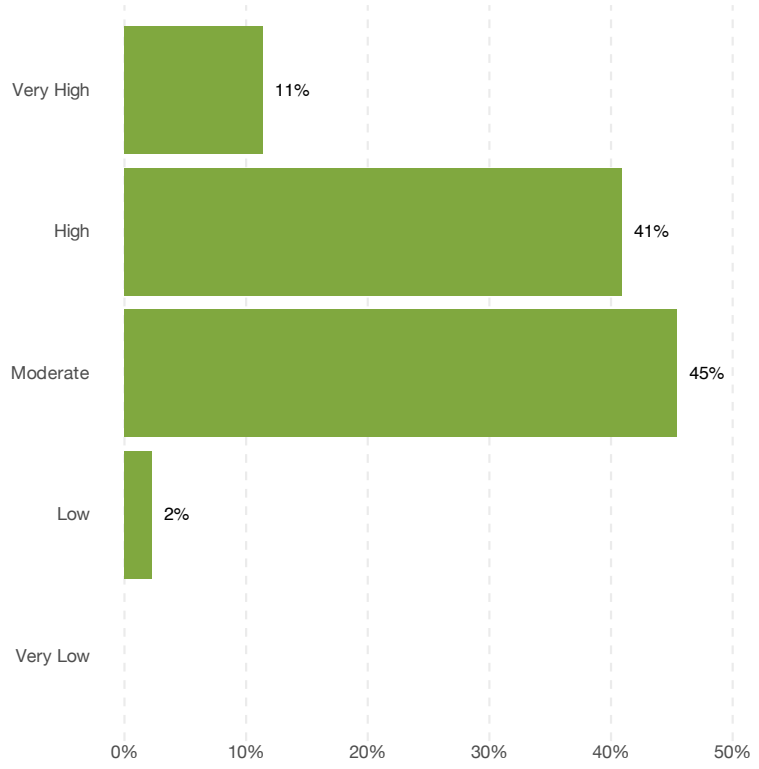
Students: Discipline of current degree (n=44)



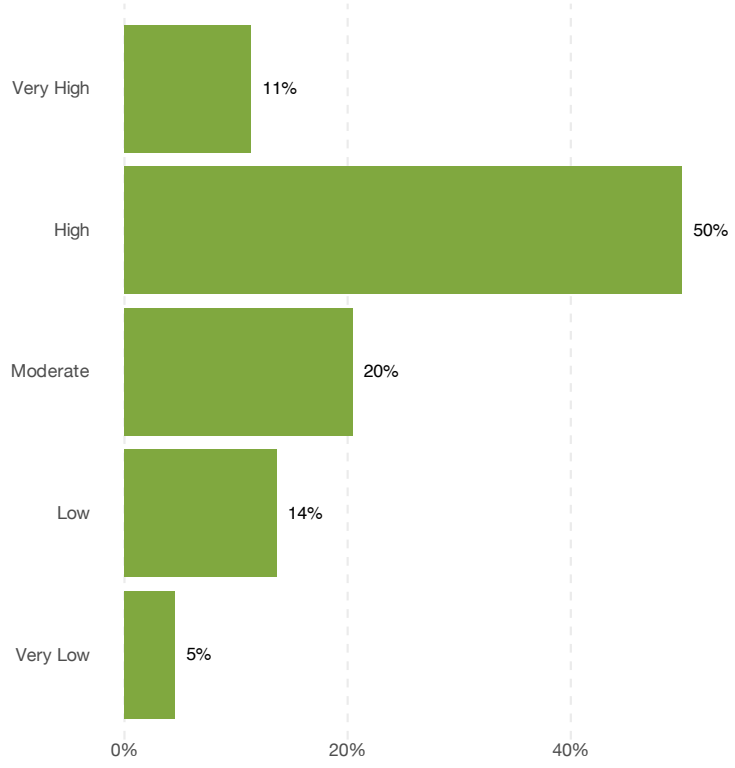
Students: Discipline content distribution of current degree (n=44)



Students: The desired amount of mathematics and statistics content in a degree in data science (n=44)



Students: To what extent are you learning the skills needed for a future career as a data scientist? (n=44)

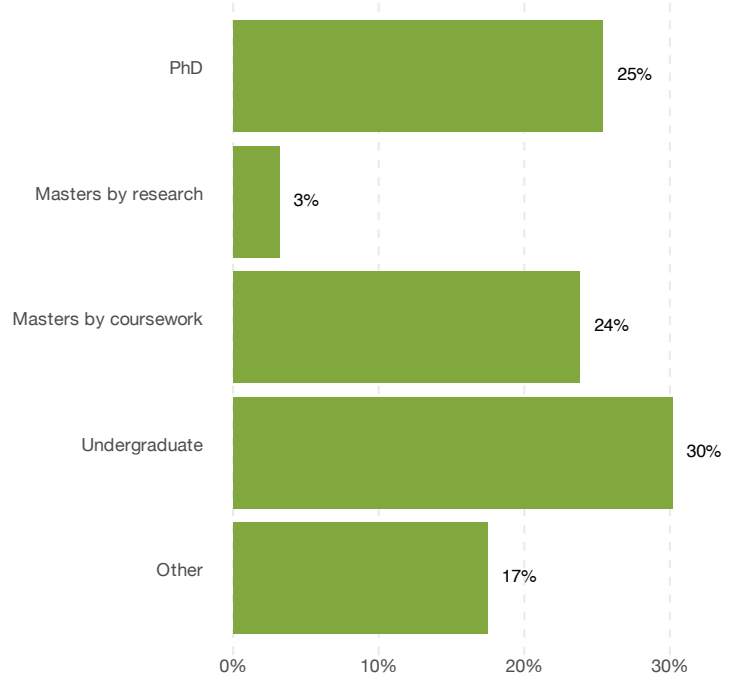


Section C: Early career professionals

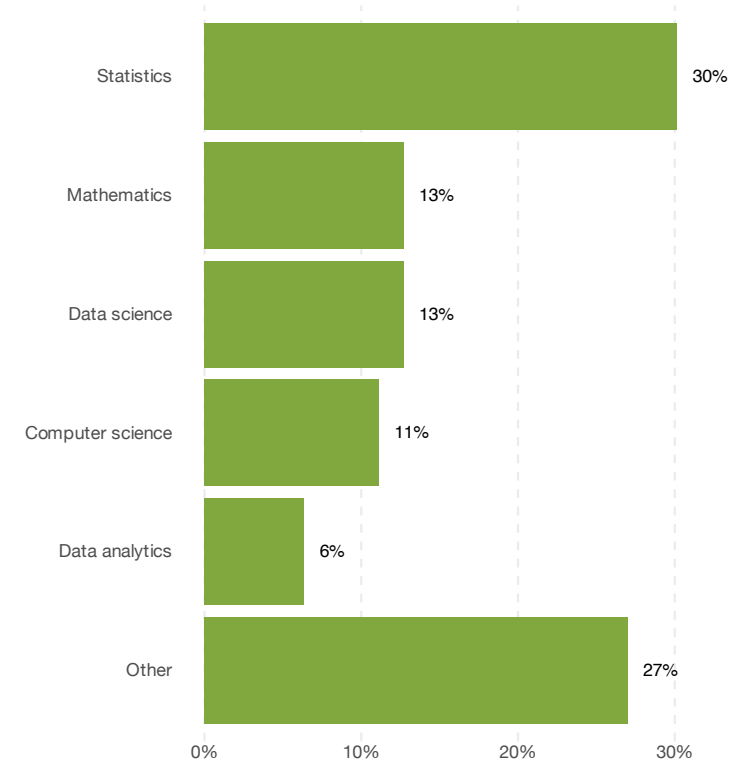
A total of 63 respondents identified as early career professionals and completed this section of the survey. The degree levels were fairly evenly spread, with about 30% having obtained an undergraduate degree, 27% a masters degree, 25% a PhD, and 17% indicating “other” degrees, which in almost all cases was further specified as a graduate certificate or diploma in data science or data analytics.

The Early career professionals: Discipline content distribution of recently completed degree (n=63) graph below list individuals on the y-axis. The individuals are ordered by discipline, using the same top to down order as in the previous graph (with Statistics on the top).

Early career professionals: Most recently completed degree (n=63)

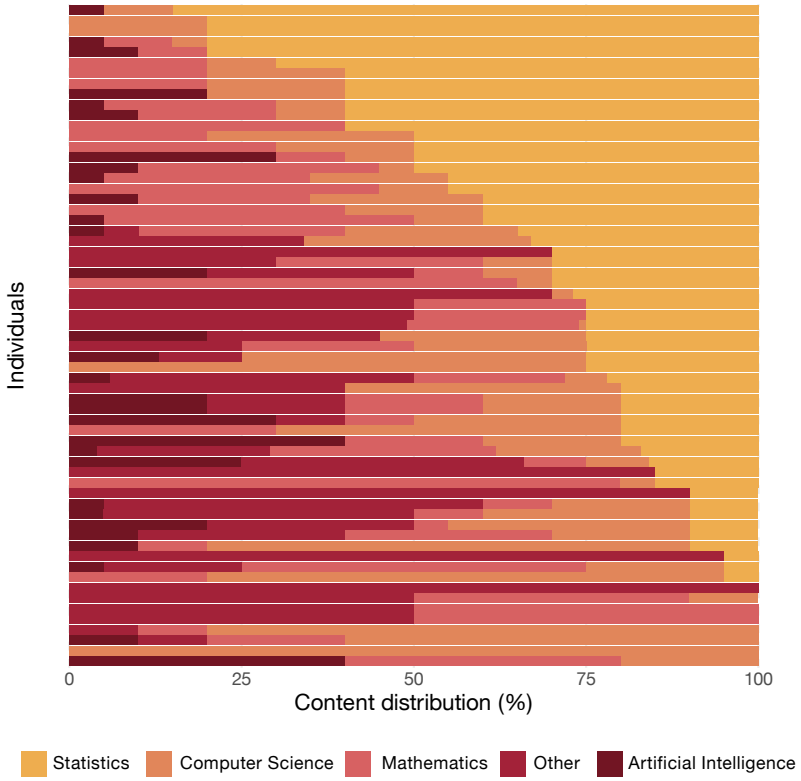


Early career professionals: Discipline of most recently completed degree (n=63)



Of the 63 early career professionals, 17 (27%) indicated they had a degree in an "other" discipline, for example in physics, psychology, social sciences, econometrics and demography.

Early career professionals: Discipline content distribution of recently completed degree (n=63)



Of the 63 early career professionals, over 84% were satisfied that the mathematics and statistics training they received at university was adequate for their current role. Following their degree, over 68% wish to undertake, or have already undertaken, further professional training post-degree. They provided further comments on the type and topics of further training.

Early career professionals: Type/topics for further professional training post-degree (open textbox) (n=63)

Type and topics of further professional training	#
Postgraduate degree	12
Short (online) courses	18
Self-directed online learning	5
Professional development via employer	4
Machine learning	4
Statistics	8
Programming/Coding	4
Visualisation/presentation	3
Survey development	2
Data linking	1
Computer Science	1
Evaluation	1
Data Science/Analytics general	3
Specific software (Python, R, PowerBI, ANN SAS)	6
Other/don't know	13

They also added their input on the access they already have, and suggestions for the access they would like to have, to ongoing professional support.

Early career professionals: Desired access to broader professional support or community of practice in data science or data analytics (e.g. professional societies) (open textbox) (n=63)

(Desired) access to professional support	#
Professional society/networks memberships: not specified	11
Professional society/networks memberships: work-related group	7
Professional society/networks memberships: SSA	12
Professional society/networks memberships: Gender or sexual orientation related	2
Professional society/networks memberships: APS	3
Professional society/networks memberships: domain specific DS related	1
Networking opportunities	9
Training/professional development opportunities: seminars and workshops	6
Training/professional development opportunities: training, courses	9
Training/professional development opportunities: career support	2
Training/professional development opportunities: people management	1
Training/professional development opportunities: online materials	1
Other	6
Don't know	8
Not needed	6

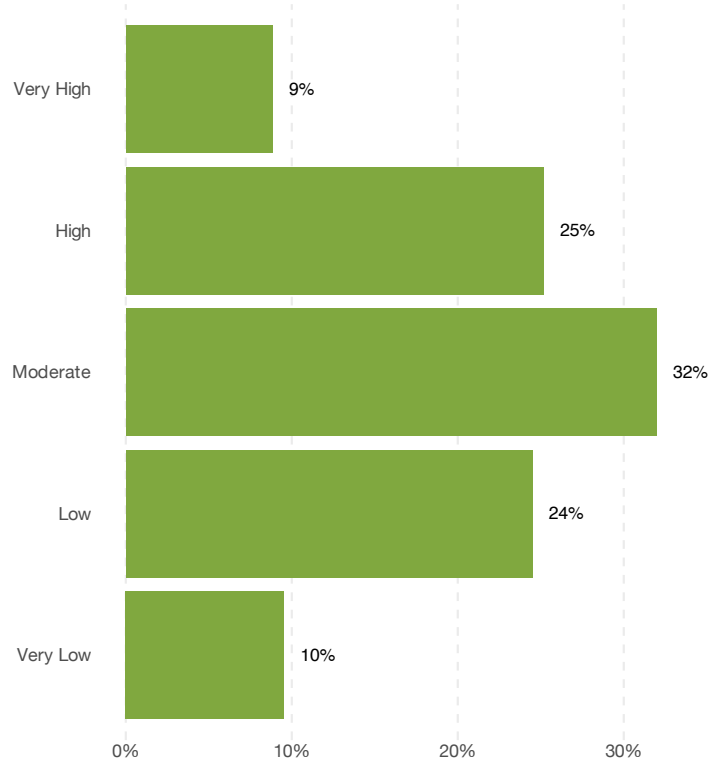
Section D: Industry/government employees

A total of 147 respondents identified as industry or government employees and completed this section of the survey.

Industry/government employees: Role of data science in their organisation (open textbox) (n=147)

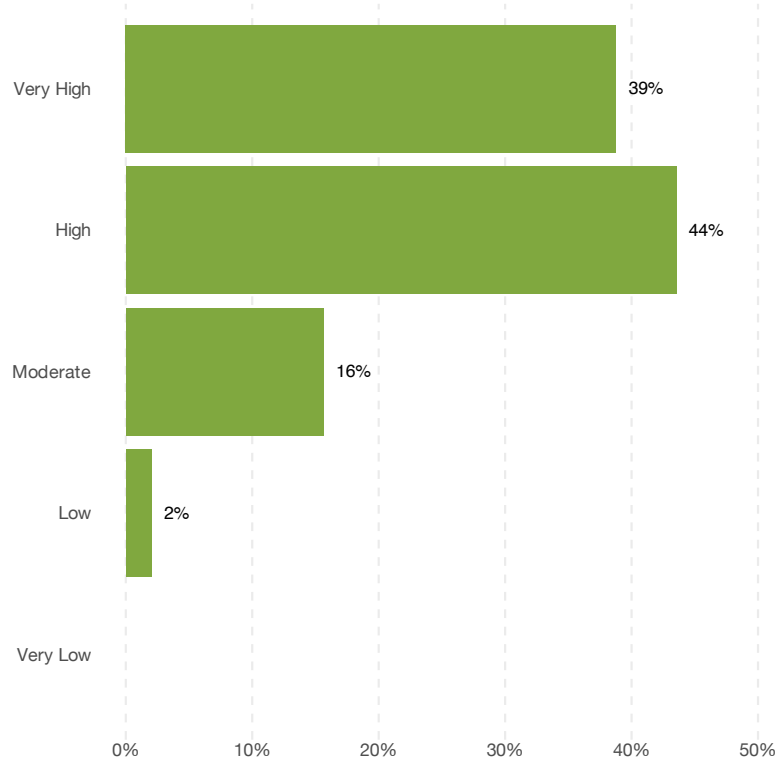
Role of data science in organisation	#
Key/core capability - not further specified	64
Decision making	18
Product/process design and improvement	15
Policy development, analysis and evaluation	14
Reporting/communicating	12
Analytics	8
Modeling	7
Consulting	7
Marketing/customer relations	5
Descriptive/exploratory analysis	4
Predicting/forecasting	4
Tax and financial management	3
Data wrangling/cleaning	3
Risk management/security	2
Survey design	2
Resource deployment	1
Optimisation	1
Visualisation	1
Other/don't know	29

Industry/government employees: Perceived level of data science knowledge in their



Of the industry/government employees, 62% indicated that the number of data scientists had grown in their organisation in the past 5 years (with 24% unsure). In terms of future demand, 52% indicated that their organisation intends to employ more data scientists in the next 5 years – however 42% were unsure of this. The overwhelming perception of the level of demand in the industry and government sectors for graduates with the required skillsets is that it is high (44%) to very high (39%).

Industry/government employees: Perceived level of demand in industry and government for data scientists with requisite skill sets (n=147)

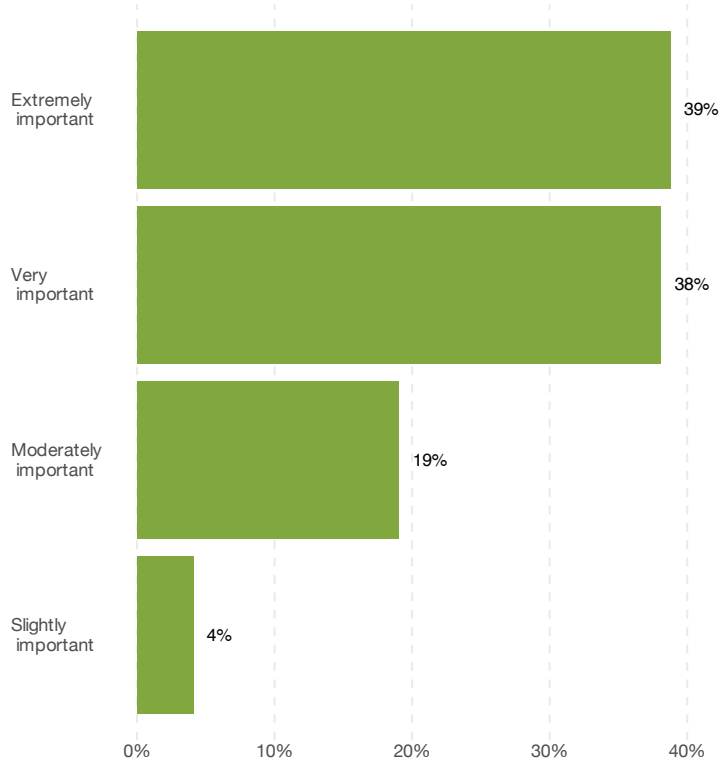


When asked if data science graduates have the requisite skillset for their organisation, 61% responded in the affirmative. However, this leaves 39% (n=57) indicating that graduates do not have the skills they need. When asked, in an open question, what skills are lacking, these 57 respondents included a variety of responses.

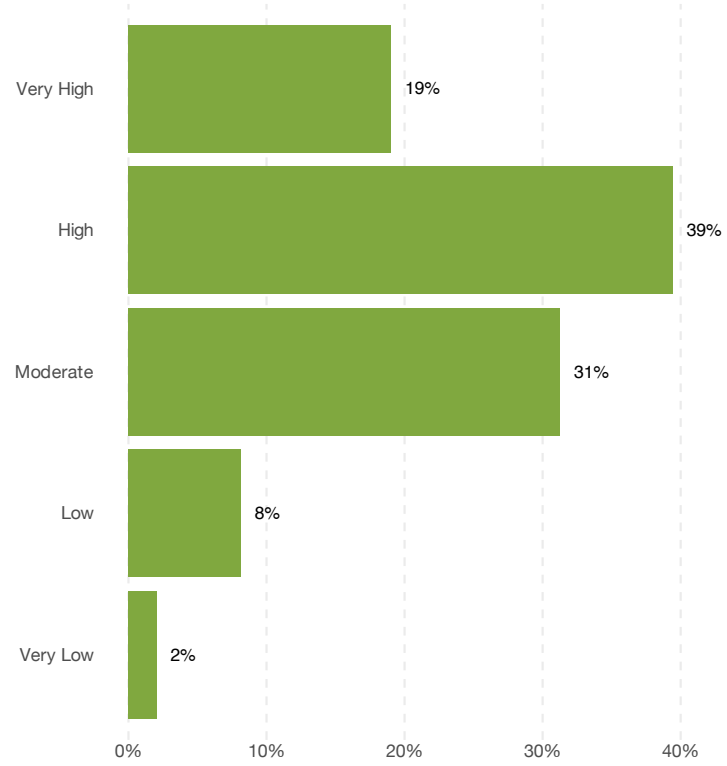
Industry/government employees: If you think data science graduates lack the requisite skills, what skills are they lacking (open textbox) (n=57)

Skills lacking	#	%
Stats skills	10	18%
Experience with real world multidimensional data	8	14%
Communication skills	7	12%
Coding/programming	7	12%
Lack theoretical knowledge	5	9%
Domain knowledge	5	9%
Don't see bigger picture/lack of strategic insight	4	7%
Black box approach	4	7%
Specific software knowledge	3	5%
Ability to self-educate	2	4%
Leadership/ program management	2	4%
Maths skills	1	2%
Data wrangling	1	2%
Stakeholder engagement/management	1	2%
Experience	1	2%
Other/don't know	19	33%

Industry/government employees: Importance of training in mathematics and statistics for data science graduates (n=147)



Industry/government employees: appetite for upskilling in Data Science among your workforce (n=147)



Appendix 6: Focus group questions

The focus group questions were used as a guideline for the discussions.

All groups

How is the growth in DS currently influencing your employment/workplace and/or study?

How do you think DS will influence and change the Australian industry and university sectors in the next five years?

What are the challenges and opportunities facing data science?

What are the respective roles for government, universities and industry in managing increasing demand for data science?

What are the skill sets that Data Scientists need?

How do we encourage more school students to study the level of mathematics/statistics needed to enter the future DS workforce?

What are the roles of the SSA and AMSI in relation to DS?

Academic

In your opinion, what should the balance between mathematics/statistics/computer science content be in DS degrees at your university?

How can a “good” balance be achieved?

What level and content of statistical training should a DS degree contain?

What level and content of mathematical training should a DS degree contain?

How can the growth in DS teaching be translated into employment and research priorities at your university?

Students

Why have you enrolled in a DS degree?

What factors have influenced your degree choice?

What level of mathematics/statistics have you studied before starting a DS degree?

What level of maths and stats do you think is sufficient background for starting a DS degree?

What job roles will you look for upon graduation?

If you are an international student, what are your thoughts about the opportunities international graduates have post-study in Australia?

Early Career Professionals

How did your previous study prepare you for your current DS related employment?

What level and content of statistical training should a DS degree contain?

What level and content of mathematical training should a DS degree contain?

What post-degree training opportunities in DS should be available?

How can the SSA support ECP in DS in terms of training and accreditation?

Industry

What role does DS play in your organization?

What is the use of mathematical/statistical skill sets for your DS employees?

What level and content of statistical training should a DS degree contain?

What level and content of mathematical training should a DS degree contain?

What role can (should) industry play in offering training opportunities to support new graduates or to upskill employees?

What initiatives are needed to resolve supply shortages in employing Data Scientists?

What is the potential role of internships or work integrated learning opportunities to university DS students?

Appendix 7: Detailed results from the focus groups

Four focus groups were held with identified stakeholder groups:

- industry and government;
- academics,
- students, recent graduates and early career professionals;
- members of the Reference Panel of this Review, comprising selected eminent data science professionals from industry, academia and government.

Participants for the first three groups were selected via a combination of purposive and convenience sampling. The Review Panel collated lists of possible participants with an active interest and involvement in data science across a range of industry, government and universities who were invited to participate in a forum group discussion. Industry and government professionals included representatives from retail and hospitality, government agencies, banking and finance, digital infrastructure, digital services and data analytics consultancy.

The focus group discussions were facilitated by members of the Review Panel with the assistance of AMSI staff. The facilitators used a list of suggested questions as a guideline for the discussions, encouraging other themes and topics to be debated as they arose. The recorded discussions were transcribed verbatim, then analysed and coded by AMSI staff with reference to the Terms of Reference of the Review.

Participation numbers in the four focus group discussions are displayed below.

Focus group	Number of participants
University Academics	9
Industry and Government	9
Students, Recent Graduates and Early Career Professionals	7
Data Science Review Reference Group	8

Common Themes

What is data science?

A common theme among all focus group discussion was the lack of a coherent working definition of what data science actually is, and the ramifications of that lack of coherence for the way data science is taught at universities and practiced in the workplace.

"We've got a perception problem because we can't define it, so our schools can't define it, universities can't define it, and industry can't define it." (Industry and Government)

"My feeling is that data science itself is just overly broad, like it really encompasses, I would say mathematics, computer science, and the subject matter. So it's really just a broad topic." (University Academics)

"…because data science is so new, it's kind of understandable that you know each university will have a different definition [of] what goes into a data science course. Some might be more mathematical, some might be more computational. And even with employers, they also are new to defining a data science, position. So, mainly they are only looking for senior people, but specifically what kind of skill sets, it's not clearly defined." (Students and ECPs)

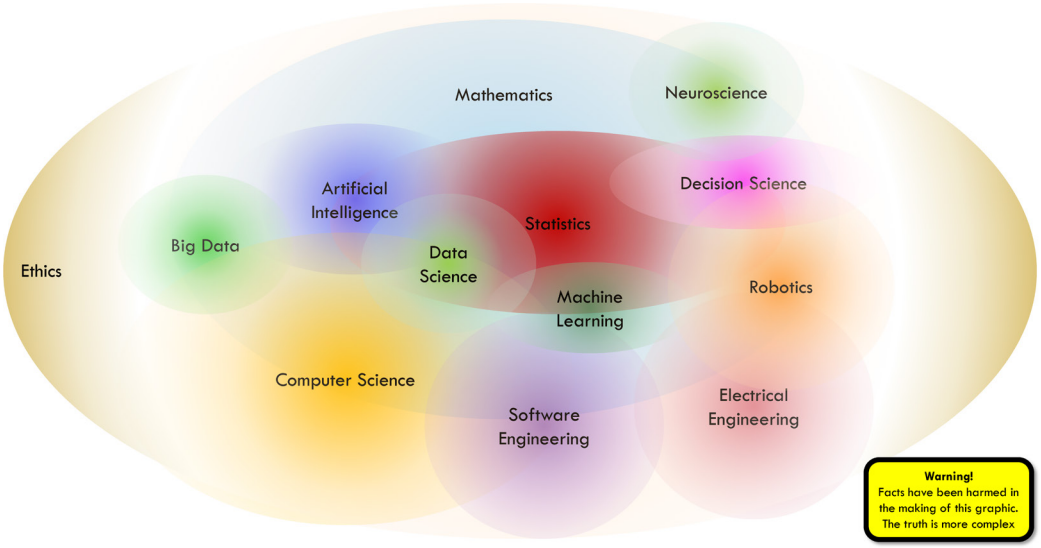
As a new field of economic activity, research and education, data science combines a number of different academic disciplines. Focus group participants emphasised the multidisciplinary and multi-faceted nature of data science – as well as the potential strengths that the different perspectives can offer.

"So basically I see three categories right, and one of them is the traditional statistician who has become a little more computer and computing savvy and has […] taken that opportunity, because […] the technical environment changes. And then you have data scientists that come from […] a bit of a mix, they're good at programming and so on, and they want to go from problem to problem and continue to solve these things from an often exploratory view. And the third category would be people who come from computer science and have a lot better skills than the traditional statistician would have to build a sustainable system." (Industry and Government)

"For me, I think data science is not just making sense of data, but also about the data management. And recently I really put in the data security as well. So it's all aspects of the data. [...] I'm not sure about the university, but the thing that people normally think of data science is just the analysis and modelling, but not much focus on the structure and the management." (University Academics)

"I think data science is a convenient term for a place where lots of disciplines can play well together." (University Academics)

A visual representation of the various disciplines that contribute to data science, volunteered by one of the focus group participants, showcases the diversity of fields involved.



(Image produced by David Lovell and published with permission)

The multidisciplinary nature of data science is also reflected in the very diverse backgrounds and fields of the focus group participants. Some illustrations of the diversity in background and people's entry into data science are below.

"so I did my undergrad in marketing, […] and when I […] went in I thought about graduating and becoming a marketer, […] but as I […] continued my program I learned about different ways, new marketing has been using big data, and how important it is to be able to take this data and produce […] meaningful business feedback." (Students and ECPs)

"so I sort of stepped into data science from the side and having a degree in a different field, quantitative psychology includes a lot of statistics but it's not a statistics degree" (Students and ECPs)

"My first degree was actually in computer science, and I worked in software for a little while, and what I found, […] is that […], clients were giving us a lot of data and collectively, our customers, we know a lot about them, but we weren't using the data in a very effective way." (Students and ECPs)

"I have an engineering and science background. Over the past few decades I've had many points of contact with data, data analysis. Originally environmental data, so contaminant concentrations in soil and groundwater, that kind of thing. And more recently, analysis of fluid mechanics systems and fluid mechanics data. […] So I'm a little bit of an outsider in the sense I wouldn't have called myself a statistician. But data science is very much central to a lot of what we do in science and engineering ". (Reference Group)

"My background is in mathematics both applied and pure, eventually I did a PhD in epidemiology by statistics." (Reference Group)

Some focus group participants pointed out that the lack of a coherent definition of data science requires the employer to be very specific about the actual skills required when recruiting data science professionals.

"… nailing down what a data scientist is is quite difficult, and […] if we were creating or advertising a role, I might well call it data science, but we'd be very likely to define quite clearly what the actual responsibilities were, because they might connect through to data, engineering, or visualization, or the statistical modeling, in quite different ways". (Industry and Government)

"And more relevant than what degree someone has, you know, maths, stats, computer science or anything, is what data structures they've worked with before. Have they worked with dense volumetric data, or have they only worked with row, column type data. And it's actually really relevant, like even geospatial data as distinct from row column data…" (Industry and Government)

One participant pointed out that without a definition of data science or coherent standards, how employer's data science needs are met depends on the current skills of their workforce and can be a matter of chance.

"I think the vagueness of data science means, at least in industry, that I guess the destination is kind of made up of what people happen to already know, and where they are naturally inclined to take things. So it's, in trying to plan, to do any kind of forward planning, it really depends on what skill sets your employees just happen to have, whether that's cloud-based, or whether that's kind of coding-based, or you know, whatever kind of details that might be." (Students and ECPs)

Something similar might apply to the shaping of data science degrees at universities.

"So part of how the data science courses are typically constructed at different universities is really a function of who's there at the time, and what their skills and background and expertise and engagement with industry really are." (University Academics)

The diversity of skills and backgrounds led some focus group participants to conclude that there is not one "type" of data science.

"…what I see with our companies is we have 3 different places that data science plays. There's the traditional decision-type space, […] which is, you know, using data to make better decisions. But then you've also got the people who very much focus on your ML (machine learning-MW) ops and Dev (development – MW) ops, and it's a very technology type, how do we make data pipelines? How do we do cloud-based analysis? And we've also got teams of people who do kind of platform work. So it's like they want to build on interactive chat bots. So they're using different API's to run different kinds of things. All those people call themselves data scientists, but I see them as three very different roles that actually require different skill sets." (Industry and Government)

The diversity in background and disciplines poses both challenges and opportunities.

"There's such a tremendous diversity of disciplines that are relevant to making sense of data […] and I can see that in some situations that plurality is a disadvantage, like if you want to enrol [in] a data science degree, you know exactly what you're going to get, you're at a disadvantage. But flipping that around, for people working in data science, I think it

gives us a great degree of freedom to strike up collaborations and connections which are quite fruitful under that aegis of data science. And you know people feel […] we've got some things in common. We might see the world in different ways, but we have enough in common and enough that's complementary that we could fruitfully work together." (University Academics)

What skills do data scientists need?

The focus group discussions indicate there is no one-size-fits-all exemplar of a data scientist. Many participants emphasise that not everyone involved in data science-related work will need the same skills or be able to operate at the same level.

"There are many hierarchies here from the advanced PhD in a subject major area that is well trained in maths, to somebody who will call themselves a data scientist and basically move things from one excel spreadsheet to another. And there's that whole spectrum." (Industry and Government)

"I guess the required skill depends on what kind of data scientists are we talking about. Are we talking about data scientists who are working in applied areas? Therefore maybe their mathematical knowledge does not have to be at the top level, but as long as they understand the techniques and they're able to apply the techniques and write the reports about them that should be good enough. They don't need to be really really top at mathematics." (Reference Group)

"… when you start working with people in industry, it's like a pyramid. There might be two data scientists, even in a pretty big organisation, but their work is underpinned by many, many people: computer scientists, people who are experts in databases, in building databases and managing databases and so on." (University Academics)

"You need that combination of skills, and I see that more as a combination of professionals, so you know, the data engineers, say the solutions architect, probably a data security professional that might be a solutions architect as well." (Students and ECPs)

In terms of the skills people need working in data science related jobs, the focus group participants' views broadly aligned with the survey results. Participants underscored that it is necessary for data science professionals to have skills such as coding and programming, data management and data wrangling. Data scientists need to be able to problem-solve, to think deeply and reason well.

Given the terms of reference, the discussions included a particular focus on the relative importance (or otherwise) of mathematical and statistical knowledge as part of the data scientists' skill-set.

Some other areas of skill and ability deserve further mention – one such area is "soft" skills and attributes, in particular around communication, teamwork, adaptability and self-directedness. Another area worth mentioning is domain-specific knowledge and contextual understanding.

The importance of mathematics and statistics

Coming up consistently among all focus group discussions is that data scientists need to be able to understand what is going on "under the hood" rather than applying the technological tools available without critical reflection. Among other things, this requires a deep understanding of the tools that are used, how they are used correctly and why – which is all based on a solid statistical and mathematical foundation.

"What we're looking at is trying to have our students positioned […] as people who understand what's going on under the hood rather than just users of the technology. People who can understand the assumptions of the technology, question those assumptions as they come up with and develop new tools, whether it be on the computer […] science side of it, […] but also have enough knowledge to unpack the theory and understand concepts like

significance […]. So really […] understanding regression and generalized linear models, so that they've got […] a feeling for what's going on under the hood […] and an ability to modify and adapt it." (University Academics)

"I think it's really easy to create a model based on something you found on the internet, run data through it, get a result that looks […] pretty, publish it and say 'aha!' you know, 'we've done data science'. But, obviously with that kind of mathematic coding and logic grounding you need to be able to understand what's going on and […] be able to be rigorous and justify what you're doing" (Students and ECPs)

"I think that kind of abstract thinking is probably really important. […] Tools go out of date very, very quickly. New technologies come along, but that way of structured, abstract thinking, I think, is what has helped me. Where I am now, confronted with a paper, I can still slowly read through it and get through it myself, because I have an understanding of the abstract thinking. Even if the technology that might implement it is changed drastically." (Students and ECPs)

"I would say, the mathematical sciences provides a clarity of thinking which is often lacking in many areas where people are working with data. And that clarity of thinking […] leads to better ideas of what the data might be saying (and) how you might approach it. And it also leads to a skepticism where a mathematician is trained to actually know what a proof is, and hence knows what an absence of the proof is." (Reference Group)

"[…] It is depending on whether we want data scientists who develop the methods further, which absolutely need the top mathematical and statistical theory skills, or the ones that there is shortage of in the industry. They don't need to be tip top mathematical statisticians or mathematicians, but they need to understand the context and everything that they can apply. They understand the problem translated in this statistical mathematical problem, and then do the analysis and interpret it." (Reference Group)

"Over the last 30 years or so of running my company, I've employed many fresh graduates, and the best predictor for how well they turn out is not what they've studied in the way of statistics or what they've studied in the way of computer science but how much mathematics they've done. It's as simple as that in my mind, and getting somebody in here without sufficient mathematics is a waste for me. So that's really what I look for. Ideally, of course, yes, I do want the statistics and the computer science as well. But if it comes without the mathematical background, then, they very quickly run out of steam." (Reference Group)

"I would just say that mathematics is essential to data science to impart rigour and to understand the foundations of the methods being used in order to obtain correct analyses and predictions." (Reference Group)

"… a lot of my circle is now people working as data scientists in some of the big tech companies, Google, Facebook or Meta, Amazon, and for those companies for a data science role the starter interview is often maths and stats. So it will be very stats 101, probability theory, and very kind of theoretical. Work it out with a pencil sort of questions. So, for those roles that sort of teaching seems really fundamental, and they expect you to have it." (Students and ECPs)

Soft skills and attributes

There was broad agreement that communication is a vital part of being a data scientist, both in order to work with others as well as to present results to stakeholders in clear and effective ways.

"… communication is one of the core skills, because […] people who […] train as data scientists rather than as statisticians are going to be looking to do applied work. And they're going to be looking at doing applied work in collaboration with someone else that is maybe not a data scientist, right? So the ability to communicate, to quickly learn, to quickly adapt to different areas of application and expertise is, I think, of great importance." (Reference group)

"[It is] very important to be able to communicate your results in a way that is understandable, and in the way that can be easily assessed by others because this collaborative approach is so vital to data science. Very vital, it's probably one of the key parts of it, as I think about it." (Reference Group)

In addition, other desirable attributes and skills relate to the rapidly evolving tools and technology, which requires data science professionals to be flexible, self-directed and able to learn and adapt to new systems quickly.

"I think […] a lot of preparation for what you're going to do on the ground is really, it will change moment to moment, and something like exactly what coding tools we're going to use, exactly which cloud platform we're going to use, what sort of industry terminology I need to get my head around to be able to clearly communicate with my stakeholder […], having a certain flexibility of approach is really the underlying skill there, because we can get too specialized. And because it is still such a broad discipline, actually having a bit of a background in how to get across a new topic really quickly was one of the most helpful things that I learned." (Students and ECPs)

"Another capability I would name is the ability to figure things out, look things up, and learn things fast." (University Academics)

One participant offered the opinion that the need for flexibility and adaptability might favour younger employees.

"And I think the growth in data science has really kind of pushed people towards being more adaptable and kind of constantly being on top of the technology. So I think maybe there's a little bit of a disconnect between what people want to do and what people have the capability to do. As an individual, it's kind of been advantageous, because people are like 'we need to do this' 'we need you to do this' and I guess the more contemporary recent graduates maybe have better skills or are better equipped to adapt to that. But, I think it does cause some tension at an organisation level, but it works out I guess in the favour of early career people." (Students and ECPs)

Importance of domain-specific knowledge

Data analysis does not occur in a vacuum but is deeply entrenched with many "real-life" fields and domains which require the data science professional to have a working understanding of. The real-life context of the data and the domain specific knowledge required elicited some comments that indicated that this is of central importance to data science.

"I work with epidemiologists, and they have a way of looking at the data and assessing the data which I learned a lot from. And so I put that under understanding the context of the data, its limitations and its strengths, and then using that to assess the answer you get from your model, or whatever you need to do. I think this is a very important skill. It's so easy to get a dataset and throw a model at it, it'll do something and get a result. You're not finished. You need to understand the context of the data." (Reference Group)

"I guess what I see as that discipline-focus, so you have got to have that strong background in a particular discipline, but then also develop those data science skills. […] So having just pure data science is a bit hard, because it's always related to something. It's not just 'I do data science'. That's great but it's often discipline-focused is what we find." (Industry and Government)

"I think, if there were to be a discipline of data science, an important ingredient in that would be coming to grips with the context and domain and motivating issues that, we get you to try and make sense of data. And I think that's a bit underdone in a strongly disciplinary model. You will learn about, the discipline of computer science, you learn about the discipline of statistics. Where you learn about synthesizing that knowledge and engaging with a problem domain is not so obvious, […] I've had feedback recently with people going oh, we had some data scientists but they were pretty useless, because you know that they couldn't get what we were trying to do." (University Academics)

What should data science degrees look like?

Both in the Industry and Government group and in the Reference Group general data science degrees received criticism for delivering graduates without enough mathematical and statistical background.

"But we do have challenges with […] the limited actual mathematical and statistical exposure that a lot of the people who come to us with straight data science degrees might have, and how that can impact on […] that data driven decisions role, and that type of data analytics role that we focus on." (Industry and Government)

"People who come out with data science degrees tend to be excellent programmers, and they know all the latest systems, and they go you know I can run this algorithm, or you know the first thing I do is I get rid of outliers. Why? Oh, because that's the first thing you do right? And you miss that kind of [...] digging into the data and getting a handle on it. So, I think it's become a very popular degree recently, but I'm not seeing high quality candidates because of it." (Industry and Government)

Participants in the Industry and Government focus group expressed some concerns whether (undergraduate) data science degrees were in fact at all useful, given that these degrees need to cover a lot of ground from diverse academic disciplines. Some suggested that it might be possible to add data science to existing degrees in science and engineering as an alternative.

"I'm just wondering if the market is demanding more skills than what a general […] undergraduate degree can offer. And if so, could either of these two solutions […] be the appropriate one? You know, the further training of existing data science degrees to include more maths, statistics, problem solving skills or something like that? Or should we then be offering data science modules as minors within existing degrees in science, in […] biology or stats or something like that? Would that […] bring people up to speed with the skills necessary to the market quicker?" (Industry and Government)

"There's a real industry issue at the moment of sorting out the good from the bad and the ugly. […] We need industry to come back […] and work with universities to ensure that the skills we want are actually being delivered in some way. So that could be […] through actually having data science not as an undergrad or a postgrad or anything else, but actually in a lot of other degrees, because it crosses. It doesn't just come through as a primary degree, it actually crosses across pretty much every school that you're having now." (Industry and Government)

In terms of what a general data science degree should contain, there is broad agreement that acquiring computer science and statistics and mathematics knowledge at an academic level should take up most space, with platform-specific technical knowledge desirable but not essential.

"You know they're (data science degrees – MW) about specific platforms and specific solutions, and I don't know that that level of integration is needed from a general degree, and they (Google, Amazon and other software platforms mentioned – MW) will offer strong certifications in their platforms anyway. So, in terms of what I'd be looking for when hiring someone, the ideal candidate, the best candidate, would probably be someone with an academic understanding of the maths and stats and computing techniques required, partnered with, in an ideal world, also some of those certifications to be efficient in a particular way of working." (Industry and Government)

"IT people are very good at understanding the structure of data. In fact, they are often superior to statisticians in that regard. But [when] it comes to the content and meaning of the data they are usually hopeless. And that is why data scientists need a significant amount of statistics. So my gut feeling is that what is sometimes taught as data science is often sort of a detuned version of computer science and with an absence of statistics. What we need is this combination of relevant *[word spoken with emphasis] computer science* and *[emphasis] the relevant parts of applied statistics."* (Reference Group)

The debate on the relative weighting of mathematics, statistics and computer science-related subject was nuanced, but with very broad agreement that the mathematical sciences contribution (at least statistics content) to the degree should be roughly the same as the computer science component.

"That pretty much matches [our] model. Where the actual […] data science degree is made up of half computer science, half maths, and there's room for another third in an external discipline, or majors in other areas. So it's not specified those other areas, but that 50-50 model". (University Academics)

"This is coming from an industry angle. Because the university degree is very short. So I'm going to put out a controversial point here. All the pure mathematics stuff, just keep it, go straight to statistics and then do computer science at at least 50% weighting. Of course I'm not saying mathematics is not important but with the three years that you have, I've got to choose something." (Reference Group)

"But you don't have to be in the top maths class to do data science at [my university] […]. At [my university] it was weighted a lot more towards stats and that probably made a lot of sense to me, I didn't go super hard into maths. If you think like a bit of a mathematician, you don't have to be particularly good at maths, but you think mathematically, and then you learn lots of stats, I think that sets you up quite well for a lot of different sorts of problems." (Students and ECPs)

University academics involved in offering data science courses indicate that data science degrees offer flexibility to meet the varying demands, with the content of the actual courses building on the strengths of the expertise that is in-house.

"I don't think there is one course or one set of content that's appropriate for everything. I think it depends on whom you're teaching. […] If you're teaching people who want to specialize in data science or machine learning or statistics, or whatever, what you're teaching them is going to be different to what you're going to be teaching users of the data who have a much more subject matter focus. Because it's such a huge area there's going to be bits that are more relevant to the different discipline areas, the content discipline areas. I personally think it's very difficult to have one concept of data science, so one course or one grouping that teaches exactly what's relevant. I think there's going to be very different courses for different cohorts of students that will be appropriate to them but not necessarily one course that is appropriate to everyone." (University Academics)

"Because if we could construct an ideal set of faculty for data science, it's unlikely that whatever we come up with is going to exist in any existing faculty. So part of how the data science courses are typically constructed at different universities is really a function of who's there at the time, and what their skills and background, and expertise and engagement with industry really are." (University Academics)

"Industry has a very broad definition. […] It encompasses a lot of different disciplines. And as universities we're not necessarily training our students to get a job, but to be able to have the sufficient skills to find a job. So if we're going to include AI and stats and all that, I'd also include optimization, because that's a huge component of what is actually done out there. […] We need flexible degree programs that allow students to have a kind of a fundamental core of what we think is appropriate for that core. And then have additional units that allow students to pursue their interests." (University Academics)

Part of the design of data science degrees involves connecting with the industry sector about their skills needs.

"The university sector's a bit slow. I mean, there are all these private operators running data science training courses. So how do we fit in with, you know, the data camps and the Courseras, and I think there's been some complaints from industry that [the] university is slow to change and give them the skills that industry in particular needs. So I think that's a conversation we need to have as well as how do we address that one as a sector, a university sector." (University Academics)

Internships and work-integrated learning

Students and early career professionals agreed on the usefulness of internships and work integrated learning, to learn some of the technical skills that are not always on offer within university courses and to gain experience within a "real-world" context.

"In running some internships for a government agency we found that because data science is still mostly vaguely defined, and the job descriptions are still very sort of abstract, students are expecting very very different things from their courses. So, they've enrolled in data science but they're really more interested in data engineering perhaps, or those sort of things, and they're, I wouldn't say they're dissatisfied, but they're looking elsewhere to further their skills rather than university courses." (Students and ECPs)

"I think it might be too much to expect a university degree to rely on all the technical skills that a data scientist would need. And so I think things like work integrated learning and placements perhaps need to play a bigger role in training the next generation of data scientists. Whether it be a university, or whatever tertiary education or institute it happens to be, but the recognition that if you're going to produce job-ready graduates, who they themselves aren't quite sure what skills they need, I think having them being embedded into organisations that will be using their skills to solve problems is probably going to be an important part." (Students and ECPs)

"There's nothing like working with real data. Obviously in your degrees you're doing projects and whatnot and you have datasets, and you are using real data in some aspects, but if you're actually at a company you're doing a real-world problem with their dataset. That's when you learn skills like data wrangling probably the most, and get the best out of it. And that's just hard to do in a degree." (Students and ECPs)

"I did want to flag, I also did the tech assessment recently where the data wrangling, I would not be able to do, based off my university experience." (Students and ECPs)

"I feel like data science and the cloud computing aspect of that does need to be paired with the security aspect, and maybe that's not commonly thought of in the same vein just yet." (Students and ECPs)

Who should deliver university data science degrees?

Despite the emergence of data science degrees at many universities, the degree delivery is often housed in different departments, schools and faculties. The university academic focus group discussed how best to bring together the different areas of expertise and the possibility of converging towards one new discipline of data science.

"I think one of the concerns with the University sector is [...] that the data science degrees are very segmented. We've got ones in business analytics. We've got one, a new master of public health in data science starting. We've got computer science running them. It's just coming from a number of different angles, and I think as a sector we need to be really clear about data science being a discipline in its own right." (University Academics)

The discussion among the academics included the topic of discipline expertise, the importance of subjects being delivered by discipline experts (this includes delivery of mathematics and statistics subjects by mathematical scientists) as well as the challenges of finding people with the right kind of expertise.

"When I had an introductory data science unit, I would struggle to get people with the appropriate technical skills to come and talk about even principles of visualization. So somebody not just who knew how to use R or Python or Tableau, but who had really thought deeply about what are some of the psychological aspects of how we visualize data, how we perceive data. So if we're actually going to do a bit of a dive into either ethics or things like visualization we do need to be partnering with others who actually have that expertise, and not just assume that the technical people already have that expertise." (University Academics)

"… in almost all our programs we have some sort of ethics component with engagement from humanities and law faculties, which serves […] a number of purposes, because it ticks off the University cross faculty KPIs, but it also has the discipline experts embedded in teaching. And that then gives us more credibility when we say, look, we should actually be teaching stats, and not the people from business school, and we're happy to do the stats in the business, master business analytics as well. […] So […] having the experts doing it works, and it increases in marketability a lot when you've got those various sort of ethics and […] philosophy and risk management […] in there." (University Academics)

The challenge for universities is to find the synergy between the interdisciplinary nature of data science while honouring the disciplines contributing to data science degrees.

"… there are notable areas of overlap. You know where you will have statisticians who've got a focus on machine learning in their research […], and you will have a computer scientist with a more mathematical focus. […]. So I would say it's about 80% discipline specific, but there are people who are being employed in computer science and mathematics schools who lean to the other side." (University Academics)

What I feel optimistic about in Australia, […] people recognize the differences in people's lived experience of coming to data science […]. And […] there's a willingness to […] respect, understand, and find out more about that, rather than retreat to a more defensive posture […] It's going to be a benefit to the mathematical sciences, a benefit to the computing sciences, and it's going to enrich people who […] might feel that they've really got a foot in both camps. They don't identify as a statistician or a computer scientist […] I have tremendous respect for those disciplines, and I want to be able to connect with and tap into that knowledge." (University Academics)

How do we satisfy the demand for data scientists?

The focus group participants cited anecdotal evidence that there is a great demand for data science professionals.

"… the market out there now for data scientists, or for cyber, or for anything in the tech field, is hot. So, it really doesn't matter what degree you have or what you put on your CV, you're probably going to get a job, and move around." (Industry and Government)

"I guess in WA […] most of the big mining companies suck up all the good talent. […] anecdotally, you know, if you're a small firm trying to find a data scientist, you can't compete with a large company that can offer huge sums of money." (Industry and Government)

"… the growth in data science is really affecting the public service's ability to attract talent that we need. With the huge amount of competition within the private sector, it's a big problem for us to try and understand how we can retain and attract people with the data science skills needed going forward." (Industry and Government)

"Even recruiting in the private sector is very difficult." (Industry and Government)

"I was talking to someone who has a start up in Sydney, who is offering graduate analysts $160,000 to $180,000 a year." (University Academics)

In terms of the level of study, Industry and Government representatives participating in the focus group preferred to employ Master's or even PhD level graduates - but this is probably not representative for all demand across employers.

"The level of people we tend to bring into the team, certainly in my team, its usually Master's, and my team's probably about a third PhDs and a third Master's, and it really is that statistical reasoning and also just being able to work with really big data." (Industry and Government)

"I'm looking for people who are PhDs in some subject, and who've actually thought about a problem deeply, can think their way through a problem, but have the right kind of computing skills and interest in data." (Industry and Government)

"So in our organization, we typically get people at the Honours level, occasionally less than that, and occasionally the PhD level, but Honours would be our most common level, or else, as an associated degree with, say, a Master's in data science." (Industry and Government)

"We're about 50 - 50 PhD and Master's, we don't really hire anybody lower than that." (Industry and Government)

There is broad agreement about the importance of making data science more visible, and motivating high school students and undergraduate university students to consider Data Science as a career through coordinated programs.

"… Queensland […] runs a gateway to schools program, where they bring industry involved to the schools. […] industry and government can fund this per state or actually at a national level, to ensure that we're getting more industry involvement in schools. And that's […] with the kids but also with the parents, so highlighting […] that parents are the decision makers generally on how kids […] go through schools. […] we need to, as a greater industry, get involved with the discussion, make rockstars essentially out of people in the industry now so that others want to follow." (Industry and Government)

"I think there's quite a lot to be done around the coordination of different initiatives […] We annually run a code camp at our offices in Melbourne, where we bring in […] a lot of local school, primarily girls and non-binary, students, sort of 14 to 16, and they spend a week […] learning about the tech industry and […] some of those broader skills that are possibly not so much in a traditional curriculum." (Industry and Government)

" I think there's probably a role there, maybe for universities, maybe for government, but it really needs to be […] partnerships and breaking down the silos to do that. And sort of also getting the incentives right, so a typical academic is not necessarily well incentivized to do sort of industry outreach or schools outreach. And typically industry, […] beyond like trying to attract the best talent, is not necessarily incentivized either. It's just sort of getting that mix right and getting the coordination to do things with impact, I think." (Industry and Government)

Some offered migration and focusing on mature age students as an alternative.

"The migration pathway is an interesting one. So, part of what government could do […] if we've got a skill shortage is to increase the priority list for data scientists coming through. Technology as a whole has been decreased on the priority list, so there is a way that the industry can come together and highlight the need for data scientists." (Industry and Government)

"There can be advantages in attracting people […] after they start working, and not necessarily straight after their degree, when they […] don't necessarily know what area they want to work in, and what the problems are. And if you can attract them back […] 5, 10 years later you can find that they are […] much more committed and finding really useful problems to work on that then help them further their career." (University Academics)

What are the merits of data science degree accreditation?

A number of participants pointed to the merits of accreditation and advocated for it for slightly different reasons.

"We're finding out from Open Days, […] mums and dads are attracted by courses that are accredited because they feel that their kids, you know that somehow, if a course is accredited by somebody or some bodies, that their kids will have a better chance of finding work. […] There's an actuarial science course that's widely regarded, and that is accredited, and it is currently the biggest competitor for data science and statistics, because there's a huge overlap." (University Academics)

"[…] The Statistical Society of Australia has an accredited statistician […] badge that they can give people that […] says this person has a profile in this area, and we kind of trust that they know what they're talking about. And because anyone can call themselves a data scientist at the moment, it might be helpful to have something like that. Also, that helps us have a goal to work towards, and a set of things that we can do to know that we're good data scientists as

well." (Students and ECPs)

"When it comes to accrediting courses, for example, I think the ACS has already gone down that path […] so we need to work with other societies as has happened in the UK. I think they're working with the mathematicians, the computer scientists and the RSS in the UK to build a kind of common understanding of data science courses. So it's less about particular methodology, but more about the kind of principles that we want to teach students." (University Academics)

"I think that it's imperative for us to provide our working definition, and what our skill sets are that we demand in our degrees and in our qualifications to meet the expectations of employers and to meet the expectations of other academics, for further downstream education of our students that exit from these degrees." (Reference Group)

What could be the role of AMSI and SSA?

Participants mentioned advocacy, career education to students, parents and teachers, creating a community of professionals, and being the contact point between industry and students as useful roles for AMSI and SSA.

Advocacy

"In my mind, the role of the SSA and AMSI is to provide a voice to the more mathematical and statistical proponents of data science. To make sure that what is being branded as data science across the University sector conforms with our understanding of data science, and is not entirely co-opted by computer scientists […] and to accreditate data science degrees in the same way that the SSA has attempted to accreditate the general statistics degrees. […] In summary, I think that the role is for us to have a unified voice as a community of mathematical practitioners within the greater data science landscape that's going to appear." (Reference Group)

"… actually to be, and this goes for AMSI as well, […] the body that also takes the discussions with media and the education role, with the communication across society, because media often asks for advice […] A bit like what the Royal Statistical Society is doing in the UK, but also the Swedish Statistical Society have tried to do, but in a much smaller extent, is to be out there and take the discussions publicly." (Industry and Government)

Career Education

"Could the society's role be to perhaps bridge the school, university and the industry sector together, […]in a data science fair, where you bring together schools, you have a bunch of different partners, and you're raising awareness for that particular career path and […] where it can lead for someone […], so […] it becomes ingrained in their mind what they could achieve in their lives through that." (Industry and Government)

"We need a layered approach. We can't directly influence the students, and because they are in contact with their teachers and parents all the time, just going to schools and talking to them once is not going to make a difference whatsoever. Instead, I think we need to look at the teachers' education and work with the Departments of Education and the Universities to educate the future teachers better. And also AMSI and the Statistical Society of Australia need to run professional development workshops for the teachers at the primary level, for all teachers at high school level. All teachers *[emphasis], not just mathematics teachers, but the science teachers and social sciences teachers, because data is everywhere. […] I also think AMSI and the Statistical Society of Australia should run workshops for high school counsellors who advise students which degree they should choose, because they are the ones who talk to kids, and when they are good at mathematics they say you could study actuarial studies, you could study finance. But they don't say you can study statistics, the don't say you can study data science, because they don't know what it is. So I think we should leave the kids to the influencers to be influenced, and we should influence the influencers."* (Reference Group)

Community of professionals

"Being part of a fraternity of professionals has been really important. […] you can be a lone data science professional, but really it is being part of a broader network and being able to have […] people with technical skills you can ask advice for and bounce off ideas. Often statisticians are the only person in their team […] I had to look outside of my studies and joining societies or going to networking events, and that's where I can actually be exposed to new technologies that may not be taught in my course. I needed a much broader professional education that probably wasn't provided in the degree." (Students and ECPs)

"Having that […] variety of presentations, that […] engage someone after a work day to […] head over or join a Zoom call, I think, is a great place to start. But really, the potential for the professional societies is to be a place where ideas can mix, and people can […] talk across those traditional silos." (Industry and Government)

Contact point between students and industry

"I think it might be too much to expect a university degree to rely on all the technical skills that a data scientist would need. And so I think things like work integrated learning and placements perhaps need to play a bigger role in training next generation of data scientists. […] The recognition that if you're going to produce […] job-ready graduates, who they themselves aren't quite sure what skills they need, I think having them being embedded into organisations that will be using their skills to solve problems is probably going to be an important part. And I think the SSA or AMSI might have a role in doing that […] because as a student how could I possibly apply to a dozen different organisations and know what they want? There's one place who can advocate for industry as much as […] advocate for students on both sides. I think that would be a great help." (Students and ECPs)

Appendix 8: Review Panel and Reference Group participants

Data Science Review Panel

Name	Affiliation
Distinguished Professor Kerrie Mengersen	Chair Queensland University of Technology
Associate Professor Jessica Kasza	SSA President (2020-22)
Professor Tim Marchant	AMSI Director
Dr Maaike Wienk	AMSI Policy and Advocacy Manager
Prof Ian Gordon SSA President	Incoming SSA President (2023-24)
Professor Inge Koch	University of Western Australia
Professor Lewis Mitchell	University of Adelaide
Professor Annette Dobson	University of Queensland
Dr Emi Tanaka	Australian National University
Dr Nicola Armstrong	Curtin University
Professor Scott Sisson	University of New South Wales

Data Science Review Reference Group

Name	Affiliation
Associate Professor Ayse Bilgin	Macquarie University
Ms Lisette Aarons	Australian Bureau of Statistics
Dr Hien Nguyen	University of Queensland
Dr Don Vicendese	University of Melbourne
Associate Professor Grace Joshy	Australian National University
Dr John Henstridge	Data Analysis Australia
Professor Marijka Batterham	University of Wollongong
Associate Professor Robert Niven	UNSW Canberra
Professor Sally Cripps	CSIRO

Glossary

ASC	Australian Skills Classification
ACS	Australian Computer Society
AMSI	Australian Mathematical Sciences Institute
SSA	Statistical Society of Australia
AI	Artificial Intelligence
STA	Science & Technology Australia
AAS	Australian Academy of Science
ADSN	Australian Data Science Network
ABS	Australian Bureau of Statistics
CSIRO	Commonwealth Science and Industrial Research Organisation
ICT	Information and Communication Technologies
ACS	Australian Computer Society

For further information
contact the Director at AMSI

enquiries@amsi.org.au

amsi.org.au/publications

AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

AMSI Members

Full Members

University of Adelaide
Australian National University
La Trobe University
University of Melbourne
Monash University
University of Newcastle
University of New South Wales
University of Queensland
Queensland University of Technology
RMIT University
University of Sydney
University of Western Australia

Associate Members

ATSIMA
ANZIAM
Australian Bureau of Statistics
Australian Institute of Physics (AIP)
Australian Mathematical Society (AustMS)
Australian Signals Directorate (ASD)
Computing Research and Education Association of Australasia (CORE)
Curtin University
Deakin University
Flinders University
Griffith University
James Cook University
Macquarie University
The Mathematical Association of Victoria (MAV)
Mathematics Education Research Group of Australia (MERGA)
Murdoch University
New Zealand Mathematical Society (NZMS)
New Zealand Statistical Association (NZSA)
Optiver
Reserve Bank of Australia
Statistical Society of Australia (SSA)
Swinburne University of Technology
University of New England
University of South Australia
University of Southern Queensland
University of Tasmania
University of Technology, Sydney
University of Wollongong
Victoria University
Western Sydney University

AMSI and its members acknowledge the significant contribution of the University of Melbourne as our Lead Agent and host

List of members as of November 2023