Data Science in Western Australia

What we do *Today* will pave the way for a better *Tomorrow*

NOVEMBER 2020

WA DATA SCIENCE INNOVATION HUB

KPMG

NERA NATIONAL ENERGY RESOURCES AUSTRALIA

METS IGNITED

Department of Jobs, Tourism, Science and Innovation

NEW INDUSTRIES WA

bankwest

FÆTHM

Hon. Dave Kelly MLA

Minister for Water; Forestry; Innovation and ICT; Science; Youth

> Western Australia is fortunate to have talented people and many competitive advantages. Together we can leverage the opportunity provided by data science to build the skills for the jobs of the future.
>
> Hon Dave Kelly MLA

Minister's Foreword

The McGowan Government's Recovery and Jobs Plan is helping to drive our economic and social recovery from the impacts of the global COVID-19 pandemic. The State Government is investing in jobs and building the skills needed for a diversified and resilient economy.

An important component of the McGowan Government's plan to diversify the WA economy is our support for the WA Data Science Innovation Hub, provided through the New Industries Fund. The hub works closely with organisations across WA to build data literacy and skills to ensure the WA workforce is ready to capitalise on the many opportunities that data science can bring.

As Minister for Science, and Innovation and ICT, I have seen first-hand how WA is building a competitive advantage in data science and analytics. Data science skills are already providing tremendous benefit to WA, whether predicting the spread of disease, distilling insights from the enormous quantities of space data coming from radio astronomy projects in the Mid West, or the ongoing technological advancements in our thriving mining industry.

This report reiterates the need to continue our investment in facilitating home-grown data science STEM skills. All stakeholders consulted for the report viewed skilled data scientists and a thriving data science ecosystem as fundamentally important. Many of these stakeholders, particularly those from industry, outlined that Western Australia will need more highly skilled data scientists in the future.

This need becomes greater as we progress further into the digital age. The State Government is committed to supporting the WA Data Science Innovation Hub, the development of STEM skills through our State STEM skills strategy and the broader data science ecosystem to create jobs in this increasingly important field.

Together, industry, academia and government can leverage these opportunities and build the skilled data science workforce required now, and for the jobs of the future.

WADSIH Foreword

Western Australia is uniquely positioned to capitalise on the opportunities that data science can bring to the State through its comparative advantage in key industries including resources, agriculture, space and health, amongst others.

The State Government, along with Curtin University, has supported the WA Data Science Innovation Hub (WADSIH) to provide a focal point for industry to work together to accelerate the uptake of data science in Western Australia.

WADSIH aims to ensure the State remains at the forefront of the digital revolution by increasing the uptake, education, training and awareness of data science in Western Australia.

The use of data to solve complex business and societal challenges is becoming a key advantage for many territories around the world, and Western Australia must ensure that it has the required skills and collaboration to take advantage of this – for the benefit of all Western Australians.

The vast arrays of data collected by governments and businesses provides an enormous opportunity for Western Australia to develop innovative solutions based on data to address challenges and seek opportunities that will impact us all.

Since the creation of WADSIH we've been working to enable the application of data science for the benefit of all Western Australians.

In this report we seek to show the potential that this has begun to unlock and how we can all work together to ensure Western Australia continues to flourish through data science.

Dr Liz Dallimore
Director

WA Data Science Innovation Hub

> "The use of data to solve complex business and societal challenges is becoming a key advantage for many territories around the world."
>
> Dr Liz Dallimore

Data Science in Western Australia

Currently there are **32,034 FTEs** in the data science family, over the next 10 years this family will experience:

Capacity gain of **6,300 FTE** through augmentation

2,783 FTE job additions to cater to the technology types

Data science ecosystem of today

Substantial variability in the size, maturity and location of data science teams

Challenges of structuring and positioning data science teams persist

Graduates from Western Australia are highly sought after

Data science ecosystem of tomorrow

93%

14/15 organisations ranked data science as very important to the future of their organisation

5 years

Organisations stated that their data science teams will likely **double** over the next five years

Organisations stated a **high demand** for citizen data scientists to grow capability

WADSIH

The WA Data Science Innovation Hub (WADSIH) is a Western Australian Government initiative, and supported by Curtin University, which aims to ensure the State remains at the forefront of the digital revolution by increasing the uptake, education, training and awareness of data science in Western Australia. The mission of WADSIH is to:

Enable Western Australia to build jobs by developing a data driven ecosystem and culture

Foster collaboration, promote expertise, advocate and enable data literacy

Work across industry, academia and government

WADSIH's mission is achieved by:

Access to specialist data science capabilities in universities and other research organisations

Generating a supply of trained graduates and provision of upskilling programs, and creating jobs

The translation of data science capabilities from mature sectors to emerging sectors

Advocating to industry, academia and government

Facilitating collaboration and building the data science community

This report helps WADSIH achieve their mission and remain at the forefront of data science in Western Australia by:

Establishing a baseline of the current Western Australia data science ecosystem

Discerning future trends expected to impact the data science ecosystem over the next five to ten years

Providing recommendations for industry, academia and government to continue building out this ecosystem

WADSIH is supported by the Government of Western Australia and Curtin University, and a number of Foundation Partners:

Department of Jobs, Tourism, Science and Innovation · New Industries WA · Curtin University · bankwest · Mets Ignited · NERA National Energy Resources Australia

> The WA Data Science Innovation Hub was established in 2018 through the support of the Government's New Industries Fund and Curtin University.

Contents

Executive summary 1

1 Scope and approach 5

1.1 Scope 6

1.2 Approach 6

1.3 Stakeholders interviewed 6

1.4 Classifications, source data and methods 7

1.5 COVID-19 8

1.6 Structure of this report 8

2 Western Australian Data Science ecosystem of today 9

2.1 Overview of the data science ecosystem 10

2.2 West Australian data scientists 11

3 Western Australian Data Science ecosystem of tomorrow 19

3.1 Data science and the ecosystem of tomorrow 20

3.2 The impact of technology 23

4 Next steps for the Western Australian Data Science ecosystem 27

Appendix

5 Notes 31

Co-Branded Disclaimer

> Data science turns complexity in large datasets into actionable hypotheses and solutions.
>
> Stakeholder from the Health Care Industry

Executive summary

Executive summary

The Western Australian Data Science Innovation Hub (WADSIH) engaged KPMG and Faethm to establish a baseline of the Western Australia data science ecosystem and future trends expected to impact it over the next five to ten years. This report provides a range of insights to both support WADSIH's role and inform the Western Australian industry, academia and government about the value of data science today and into the future. These insights were provided by 15 organisations across eight industries that WADSIH considers to be industry leaders in data science in Western Australia. This was supplemented with the alignment of occupation related classifications and analysis of a range of detailed and publicly available data to build a range of projections within the Faethm platform.

Western Australian Data Science ecosystem of today

Across the 15 organisations interviewed, there was broad agreement that data scientists were a relatively new role within their organisations. Stakeholders generally described the role of the data scientist as one that looks at problems through the patterns in data and applies mathematical constructs to the data to generate knowledge. Stakeholders often added that a data scientist was professionally competent across four domains - statistics and/or mathematics, data analysis, communication and subject matter knowledge.

The occupation of "data scientist" is not formally listed within the Australian New Zealand Standard Classification Occupation (ANZSCO). This is not surprising – new occupations take time to be recognised and added to ANZSCO. However the role of a data scientist is generally recognised as a skilled profession and one that is likely to be backed by a tertiary qualification.

Using ABS Census data related to a person's occupation and job advertisements from an array of websites, Faethm estimated that there are 32,034 FTE "data science family[1]" roles within Western Australia in 2020. The majority of this workforce is male (63 per cent), the age band with the largest number of data scientists are aged 34 years and under (37 per cent) and the Professional, Scientific and Technical Services industry employs the largest number of FTE (31 per cent). Over 90 per cent of the data science workforce are estimated to be in Perth and surrounds with the remainder across the regions.

Almost all of the organisations interviewed stated that data science was very important, extremely important or critical to their organisation. In recognising this importance, organisations noted that they face a range of opportunities and barriers in their organisation and industry. These opportunities related to data science being a growth industry, the potential to save money, better target investment and improve the sharing of data across their organisations. Barriers related to access to infrastructure, not releasing more data, improving data linkage access and access to talent.

Gender

FEMALES: 11,800 FTE

MALES: 20,234 FTE

Age range

< 34 and below: 11,952 FTE

= 35-44: 10,234 FTE

> 45 and above: 9,849 FTE

Industry (top 3 ANZSIC)

Prof., Sci. & Tech Services: 9,917 FTE

Public Admin. and Safety: 3,326 FTE

Fin. & Insurance Services: 3,170 FTE

1 Feathm has defined that the "data science family" is comprised of 116 roles across a range of industries classified as being similar to data scientist based on a scoring system that compared abilities, interests, knowledge, skills, work activities, work context, work styles and work values.

Western Australian Data Science ecosystem of tomorrow

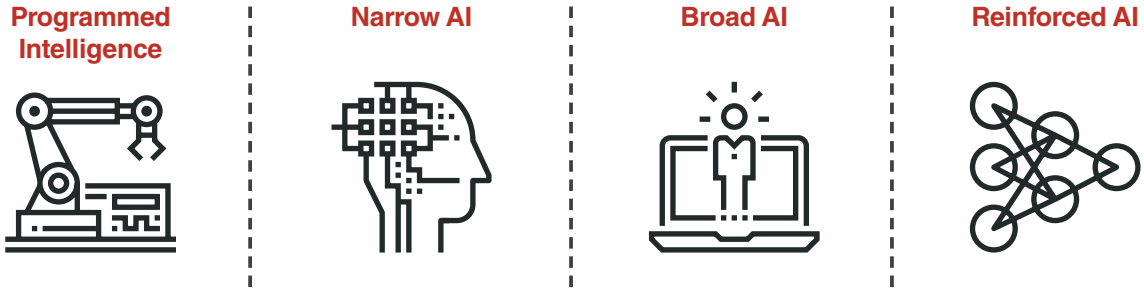
As we progress deeper into the digital age, and into the data science ecosystem of tomorrow, Western Australia will need more data scientists – both in quantity and quality. Work across all industries is becoming more data-driven, affecting both the skills and jobs required. In recognition of this, organisations stated that data science was very important, essential or critical to the future of their organisation. With one organisation expressing that data science would become '*more and more important as our executives continue to buy-in*'.

All organisations envisioned their data science team growing, with some doubling or tripling the size of their team within the next five years. One organisation envisioned a future where '*every team will have a data scientist in it*'. To meet this demand, the data science ecosystem of tomorrow will require data scientists, citizen data scientists and all talent embodying data science literacy.

Over the next ten years, several technology types will impact almost all occupations, including that of a data scientist. These impacts may result in the automation of occupations (i.e. the loss of jobs), augmentation of occupations (i.e. increasing the capacity) and add occupations (i.e. increase in jobs). Data scientists are not immune from some of these technology types automating their current roles. Although forecasts generated by Faethm have found that many data scientist roles will be "augmented" by these technologies. These technologies are broadly classified as programmed intelligence, narrow AI, broad AI and reinforced AI.

Augmentation of an occupation is defined as the capability of technologies to supplement the efficiency of a job and in doing so, enabling a worker to gain capacity to do higher value work. It is estimated that a total of 42 per cent of roles in the data science family will be augmented by these technologies over the next ten years. Augmenting a role contributes to capacity gain i.e. a person in the data scientist occupation is able to do more and hence contributing to an increase in capacity gain. The Faethm platform predicts that this capacity gain amounts to a total gain of 6,300 FTEs within Western Australia over the next ten years.

Figure 1. Technology types that will impact the data scientist role into the future



Next steps for the Western Australian Data Science ecosystem

As we progress deeper into the digital age, and into the data science ecosystem of tomorrow, Western Australia will need more data scientists in both quantity and quality.

Given this, a range of recommendations for industry, academia and government are made surrounding building out this ecosystem. WADSIH has begun to and will continue to play a key role in delivering these next steps, through increasing the uptake, education, training and awareness of data science.

“WADSIH has a role in influencing policy, higher education and in the definition of the data science language.”

Stakeholder from the Electricity, Gas and Waste Water Industry

Scope and approach

1. Scope and approach

1.1 Scope

The Western Australian Data Science Innovation Hub (WADSIH) engaged KPMG and Faethm to establish a baseline of the Western Australia data science ecosystem and future trends expected to impact it over the next five to ten years. This report supports WADSIH's role, which aims to ensure Western Australia remains at the forefront of the digital revolution by increasing the uptake, education, training and awareness of data science. In addition, the insights provided in this report inform industry, academia and government about the value of data science today and into the future.

1.2 Approach

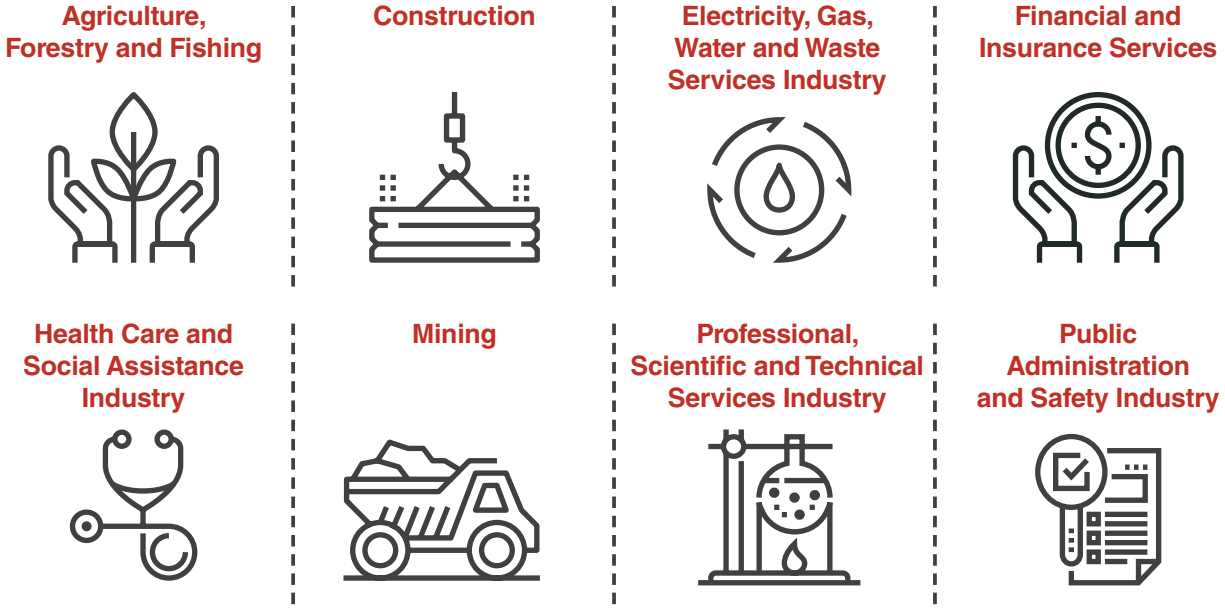
The approach to this engagement involved the following stages:

- **Stage one:** Development of an interview guide and identification of key Western Australian based organisations from across eight industries. In this stage, KPMG, Faethm and WADSIH identified and discussed key datasets, classifications (ANZSCO and O*Net) and statistical methods. Key datasets identified included the use of Australian Bureau of Statistics (ABS) Census occupation data and data available from online job sites (i.e. data scientist advertisements) to enable Faethm's platform to generate insights on current and future workforce size and distribution across Western Australia.
- **Stage two:** Stakeholder interviews with 15 organisations focused on the current and future state of data science within their organisation. A series of questions focused on the current state of the data science ecosystem worked to gauge how organisations currently define data science and the role of data scientists, including competencies of data scientists currently employed. Questions focused on the future state involved asking organisations the data science workforce upskilling requirements and how important they see data science to the future of their organisation. The roles of the individuals within these organisations ranged from Chief Data Officer to Data Scientist. These individuals had a sound basis on which to comment on the data science ecosystem within their own organisation and more broadly within Western Australia.
- **Stage three:** The analysis of the qualitative interview material and quantitative data. The qualitative component involved the synthesis, thematic coding and analysis of the information gathered through the interviews with organisations. The analysis of the quantitative data involved the collation, mapping and analysis of the key data sources. This data was then loaded into Faethm's platform. This qualitative and quantitative analysis is the basis for the insights into the Western Australia data science ecosystem today and over the next ten years.

1.3 Stakeholders interviewed

A series of semi-structured interviews were undertaken with 15 organisations from a range of industries which WADSIH and KPMG considered to be key industry stakeholders in the data science ecosystem. These organisations were asked questions regarding the current and future state of data science within their organisation. Due to privacy purposes, these organisations are not named throughout the report. However, their input has been aggregated under their representative industry (e.g., Health Care and Social Assistance Industry). While a broad scope of public, private and government organisations across industries were sought, limitations around availability constrained participation. Industries to this report include:

Figure 2. Western Australia data science industries to the report



1.4 Classifications, source data and methods

Two occupation classifications have underpinned the work presented in this report. They are the Australian Bureau of Statistics occupation classification, Australian and New Zealand Standard Classification of Occupations (ANZSCO) and O*Net. ANZSCO is a skill-based classification of occupations, developed as the national standard for organising occupation-related information for purposes such as policy development and review, human resource management, and labour market and social research. ANZSCO includes all jobs in the Australian workforce.

The other occupation is O*Net Online. O*NET OnLine has detailed descriptions of the world of work for use by job seekers, workforce development and HR professionals, students and researchers. Faethm mapped these two classifications for the occupations most closely related to "data scientist." This occurred between the O*Net occupation and the ANZSCO four-digit level.

Along with the occupation classifications, two key data sources underpin the report. They are the ABS Population Census occupation counts (2011 and 2016) and job advertisements (2017 to mid-2019). The figures in this report provide estimates of the number of "data scientists" as of 2020 in Western Australia. These numbers were derived from the ABS Census 2011 and 2016 figures and utilising a growth rate to forecast figures for 2020.

While job advertisement information was solicited by 'scraping' detailed data for "data science" roles from various online job boards (e.g. Seek and Indeed). Advertisements were analysed and information extracted using natural language processing techniques and removing references to the same job advertisement. Within this, information on skills listed within the advertisement was also extracted. Clustering algorithms were then deployed to arrive and classify data points within a specific data science ontology. For example, a job advertised as "AI Research Scientist" was a high match to data scientist.

1.5 COVID-19

The timeline of this project was interrupted by the unprecedented event of COVID-19. Stages one and two involved gathering the qualitative and quantitative data to develop insights around the Western Australia data science ecosystem. The majority of this work took place prior to the full and ongoing impact of COVID-19. For example, while the first stakeholder interview took place in December 2019, the final stakeholder interview occurred in mid-March 2020. Stage three of the project, which involved the qualitative and quantitative analysis took place during and after COVID-19.

Qualitative and quantitative insights were gathered and developed prior to the full and ongoing impact of COVID-19. KPMG were unable to re-interview stakeholders and ask whether they still anticipate their data science teams growing, or whether COVID-19 had directly impacted the size and geographic distribution of their data science teams. Similarly, the growth rates and impacts of technology that underpin the Faethm analysis do not account for the impact of COVID-19. The reader of this report should keep this in mind when interpreting the results that follow.

Whilst COVID-19 has impacted organisations in many and varied ways, there has generally been an acceleration of the digital transformation of many organisations. While digital transformation strategies are designed by and for the organisation, they essentially involve embedding technology to allow employees to solve traditional business challenges through the use of technology. These digital transformations often result in exponential growth in the volume, variety and velocity of data available to provide insights to these organisational challenges. Data scientists are uniquely positioned to generate insights and help solve old and new challenges of these digitally transformed and data-rich organisations.

1.6 Structure of this report

This report is structured as follows:

- **Scope and approach** (*this section*) provides an overview of the approach to the engagement, and the approach taken by KPMG and Faethm to develop the findings outlined in this report.
- **West Australian Data Science ecosystem of *today*** outlines the current state of the data science ecosystem in Western Australia today based on the views of stakeholders and the insights from Faethm.
- **West Australian Data Science ecosystem of *tomorrow*** details the future state of the ecosystem over the next five to ten years based on the views of stakeholders and the insights from Faethm.
- **Next steps** provides a series of recommendations for consideration by government, universities and businesses.

Western Australian Data Science ecosystem of today

> The most valuable algorithms in the business are the simplest.
>
> Stakeholder from the Mining Industry

2. Western Australian Data Science ecosystem of today

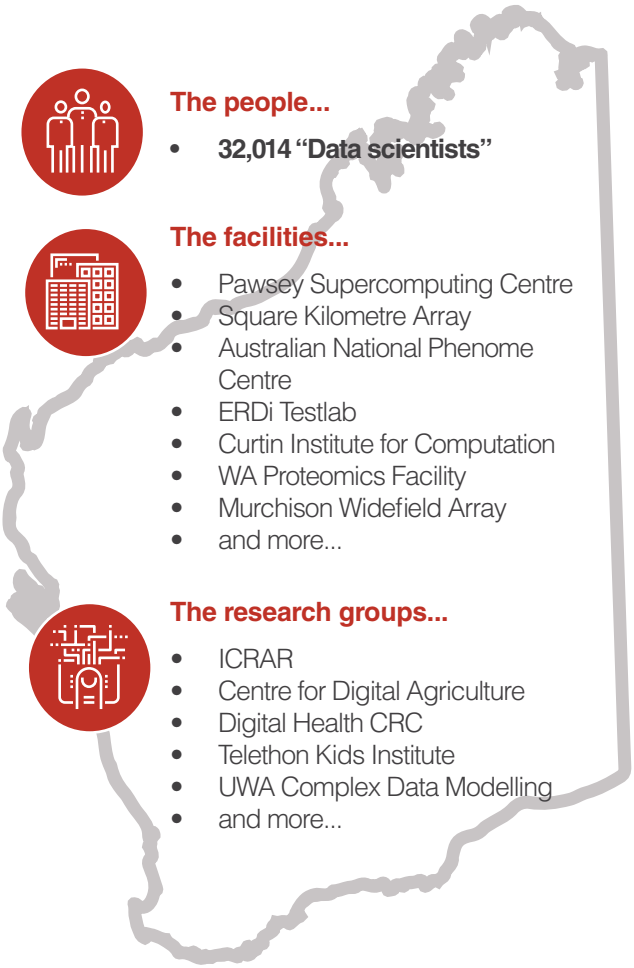
2.1 Overview of the data science ecosystem

More data is being generated than ever before and a thriving data science ecosystem is essential in navigating the complexity and opportunity presented. At its most simple, the data science ecosystem is comprised of data, people (i.e. data scientists and research groups) and physical infrastructure (i.e. facilities) that are connected through the shared use of data science.

Within Western Australia the data science ecosystem infrastructure is characterised by data-intensive research groups, facilities that support and enhance Western Australia data science projects and community groups who work with both individuals and organisations to enhance understanding and connections within the data science community. Many of these research groups and facilities were discussed by stakeholders throughout conversations.

> Two of the three fastest supercomputers are located within Western Australia.
>
> Stakeholder from the Health Care Industry

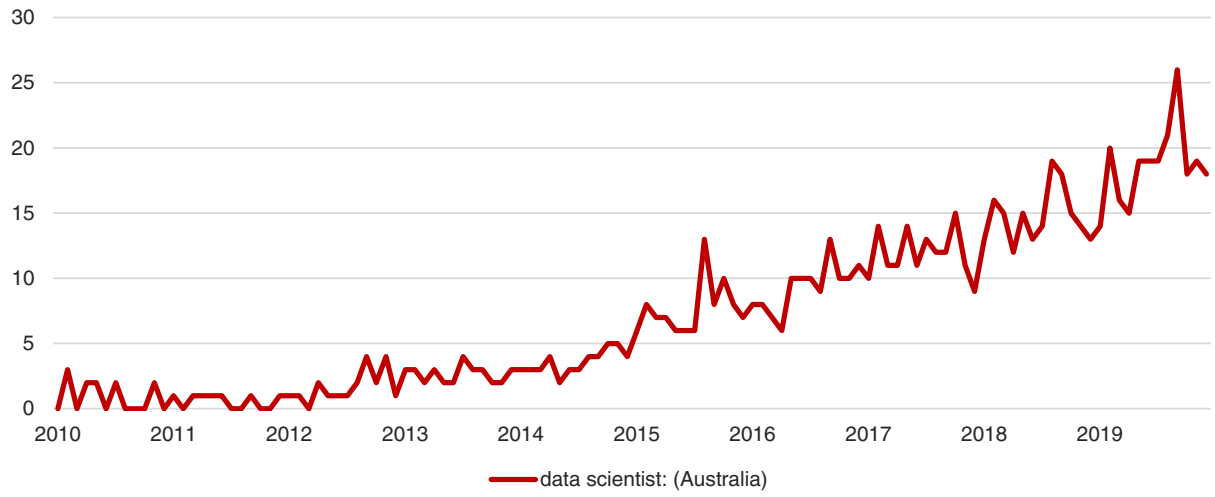
Figure 3. A high level snapshot of the people, research facilities and groups that comprise the Western Australia data science ecosystem



2.1.1 Classifying data scientists

The occupation of data scientist is not listed within the Australian Bureau of Statistics occupation classification (Australian and New Zealand Standard Classification of Occupations aka ANZSCO). This is not surprising, when even a quick google trend analysis shows that the phrase did not start showing much search interest until late 2015 (See Figure 4). Despite this there has been some criticism of ANZSCO not being up-to-date and not recognising quickly growing professions such as data scientists.[2] However the ABS and the Department of Immigration recently advised (late 2019) that the new and emerging occupation of Data Scientist should use the code 22499 - *Information and Organisation Professional Not Elsewhere Classified (NEC)*.[3] The Occupations within this ANZSCO unit group of *Other information and organisation professionals* are likely to have a level of skill commensurate with a bachelor degree or higher qualification.[4] Persons within this occupation are likely to have at least five years of relevant experience i.e. the role of data scientist is recognised, somewhat indirectly, as a skilled profession and one that is generally backed by a tertiary qualification.

Figure 4. Google search term interest in the phrase "data scientist"



Source: Google trends. Data accessed March 2020.

2.2 Western Australian data scientists

Faethm is an analytics platform that enables companies and governments to understand the impact of emerging technologies on the future of work. Faethm provides an estimate of 32,034 FTE "data scientist" like occupations within Western Australia[5]. The methods used to derive this breakdown are outlined in *Classifications, source data and methods* under the scope and approach section within this report. The breakdown by gender, age group and top three industries is outlined in Figure 5. The majority of the workforce is male (63 per cent), the age band with the largest number of data scientists are aged 34 years and under (37 per cent) and the Professional, Scientific and Technical Services industry has the largest number of FTE (31 per cent). These baseline figures for 2020 provide a solid basis for understanding the future size and distribution of the data science workforce in Western Australia. The Faethm platform also provides predictions of which data science roles and industries are likely to be impacted by technologies into the future.

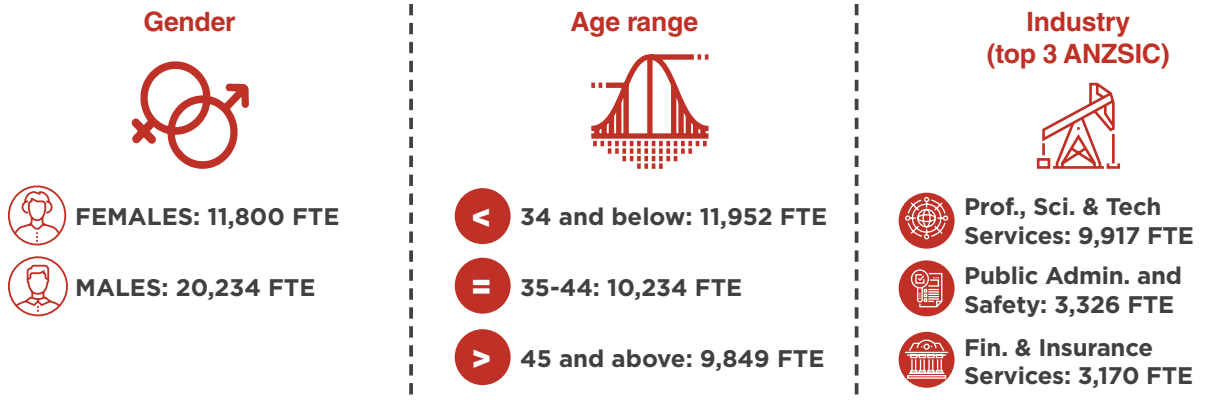
2 https://www.abc.net.au/news/2019-08-17/anzsco-occupation-lists-need-updating/11413518, viewed March 2020.

3 https://immi.homeaffairs.gov.au/what-we-do/skilled-migration-program/skilled-visa-newsletters/september-2019, viewed March 2020.

4 https://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/E8131A0C7C678C4FCA2584A8000E78C2?opendocument, viewed March 2020.

5 See Appendix A for a detailed description of Faethm.

Figure 5. Western Australia data science roles by gender, age range and industry



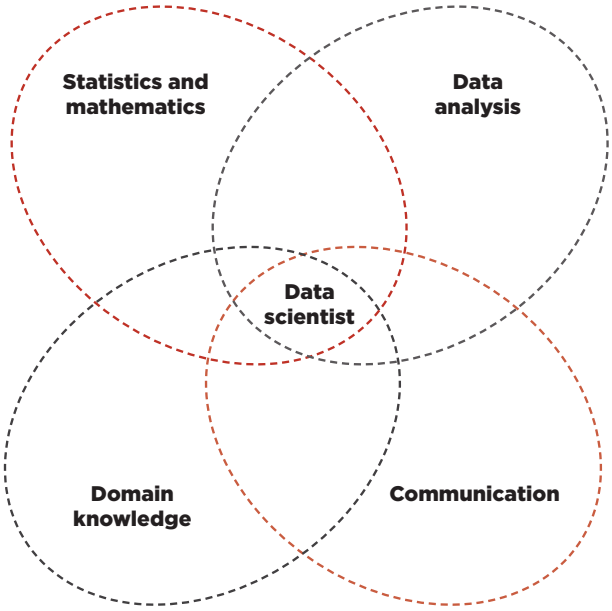
Job advertisements were "scraped" from a range of websites (e.g. SEEK) and job advertisement classification was performed with a BERT based text classifier. The list of ten descriptions from the Faethm ontology that made up the "data science specialisation" list included, Actuaries, Data Scientists, Data Science Managers, Machine Learning Engineers, Analytics Managers, AI Research Scientists, Bioinformatics Scientists, Biostatisticians, Epidemiologists and Statisticians. The job advertisements for the first six months of 2019 identified **215 distinct** "data science specialisation" like jobs. This was nine per cent higher than the same period in 2018.

> Not enough data scientists.
> And they aren't cheap.
>
> Stakeholder from the Professional, Scientific and Technical Services Industry

2.2.1 Organisational view of data scientists

Across the 15 organisations interviewed, there was generally broad agreement of data science and data scientists as a relatively new role within their organisations. Stakeholders described the role of data scientists as one that looks at problems through the patterns in data and applies mathematical constructs to the data to generate knowledge. Through looking at the data, they can find solutions buried within and transform this into a transparent solution to solve complex problems, discover insights and generate value for the organisation.

Organisations generally discussed a data scientist needing to be competent across four domains. Those domains were Statistics and/or mathematics; Data analysis; Communication; and Domain (or subject matter) knowledge. Organisations would then discuss that the data scientist sits at the intersection of these domains, although two to three organisations emphasised that without the statistical and mathematical skills the individual could not readily call themselves a data scientist.



A data scientist looks at problems through the patterns in data and applies mathematical constructs to the data to generate knowledge. Through looking at the data, they find solutions buried within and transform this into a transparent solution to solve complex problems, discover insights and generate value for the organisation.

Figure 6. Domain skills of a data scientist

Similarly stakeholders recognised that some particular industry sectors (e.g. healthcare) require the data scientist to acquire a depth of domain knowledge. Within healthcare there are an extensive list of classifications (e.g. ICD-AM) that do not remain static and that having machine learning skills does not necessarily mean you can readily switch to become a data scientist within the healthcare industry.

The value of a data scientist being able to clearly communicate the results was emphasised by several stakeholders. This included a range of skills from data visualisation, report writing and the communication aspect of transforming the data into a transparent solution to people within the organisation.

> Data science requires skills from computer science, data analytics and metadata, machine learning – a broad - brush of skills.
>
> Stakeholder from the Professional, Scientific and Technical Services Industry

2.2.2 Number, maturity and location of data scientists

Across the 15 organisations interviewed the size, maturity and location of the data science teams varied substantially. For some organisations the number of data scientists could be counted on one hand. For other organisations they had both a data science "Centre of Excellence" and data scientists embedded within smaller teams throughout the organisation.

Within one organisation they recognised the value of data science within their business but had not yet established a "data science" team. On a scale of one to five they viewed the maturity of their data science team at a one. Other organisations rated the maturity of their data science team at a four or five out of five. These organisations also tended to have more sizeable data science teams. For example one organisation had a team of approximately 15 data scientists, which were complemented by a range of technical analysts. These technical analysts included spatial scientists, data governance roles and computer scientists.

> For data science teams to be successful within an organisation, they need to be structured correctly within the organisation. Often data science teams fail as they're placed within the technology side of the business - too far away from day-to-day business operations.
>
> Stakeholder from the Health Care Industry

The geographic location of data scientists varied from organisation to organisation. One organisation noted that their data scientists are globally distributed and they "hire" support functions from particular locations. Others noted that their data scientists are a team of three that are spread throughout teams within their office in Perth. Interestingly while data may have been generated within regional areas of Western Australia (e.g. the Square Kilometre Array, mine sites, regional health services, digital farms etc.) the data scientists tended to be located either within Perth or other Australian capital cities.

2.2.3 Organisational structure of data science teams

Organisations spoke to the challenges of structuring and positioning their data science teams. In general, organisations tended to adopt either a centralised or decentralised structure for their data scientists. Under a centralised reporting structure organisations had data scientists reporting to a head, or equivalent of, data science within their organisation. For many organisations beginning their data science journey, a centralised operating model appeared to be the preferred structure due to the accessibility and technical collaboration opportunities. It also allowed the data science team to

'establish a profile' within the organisation.

> The operating model is really important for data science. Specificity is key.
>
> Stakeholder from the Financial Services Industry

For those organisations that had a more decentralised structure to their data science teams they tended to be further progressed in their data science journey. That is to say they could embed a data scientist into a particular team within their organisation. That team would know how to work with the data scientist and utilise their skills and insights. Similarly the data scientist knew when to "reach back" into the organisation's central team for support and guidance.

One such organisation highlighted the importance of not viewing centralisation or decentralisation as the only two available operating models. This was particularly the case as organisations progressed in data science maturity. Several opportunities for a decentralised reporting structure were noted, the largest being the proximity of the data scientists to the business unit and use case. Highlighting the importance of data scientists being '*in the field close to the problem*'.

> In a high risk world internally, it's difficult to chart new ground as you have no experience to base it on.
>
> Stakeholder from the Financial Services Industry

Getting this organisational structure right would be a complex challenge for government. One organisation pointed out the complexity associated with where a data science specialist team might "centralise" within government. This was due to both the breadth of data that government generates but also the domain expertise required, for example, in healthcare. Another organisation discussed whether data science was a whole of government, central agency or departmental issue. They further noted that even at the Departmental level there are challenges for the data science team structure. This stakeholder spoke of the specific example of health data and whether the data scientist worked at a hospital, health service provider or Department of Health level.

2.2.4 Focus areas of data science teams

The focus areas of data science teams were, somewhat unsurprisingly, dependent on the organisation's industry. Stakeholders suggested a broad array of industry specific focus areas, technologies and mathematical techniques. Underpinning the specific focus areas or use cases were key points around there needing to be clear business cases for specific data science projects i.e. the data scientists cannot be left alone building technical models that no one understands or that cannot be "monetised." Stakeholders would also balance this point by stating that data scientists were not expected to be subject matter experts. They did however need to be able to work closely with subject matter experts and quickly understand key concepts.

Within public administration and safety stakeholders discussed the benefits and focus areas related to data linkage. Specifically discussions related to distributed data linkage models and the value of the Social Investment Data Resource (SIDR)[6] to understand the government services provided to vulnerable families. Data scientists were also focused on using these linked datasets to assist in the forecasting of the demand for and long term cost of government services.

> Focus areas? Surveys. Epidemiology. Bayesian adaptive clinical trials. Genomics. Rare diseases. Data linkage.
>
> Stakeholder from the Health Care Industry

6 https://www.wa.gov.au/organisation/department-of-treasury/case-study-social-investment-data-resource, viewed April 2020.

Data scientists working in electricity, gas, water and waste services were more focussed on asset optimisation in terms of operations for anomaly detection, assets and usage. Similarly within mining conversation often turned to prediction and optimisation but across all areas of the business e.g. from trucks carrying ore to human resources. While data scientists within the Healthcare and social assistance industry were focussed on using analytics to predict demand for health care services. One stakeholder spoke of the application of Bayesian techniques to support clinical trials.

Figure 7. Western Australia data science focus area snapshot by stakeholder industry

Agriculture, Forestry and Fishing

- Digital soil and crop mapping in combination with disease detection and pest management

Mining

- Prediction and optimisation across many business areas from HR to heavy machinery

Construction

- Construction simulation, design issue prediction and warranty analysis

2.2.5 Importance of data science to organisations

Almost all of the organisations interviewed stated that data science was very important, extremely important or critical to their organisation. There was an outlier that identified that data science was not yet very important to the organisation but that it soon would be. For this particular organisation they required executive "buy-in" to data science and for the organisation to progress their digital transformation before being able to state that data science was very important to them.

> Very important. Actually, extremely important.
>
> Stakeholder from the Mining Industry

2.2.6 Opportunities and barriers for data science

Organisations identified a range of opportunities and barriers for data science within their own organisations and more broadly for their industries within Western Australia. These opportunities related to data science being a growth industry, the potential to save money, better target investment and improve the sharing of data across their organisations. Barriers related to access to infrastructure, not releasing more data, improving data linkage access and access to talent.

> We still have the ability to make a large improvement. Saving millions of dollars, increasing up-time a visible effect.
>
> Stakeholder from the Mining Industry

All stakeholders perceived data science as a substantial opportunity for Western Australia, with one stakeholder simply describing the opportunity of data science as "*huge*." Also highlighted was the importance of government investment in continuously developing and capitalising off this substantial opportunity. One stakeholder iterated that as data science is an important growth industry, there is a simple argument for further government investment.

The importance of improving data linkage to drive organisational value, especially across scientific and government services was noted. One stakeholder pointed out that while Western Australia was once a leader in this area, it could no longer be considered a leader.

Within their own organisations stakeholders saw potential for further savings and that as the executive within their organisation saw the value of data science it was becoming increasingly important.

> To put bureaucracy in place is hard and to still have the freedom to innovate is difficult.
>
> Stakeholder from the Construction Industry

When it came to barriers to data science within their organisation stakeholders identified a range of issues. These barriers included:

- A lack of executive awareness of the value of data science
- Private sector investment in particular available infrastructure
- Bureaucracy
- The need for greater coordination across sectors and
- Sourcing data science talent

In relation to data science talent one stakeholder identified that there are "*not enough of them.*" Other stakeholders provided more nuanced feedback. In particular that there was an opportunity for data scientists to move beyond building great theoretical models. Data scientists need to demonstrate some commercial acumen as a theoretical model won't save the business money or identify where money should be invested. A particular stakeholder provided an example of an extremely sophisticated model integrating a variety of data sources providing outputs that were ultimately very similar to the existing simple model.

The "*not enough of them*" comment was echoed by several stakeholders. These stakeholders discussed that when searching for data scientists that their talent search was not restricted to candidates within Western Australia or Australia. That the search for the right talent was often global. This was due to the fact that the required data science talent needed to have that combination of statistics and mathematics, domain knowledge, communication and analytical skills. Finding this combination within Australia was often just not possible.

> A coordinated approach with a digital drive behind infrastructure delivery.
>
> Stakeholder from the Construction Industry

Stakeholders pointed out that a raft of data is generated behind infrastructure projects but that this information is not readily available e.g. technical plans and drawings for new roads, assets, tunnels, buildings etc. One stakeholder suggested that this data, where it is not commercially sensitive, should be released upon such projects being completed.

> ... because of under-investment in data linkage infrastructure and unrealistically risk-averse regulatory environment…. It's too exhausting to do this data linkage.
>
> Stakeholder from the Health Care Industry

2.2.7 Training, sourcing and development of data scientists

Organisations outlined that they sourced their data scientists through a range of methods. This included organically (i.e. internal role having a data science focus), through networks and third parties. Often a wide range of applicants with varying levels of competencies and backgrounds applied to data science roles. Given this, some organisations used challenge statements and activities to demonstrate competency.

Organisations were most likely to hire specialist data scientists internationally and through a third party. Similarly, when sourcing senior data scientists organisations reported a lack of talent within Australia. Finding this international talent required a third party.

> We run people through certified analytics training from the US and support this internally.
>
> Stakeholder from the Public Administration Industry

Some organisations leveraged their internship programs for finding and tracking talent throughout their university careers, with one such organisation tracking talent from Australia, the United Kingdom and the United States. While many organisations hired external talent, a significant number relied on organic growth and up-skilling internal talent. In up-skilling potential and current talent, organisations reported a variety of approaches. These included a reliance on communities of practice (as listed on WADSIH), open source platforms (e.g. Coursera) and in-house development programs comprised of data focused modules. International and online training was often cited due to their affordability and accessibility.

Organisations with mature data science teams mentioned they used internal programs to develop the skills of their data scientists. Such programs included job rotation to provide more organisational domain knowledge, internal and external training. However, some organisations' training programs appeared to be more conceptually focused, looking at what you can do with data science and data more generally. With scope to evolve to a more formalised curriculum into the future. Notably, an organisation of self-assessed lower digital maturity, attempted to build a data science environment by proposing a data literacy program to their executive team. While this attempt was unsuccessful, they noted it was an essential first step in furthering a data focussed innovation culture.

> There's a lot of self-learners who don't want to go back to university, it is more skill than degree driven.
>
> Stakeholder from the Construction Industry

In terms of the quality of graduates from Western Australia, and Australia more generally, its graduates were reported to be highly sought-after. However, organisations mentioned that retaining this talent locally was problematic as a large amount move overseas to work for technology companies. Organisations reported following individuals through their university career, only to lose them to technology companies oversees at the end of their university career. Often this was due to the reputation of the industry itself or the location of work.

> Every team will have a data scientist in it.
>
> Stakeholder from the Mining Industry

Western Australian Data Science ecosystem of tomorrow

3. Western Australian Data Science ecosystem of tomorrow

This section discusses some key findings based on conversations with stakeholders and the analysis of a range of data sources by Faethm. Organisations reflected on the importance of data science to organisations, the anticipated growth in the number of data scientists and specific up-skilling requirements of data scientists into the future. Analysis by Faethm outlines the impact of automation, the augmentation and addition to the data science family of occupations.

3.1 Data science and the ecosystem of tomorrow

3.1.1 Importance of data science to the future of organisations

Almost all of the organisations interviewed stated that data science was very important to the future of their organisation. With one organisation expressing that data science would become '*more and more important as our executives continue to buy-in*'.

> How important is data science to the future of our organisation? Extremely important.
>
> Stakeholder from the Professional, Scientific and Technical Services Industry

There was one outlier who reported that data science was of medium importance to the future of their organisation. Primarily as data science '*was not a core function, but helps [the organisation] be better at what they do*'.

> Very important, health care has a deep need for more data scientists.
>
> Stakeholder from the Health Care Industry

3.1.1 Organisational structure of data science teams into the future

Organisations envisioned that over the next five to ten years, the structure of data science teams would likely ebb and flow between being centralised (mainly through a Centre for Excellence) and being distributed. With one organisation stating that while some '*parts will stay centralised; others might become more geographically decentralised*'. The opportunities provided by a centralised structure included networking, adopting new technologies and understanding issues that impacted the business. Some organisations spoke of the opportunities for decentralisation as the maturity of their data science capability increases alongside the volume, variety, velocity and veracity of their data. Notably, one organisation imagined that '*users of data and data visualisation tools into the future being as frequent and well accessed as the current users of PowerPoint*'.

3.1.2 Growth in data science teams in the future

All organisations envisioned their data science teams growing, with some doubling or tripling the size of their data science teams. One organisation emphasised the importance of a data-focused innovation culture to the longevity of their organisation, stating '*There would have to be growth, or we won't be around*'.

> Looking to increase in size, there's recognition of the value it can provide.
>
> Stakeholder from the Mining Industry

One organisation envisioned a future where '*every team will have a data scientist in it*'. This future involved a view of data science as a continuum with data scientist roles differentiated and embedded within assets and functions. In a practical sense, this translated into a future where employees take on a complementary role to data scientists. Building the base of these employees, or "citizen data scientists", would be targeted through communities of practice. Other organisations also stated the importance of building up data scientist roles and using specialist data scientists to uplift capability more broadly.

> Increase, through data science as a direct role and through using citizen data scientists by increasing data literacy across the company.
>
> Stakeholder from the Electricity, Gas and Waste Water Industry

Another organisation stressed the importance of building internal capability while testing the talent market. This organisation highlighted the need for more data scientists within their organisation but noted the competitive advantage of emerging countries and specialist firms. In acknowledging this, the organisation was exploring alternative models for accessing talent. While alternative models were being explored, they emphasised the importance of having a solid internal capability to understand and question their own capabilities as they were only in the early stages of understanding what data can do for their organisation.

> Yes, but this growth depends on the use case and exploratory data science can help with this. We don't see a drastic increase. As not every use case has a need for deep data science expertise. We're investing in this capability which will mean a couple more heads and a lot more training..
>
> Stakeholder from the Health Care Industry

Data science opportunities in government also plan to increase; however, one organisation stressed the importance of a government data strategy to enable this. Another organisation emphasised that as access to data within government grows, and the sharing of data between agencies continues to grow so will the opportunities for data scientists.

3.1.3 Up-skilling requirements of the future

There is no doubt that the recent explosion in data, connectivity and computing power has sparked a huge interest in data science among industry, academia and government. To take full advantage of this explosion a growing need of data scientists, citizen data scientists and employees with some data science skills is required. In response to this identified requirement industry, academia and government each need to respond individually and collaboratively.

In Western Australia, government policy makers are already responding to trends around how the world of work is changing and proactively preparing the State and its workers for the employment ecosystem of tomorrow. Contributing to this proactive preparation is a range of future-focused policy architecture such as the *Future jobs, future skills – Driving STEM skills* in Western Australia Strategy. The Strategy embodies an ambitious framework for building Western Australia's STEM skills to address the needs of the future workforce and capitalise on the creation of job opportunities. The execution of this Strategy, and partnership

with industry, to achieve an additional 150,000 jobs in Western Australia by 2023-24 as outlined in *Our Priorities: Sharing Prosperity* cannot be underestimated.

> Data science thread through courses, for example, commerce and within that thread.
>
> Stakeholder from the Mining Industry

From the interviews it was clear industry are leaders, understanding the importance of data and integrating data scientists into their existing workforce. However, up-skilling concerns were noted regarding who, how and what to specifically upskill in. For example, one organisation mentioned the challenge between choosing to upskill their data scientists with SME skills or choosing to upskill SMEs with data science skills. In this instance, the organisation was most likely to use open source platforms such as LinkedIn Learning and Coursera, but was unlikely to use a local university course as '*they are too much of a slow burn*'.

One organisation, who had a number of entry-level data scientists, planned to focus less on the technical and analytical side and more on domain knowledge and communication skills, highlighting that most of their data scientists are straight from university, and bring technical skills such as coding and building models. Hence, their plan to focus on domain knowledge and communication skills such as problem-solving, understanding their organisation and influencing solutions.

> Less on the technical and analytical side, more on domain knowledge and communication skills.
>
> Stakeholder from the Finance Industry

Many of those considering a career in data science, or simply wanting to upskill, use courses and programs facilitated through academia. Courses are available at the vocational, undergraduate, postgraduate and executive level. Some courses may be focused on computer science, applied mathematics, or analytics but have a data science thread throughout. One organisation, spoke to the importance of more courses such as the common Bachelor of Commerce having a data science thread throughout in meeting the needs of the data science ecosystem of tomorrow, where every team has talent with data science skills.

Over the last few years in Western Australia, data science specific undergraduate and postgraduate degrees have been created. These degrees are multidisciplinary and combine studies in computing, statistics, and mathematics and internet technologies. One organisation spoke to how academia instils graduates with strong technical skills like coding and building models but cited concern that '*more of this is going to be automated in the future*'. Highlighting the need to continuously update curriculums to ensure their currency due to the uncertainty of the data science ecosystem of tomorrow. Another organisation mentioned that many of their data scientists '*can explain the model, but not how to use it*'. They suggested that commercial applicability could be a gap within the curriculum.

Through WADSIH's engagement with industry, it is clear that broader data literacy and capability building across existing employees will be critical for the advancement of data science within organisations. WADSIH is working closely with industry, universities and training providers to ensure Western Australian organisations have access to opportunities to develop in house capabilities.

3.2 The impact of technology

A range of technologies will impact the data science ecosystem of tomorrow. To understand this impact, analysis has been undertaken on the data science family of roles in Western Australia. Given this, the analysis is only relevant to Western Australia and does not extend to roles defined outside of the data science family taxonomy. The data science family was curated based on the methodology outlined in *Classifications, source data and methods* under scope and approach in this report. All analysis on automation, augmentation and addition of data science jobs is in reference to the data science family unless otherwise stated.

Figure 10. Data science family and specialisation definitions

Data Science Family

- The "data science family" is comprised of 116 roles across a range of industries classified as being similar to data scientist based on a scoring system that compared abilities, interests, knowledge, skills, work activities, work context, work styles and work values.

Data Science Specialisation

- A "data science specialisation" is comprised of a subset of ten roles which are deemed to be either directly transferable to, or a specialisation of "Data Scientist".

Source: Faethm

Data science roles will be impacted by a range of technology types over the next ten years. This technological advancement is dependent on a range of factors such as the extent of human input and AI maturity. To this end, some technologies will impact the data science ecosystem of tomorrow more than others. All analysis on automation, augmentation and addition of data science jobs is in reference to all of the below technology types, and projections are dependent on their implementation within certain time parameters. A list of technology types is provided adjacent.

Figure 11. Technology types definition

Programmed Intelligence

Acts entirely on human input. These technologies can perform highly structured and simple process tasks by employing rules-based logic, processes, instructions and simple robotics.

For example, programmed code with no memory that employs rules-based logic.

Technology types:

- Process automation
- Fixed robotics
- Mobile robotics

Narrow AI

Acts semi-autonomously when prompted by humans. These technologies perform structured, familiar tasks in defined domains by using machine learning to interpret certain problems.

For example, reactive tools with limited memory that employ machine learning.

Technology types:

- Predictive analysis
- Recognition vision
- Voice response
- Suggestion provision

Broad AI

Acts to self-initiate actions with no human input. These technologies perform unstructured tasks and engage with their environment using perception and sensory processing of external input data.

For example, proactive systems with increased memory that employ deep learning.

Technology types:

- Sensory perception
- Decision generation
- Conversation exchange
- Dexterous robotics

Reinforced AI

Acts independently from experience to perceive and complete new tasks. Performing creative, unfamiliar actions across domains by using reinforced learning.

For example, self-improving agents (deep memory) using reinforced learning.

Technology types:

- Navigation robotics
- Collaborative robotics
- Solution discovery
- Generative design

Source: Faethm

Over the next ten years, a number of technology types will work to automate jobs or large components of jobs, causing the need to redefine jobs and/or re-deploy people, whilst ensuring they have the skills needed to pivot into new roles. The main impact of automation to the data science ecosystem of tomorrow over the next ten years will be experienced through process automation which is from the programmed intelligence technology horizon. More specifically, process automation focuses on code programmed to complete pre-defined, logical and rule based processing tasks. It is estimated that 33 per cent of FTE roles in the data science family in Western Australia will be automated. This is followed closely by predictive analysis from the narrow AI technology horizon, which predicts 31 per cent of roles will be automated.

3.2.1 Automation, augmentation and additions to the data science family

Figure 12. Automation, augmentation and addition definitions

AUTOMATION

Job reduction

- The capacity of technologies to entirely replace a job or large components of a job, causing the need to redefine jobs and/or re-deploy people.

AUGMENTATION

Capacity gain

- The capability of technologies to supplement efficiency of a job and in doing so, enable a worker to gain capacity to do high value work.

ADDITION

Job creation

- The addition of more existing jobs – or entirely new jobs – to your workforce to implement, run, maintain and govern emerging technologies.

Source: Faethm

Historically, automation and augmentation have touched every industry and occupation. Given this, a binary assessment of the impact of technological change associated with advancement in technology, reduced to jobs lost, added and reduced misses the point and contributes to bleak forecasts and fear-based discussion on the impact of technology on work. As the augmentation of existing jobs and the requirement of people in those jobs to attain new skills to apply these new technologies will be one of the biggest changes experienced in the data science ecosystem of tomorrow.

3.2.2 Gender impact across the data science family

Currently, in Western Australia, there are 32,034 total FTEs in the data science family. Like many other scientific and technical fields, the data science family is predominately male-dominated. Currently, males comprise 63 per cent (20,234 FTE) and females 37 per cent (11,800 FTE) of current data science family roles in Western Australia. A variety of future-focused initiatives are already being undertaken by industry, academia and government to increase female participation in STEM. To this end, Pillar 4 Diversity in STEM within the *Future Jobs, Future Skills Strategy*[7] focuses on enhancing women and female students' participation in STEM education and careers.

3.2.3 Job impact and location of the data science family

The majority of the 32,034 FTEs in the data science family is located in Perth (31,126 FTE) while the remainder is spread across the regions (909 FTE). Of the data scientists spread across the regions, the top three regions with the greatest number of data science family roles are the Pilbara (262 FTE), South West (218 FTE) and the Goldfields-Esperance (167 FTE) region.

Given the geographical spread of FTEs, data science family roles in Perth will experience the largest number of FTE roles automated and augmented over the next ten years. Overall a larger number of roles will experience augmentation than automation, meaning there

7 Ibid.

will be a requirement of talent in existing jobs to attain new skills to apply new technologies to meet demand.

The 909 FTEs in the data science family located in the regions can expect to experience a higher overall percentage of automation and augmentation than their Perth counterparts. For automation, data science family roles in the Pilbara (11 per cent) and Goldfields-Esperance region (11 per cent) will experience the highest overall percentage of automation. Half of the FTEs in the data science family currently will be impacted by automation and augmentation in some way. A substantial need to continue upskilling to meet demand is required by industry, academia and government, and the individual to meet demand.

3.2.4 Disruption to the data science family

All industries and occupations within the data science family will be disrupted through automation and augmentation in some way. It is estimated that a total of 42 per cent of the total workforce will be impacted by augmenting technologies over the next ten years and hence contributing to an increase in capacity gain. The aforementioned capacity gain amounts to a total gain of 6,300 FTEs.

The opportunities brought by automating technologies over the next ten years, could amount to a reduction in total salary cost and FTEs. An estimated 9 per cent of the total data science family workforce could be automated. However, this is dependent on the implementation of technologies throughout the next ten years.

Some occupations and industries within the data science family are at higher risk of disruption and change due to automation and augmentation. For example, over the next ten years the industry within the data science family with the highest number of FTE at risk to automation is the professional, scientific and technical service with 905 FTEs (9 per cent) of this industry's total workforce within automatable roles.

Simultaneously, the professional, scientific and technical service will gain the most additional capacity through augmentation. A total of 4,500 roles are expected to be augmented within this industry for a total capacity gain of 2,100 realisable FTE task time.

It is important to note, however, that impact is highly specific to each industry, and occupation across the data science family as well as the ability of various stakeholders within to manage change. For example, there is a large discrepancy between how workers in different jobs experience automation. For some workplaces, automation could lead to consolidation of jobs, while in others, it will alter the way talent use their time within their existing roles.

3.2.5 Additional jobs to the data science family

The data science family of occupations is expected to grow over the next ten years. This is due both to the increase in data and the implementation and operation of the aforementioned technology types. Projections estimate that 2,783 FTE are required within Western Australia to cater for this over the next ten years.

It is important to note that this figure is only in reference to jobs created through the introduction, implementation and maintenance of the technology types. The figure is not cumulative; hence it does not take into account jobs lost and jobs created outside of the introduction and implementation and maintenance of the technology types.

Following on from this, the technology horizon that will see the largest number of job additions is Narrow AI with 1,295 FTEs. Of these, 1,083 FTE are directly attributed to the predictive analytics technology type. The predictive analytics technology type focuses on tools that reactively use machine learning to conduct narrow analysis and make related predictions. And, is expected to come to fruition over the next five years.

Indeed, as others have noted, the need for data literacy skills in roles outside of the data science family will be considerable, however the availability of data did now allow for these projections.

> People with data science skills are not easy to find – there aren't enough of them.
>
> Stakeholder from the Professional, Scientific and Technical Service Industry

Next steps for the Western Australian Data Science ecosystem

4. Next steps for the Western Australian Data Science ecosystem

As we progress deeper into the digital age, and into the data science ecosystem of tomorrow, Western Australia will need more data scientists with deep technical skills as well more citizen data scientists. The establishment of the WA Data Science Innovation Hub through the State Government's New Industries Fund and Curtin University is a positive step in addressing how Western Australia addresses this issue. In addition, the following recommendations have been devised from findings from the stakeholder interviews and Faethm projections for building WA's data science ecosystem of tomorrow.

A summary of the recommendations for industry, academia and government are below in addition to a brief description of what is currently being done to address these issues presently.

	Actions taken to date
Industry	
Recommendation: look to "build out" local data science capabilities using both home grown and international talent.	• A number of organisations have technical upskilling programs. WADSIH offers a free program for startups and SMEs • Developing a freely available data literacy program to develop citizen data scientists • Accelerating the uptake of data science within organisations through advisory services. WADSIH provides a free service to small to medium companies to help them navigate how to establish data science within their business • Webinars with national and international experts • Resources catalogue of WA data science related product and service providers provided on the WADSIH website
Recommendation: adopt a data science capability framework.	• WADSIH is working with industry to develop a data science capability framework
Academia	
Recommendation: encourage a basic understanding and competence of data science skills.	• Universities, in collaboration with WADSIH, work with industry to bridge the gap in competencies required • Promote industry/academic collaboration. WADSIH works with existing initiatives such as Innovation Central Perth to promote this collaboration • Various internships and industry student projects • Sponsoring the UWA Data Science Club's Data Science Industry Insights night

	Actions taken to date
Academia (continued)	
Recommendation: evolve a range of educational pathways, through continuing and maintaining the flexibility to share courses and materials.	• Listing education, training and career opportunities on the WADSIH webpage
Recommendation: WADSIH should continue to act as a 'go-between' for industry and academia.	• Meeting with industry and government regularly • Establishing an internship portal • Evolving the competency framework model required by industry
Government	
Recommendation: refresh the Open Data Policy.	• Data science case studies that highlight recently released data
Recommendation: that government continue to build innovative partnership models with academic institutions and industry.	• Existence of WADSIH as a result of the New Industries Fund and Curtin University • Foundation partners of WADSIH such as NERA, METS Ignited and Bankwest • Partnership between WADSIH and the Pawsey Supercomputing Centre for the Australian Space Data Analysis Facility
Recommendation: WADSIH should continue to act as a "go-between" by working across government to ensure knowledge from industry is leveraged.	• Meeting with industry and government regularly • Providing input into government consultations and working with ministers to realise the value of data in government • WADSIH providing industry insights to government to share knowledge and learnings

4.1 Industry

Many organisations spoke of the challenge of finding the right data science talent. Further, these organisations would often seek out such talent internationally to work in Western Australia. The data science talent pool within Western Australia will likely never be deep enough to cover all industries and technologies. However, industries should look to "build out" local data science capabilities from these international talents.

Recommendation: industry organisations that bring talented data scientists to Western Australia need to ensure that these data scientists build out local capabilities.

To ensure that industry and organisations are able and ready to embrace changes, proactive planning for the data science ecosystem of tomorrow is already underway and should continue. To ensure consistent and defined language around roles, WADSIH is developing a data science capability framework in collaboration with Industry. The framework will recognise the granularity of expertise within the data science family, and what it takes to advance from one level to another. To this end, there is scope for industry to continue to work with WADSIH to identify data science talent, and for WADSIH to provide a platform in which knowledge can be exchanged.

Recommendation: industry organisations adopt a data science capability framework. WADSIH is currently working with Industry to develop a relevant framework to assist both employers and employees.

4.2 Academia

Currently, a 'data scientist' is described as someone who 'looks at problems through the patterns in data and applies mathematical constructs to the data to generate knowledge.' However, the data science ecosystem of tomorrow will be dependent not only on experienced data scientists but also "citizen data scientists," i.e. almost all occupations will likely require some base level data science skills. Many organisations envisioned growth in their core data science team but also in their "citizen data scientist" capabilities. Given this, in academia, many more students would benefit from awareness and competence in data science skills.

Recommendation: to prepare talent for this new digital data driven era Western Australia based academic institutions should encourage a basic understanding and competence of data science skills across many more educational pathways.

Courses in data science work cross-disciplinary drawing from existing courses in computer science, statistics, mathematics, statistics and information technology are already in place. Graduates from these courses work across all industries in a variety of roles. Organisations spoke of seeking out online courses for their data scientists rather than courses offered by local Western Australia based academic institutions.

Recommendation: as most data science courses are in their infancy with Western Australia based academic institutions should be prepared to evolve a range of educational pathways, through continuing and maintaining the flexibility to share courses and materials across programs, faculties and delivery modes.

In recognition of the evolving educational needs of data science talent WADSIH has continued to extend itself outside of traditional delivery modes of education. One example is through running data literacy pilots, another is through delivering upskilling training in both technical data science and broader data literacy with the initial focus of this upskilling training on SME businesses, with an extension to larger organisations.

Recommendation: WADSIH should continue to act as a 'go-between' for industry and academia, ensuring the universities are delivering relevant modes of learning.

4.3 Government

As the Open Data Policy points out, data is a major resource that can help government make data-led decisions and develop innovative solutions to wicked policy problems. A variety of initiatives are well underway within Western Australia and nationally to improve access to data. This includes the data.wa.gov.au website and its national equivalent data.gov.au. Within Western Australia there have been new data linkage initiatives such as SIDR, which sit comfortably alongside the Open Data Policy.

However, the Open Data Policy was released in 2015. At a similar time when the phrase "data scientist" was beginning to trend in google searches. It may be time to refresh this policy and ensure the exponential growth in data available to Western Australia businesses, academia and the community. This may require setting a target number of agencies or datasets beyond what is currently available on wa.data.gov.au. Many stakeholders pointed out that there is still a wealth of data held by Western Australia Government agencies that could be released to be used by industry and the community more broadly.

The discussion paper and consultation process on privacy and responsible information sharing is a core first step in the refreshment and modernisation required to build the data science ecosystem of tomorrow

> The Western Australian public sector collects and uses a vast array of data in the course of its everyday operations. This data is an important strategic asset and, when managed well, is a source of significant value to the State.
>
> Whole of Government Open Data Policy[8]

Recommendation: refresh the Open Data Policy to support growth in the data available to further deliver value and benefits to all West Australians.

Generating insights from an abundance of data will require a strong talent data science base. Sourcing talent to fulfil data science roles into the future will require continued and new collaborations across government, industry and academia.

Recommendation: that government continue to build innovative partnership models with academic institutions and industry, which seek to develop the data science workforce of the future. The development of WADSIH as a Hub is an example of this.

In building the data science ecosystem of tomorrow, a coordinated response to the collaboration of industry and government is essential.

Recommendation: WADSIH should continue to act as a "go-between" by working across government to ensure knowledge from industry is leveraged.

8 Department of Premier and Cabinet (DPC) 2015. Whole of Government Open Data Policy. DPC, Perth. Western Australia.

> Data science brings a sizeable opportunity to add value.
>
> Stakeholder from the Construction Industry

Appendix: Notes

5. Appendix: Notes

5.1 WADSIH

WADSIH is at the forefront of data science in Western Australia. Their mission is to enable Western Australia to build jobs by developing a data driven ecosystem and culture through fostering collaboration, promoting expertise, advocating and enabling data literacy across the community, industry, academia and government.

WADSIH executes this mission through:

- Facilitating access to specialist Data Science capabilities in universities and other research organisations
- Generating a supply of trained graduates and provision of up skilling programs, and creating jobs for Western Australia
- The translation of Data Science capabilities from the mature sectors to emerging sectors
- Advocating the importance of Data Science to government, industry and the community
- Facilitating collaboration and building the data science community in Western Australia

5.2 KPMG Management Consulting

KPMG's management consulting practice houses a series of professional advisers assisting both public and private sector organisations to understand their most important value drivers. Through working shoulder to shoulder with these organisations KPMG's professional advisers help organisations achieve tangible and lasting improvements in performance.

5.3 Faethm

Faethm is a globally unique SaaS Analytics Platform that enables companies and governments to understand the impact of emerging technologies on the future of work. Faethm distributes both directly to Government, Education, and Enterprise client globally as well as via its partner ecosystem consisting of consulting and technology firms.

Faethm's AI SaaS platform shows the future impact of emerging technology on any economy, industry, company or job.

Contact details

Dr Liz Dallimore
WADSIH Director
liz.dallimore@curtin.edu.au
Mob: +61 8 9266 2795
Tel: +61 400 019 969

James Arnott
KPMG Project Partner
jamesarnott@kpmg.com.au
Mob: +61 418 281 752
Tel: +61 8 9263 7109

Kim Hawthorne
KPMG Project Manager
khawthorne@kpmg.com.au
Mob: +61 422 072 108
Tel: +61 8 9263 7224

kpmg.com.au