



A **research-enabling platform** that encompasses the following important elements to accelerate research performance and dissemination of research results:

- Data management
- Underpinning capacity
- Staff and student capability
- Communities of practice
- Data Science/Research careers

The platform involves systems, structures, processes, support, capabilities and culture that allow advanced technologies to serve research outcomes. It reduces risk associated with data storage and management, and enhances compliance with funding body requirements, enables increased size and scope of research datasets, improves data collection, analysis and results, enables collaboration and enhances our ability to conduct world-leading research.

## INTRODUCTION

---

This strategy progresses the eResearch Strategy Executive Paper (December 2015) and the subsequent eResearch Discussion Paper circulated for comment and feedback in the first quarter of 2016.

The strategy proposes a *Data Science<sup>1</sup> and eResearch Platform* to provide Macquarie with a foundation of enabling data science and eResearch expertise, systems, policies, technologies and support to accelerate world-leading research.

Following comprehensive consultation externally and across the University, it is recommended the University focus on building five key data science and eResearch elements to further *Our University: A Framing of Futures Priority 2 – an accelerating and impactful performance in discovery*. These five elements are:

1. Best practice in **data management** to support data sharing, re-use and collaboration; to ensure Macquarie is compliant with funding body requirements; and to harness the latent opportunities for discovery in areas such as medical health data research.
2. Build **underpinning capacity** through leadership, systems, processes, policy and support structures that enable researchers to ‘focus on what they alone can do’<sup>2</sup>.
3. Grow **staff and student capability** by developing a strong data science learning culture, contributing to Macquarie University as a destination of choice for researchers and higher degree research candidates.
4. Support a vibrant and active **community of practitioners** to enable opportunities for cross-disciplinary research that lead to novel and diverse solutions to problems.
5. Investigate alternative **researcher career pathways** and data science research careers, attraction and retention strategies.

To do so requires the collaborative work and support of key stakeholders across existing portfolios:

- Pro Vice-Chancellor, Research Performance and Innovation (RP&I)
- Faculties
- University Library
- Information Technology
- Research Office
- Dean HDR

and proposes a new **Data Science and eResearch Team** and a **Data Science and eResearch Advisory Group** (consisting of representatives from the groups above) to guide and drive implementation.

---

<sup>1</sup> “At its core, data science involves using automated methods to analyze massive amounts of data and to extract knowledge from them. With such automated methods turning up everywhere from genomics to high-energy physics, data science is helping to create new branches of science, and influencing areas of social science and the humanities” (Data Science at NYU <http://datascience.nyu.edu/what-is-data-science/>)

<sup>2</sup> Our University: A Framing of Futures

## BACKGROUND

---

The rate at which research data is being generated and processed has increased markedly in the past decade. Larger amounts of data, faster processing times and innovative means of collection, analysis and presentation provide enormous opportunities for Universities, researchers and end-users. They also provide a challenge to maintain a state of readiness in the systems, structures, and support that allows realisation of these opportunities.

Macquarie University has some underpinning capacity to meet this challenge. The University has existing strengths in its researchers, library data management expertise, proposed information technology solutions<sup>3</sup> and a growing community of practitioners, some of whom have made impressive advances in this area. This strategy provides direction for the University to develop its research support capabilities further and to harness existing capabilities.

It provides a framework for agility, responsiveness and forward thinking in our research practices and support systems. It is part of a larger piece of work, articulated through *Our University: A Framing of Futures; World-leading Research; World-Changing Impact Strategic Research Framework*, and the draft *Research Commercialisation, Innovation and Impact Framework*, striving for pervasive best practice in research and research support. Data Science and eResearch are enablers of all Key Research Objectives of the Strategic Research Framework and crucial for our burgeoning research engagement, innovation, commercialisation and impact imperatives. Fulfilment of this strategy will have significant impact on research performance, the quality and readiness of our HDR candidates, our capacity to collaborate and to deliver high-impact research.

## VISION AND AIMS

---

Our vision is for Macquarie University to embrace, embed and drive Data Science and eResearch practice within the research lifecycle as an enabling platform for accelerated research performance.

To achieve this vision we will focus our attention on:

- **data management** as a cornerstone to Data Science and eResearch
- the University's **capacity** to provide **enabling** systems, structures, policies, processes and support, increasing the ease by which researchers can access and use these to drive research outputs, engagement and outcomes;
- developing the **capability** of our staff and students to optimise research performance and outcomes;
- building **community** within Macquarie and externally in pursuit of shared goals for raising the standards of research, collaboration, engagement and communication;
- re-thinking **career** paths, creating new opportunities for excellence and innovation in research.

---

<sup>3</sup> Information Technology has recently signed a commercial agreement for high performance computing (HPC), is preparing a call for tender for data storage, and plans to hire two eResearch engineers in the proposed new structure.

## INTERNAL CONTEXT

---

The ambitions articulated for Macquarie University in *A Framing of Futures* and the *Strategic Research Framework* provide the motivation for pursuing the transformative possibilities of a Data Science and eResearch Platform. The objective of accelerated and impactful performance in discovery will be further enabled by this strategy, and so too will the University's capacity to develop transformative learning, to innovate and to build reputation and presence locally and internationally.

Developing Data Science and eResearch capacity and capability is integral to improving research outcomes and extending the influence of the research undertaken at Macquarie. Developing and driving this strategically is important for all four of the *Strategic Research Framework* Key Research Objectives and Goals. In particular this strategy will position the University with relation to the following Strategic Research Framework goals and supporting strategies:

- identify the key research programs currently pursuing breakthrough research and/or seeking to contribute significantly to research breakthroughs and explore opportunities to accelerate their advancement;
- ensure the University facilitates research, its dissemination and re-use through open-access, and research data management and storage;
- expand and strengthen induction, mentoring and research skills support for early career researchers and HDR candidates;
- raise the profile of Macquarie University and the impact of our research at State and Federal Government levels and respond comprehensively and with agility to government funding announcements;
- improve policies, processes and systems that support the pursuit of excellence and integrity in research and scholarship.

### Strengths

This strategy builds on Macquarie's strengths by providing focus and strategic direction for Data Science and eResearch as platform of research support. Among the University's existing strengths are:

- Research teams developing digital data collection tools in Ancient History and Medicine and Health;
- Those working with national capabilities such as NeCTAR, Bioplatforms Australia and the Australian National Data Service;
- Programs aimed at broadening and enhancing the skill sets of HDR candidates such as the Digital Humanities program (FOAR705), Computer Science (COMP777), and Faculty-based programs such as those in the Genes to Geosciences research centre;
- A number of research staff and HDR candidates engaged in data science through communities of practice and professional development;
- Potential/planned expansion of our Data Science coursework offerings to post-graduate students.

## EXTERNAL CONTEXT

---

Data Science and eResearch tools, applications and practices are increasingly on the national and international radar as being core areas of growth and investment. From the National Science and Research Priorities to the growing future of Data Science jobs, there is an obvious need to navigate these spaces with a view to remain on the front foot. This section summarises the core external contexts.

### **National Science and Research Priorities**

Many, if not all, of the National Science and Research Priorities rely on Data Science and eResearch tools, applications and practices. The Australian Government acknowledges this by addressing elements such as big data, research infrastructure, workforce capability and collaboration in a number of government strategies. New measures of impact and engagement are emerging, and our capacity to grow our data-intensive research will have an important effect on our ability to maintain our high levels of research excellence against these measures.

### **Australian Government Department of Education and Training eResearch Framework: 2016 and beyond (in draft)**

Commissioned by the Department of Education and Training, the eResearch Framework establishes a planning framework based on four enduring components: national computational facilities, a cooperative system allowing connection and use of data to enable new research, creation of research applications to enhance infrastructure, and service providers enabling 'frictionless connectivity' and greater collaboration.

The Framework acknowledges that "to be globally competitive, opportunities must be identified and exploited by Australian research at least as quickly as by others and where affordable in advance of them". Priority for certain components delivered by the Framework will be given to NCRIS (National Collaborative Research Infrastructure Strategy) capabilities and centres of research excellence that use them. While some research teams use NCRIS capabilities<sup>4</sup>, Macquarie University must strengthen its position to capitalise on national infrastructure. Universities will be accountable for agreeing, contributing to and maintaining standards for research data integrity based on principles of "find-ability", accessibility, interoperability and re-useability (FAIR). The Framework points to the leadership role Universities can play by developing digital skills as part of researcher training programs.

### **National Collaborative Research Infrastructure Strategy (NCRIS)**

The Australian Government has recently recommitted funding of \$1.5 billion to the National Collaborative Research Infrastructure Strategy (NCRIS) over the next 10 years. It provides a national approach to investment, funding research infrastructure that supports Australian research and Australian research collaborations world-wide. In 2016 Australia's Chief Scientist is chairing an

---

<sup>4</sup> For example, the ARC CoE for Core to Crust Fluid Systems and the NeCTAR-funded UniCarbKB initiative, APAF (BioPlatforms), amongst others

expert group to identify national research infrastructure capability needs and provide a roadmap for the delivery of future requirements.

### **National Innovation and Science Agenda**

The Government has delivered a commitment to research and data-driven innovation through the National Innovation and Science Agenda (NISA). The NCRIS funding and strategy described above is one example. The Government's public data strategy, whereby the Government has pledged to connect and provide free and open access to non-sensitive government datasets is another. Demonstrating the pervasiveness of this move toward open access data across all sectors, The *Productivity Commission Inquiry into Data Availability and Use* has recently released a paper and called for feedback on the value and costs of making both public and private sector data more open and available for use. Other NISA initiatives include partnerships between PhD candidates and industry through Data61; building a Data Research Network to link businesses with data researchers; and furthering the Digital Technologies curriculum in schools.

### **Growth in Data Science jobs**

Data Science as a career has emerged as a huge growth area in academia, industry and government. Demand exceeds supply, with a MIT Sloan Management Review survey showing that 43% of companies see lack of appropriate analytical skills as a key challenge. The challenge for Universities will be attracting and maintaining a talented pool of data-science skilled researchers. The UK Open Research Data Forum highlights a 'major skills gap' for data scientists and suggests university training should focus on: data literacy for all; the development of generic data specialists, who might be best located as part of a library function; and specialists with domain expertise, who should be located in operational units (departments, centres) (2014). General Assembly, an international training organisation focusing on digital skill sets has recently run a forum in Sydney titled "*Transition to Tech: opportunities for academics in data*" demonstrating career pathways for academics to data science jobs in industry. This suggests both an opportunity and a challenge for universities – offering new career pathways for PhD graduates and a potential 'drain' of research staff to private-sector data science jobs.



## 1. Data management

### Why?

Research data is an asset which must be well managed if it is to be an asset of value. Good data management practice enables researchers to validate, reproduce, re-use and share (with consideration to ethical and privacy restrictions) their research data. The move toward open data is gaining momentum. Over time researchers will have increasing access to government, research and possibly commercial data and the collaboration possibilities that come with this.

For individual researchers and teams, data management allows validation of results and re-use of data for unique research problems. By publishing data, researcher citations are increased<sup>5</sup>. By describing and curating data, opportunities are created for collaboration and to communicate and showcase research.

When shared, research data has value to other researchers both within and outside the discipline it was created for – sometimes to progress the solution for which it was created, sometimes applied in unanticipated ways. Sharing data enables collaborations that contribute comprehensive solutions to the world’s pressing issues.

### Where do we want to be?

Where practicable, and where ethical and privacy restrictions do not apply, Macquarie University researchers will make their research data open. Researchers and students will have the capacity and support to optimise the contribution of their research data to the research and broader community. Good data management practices as described by the Macquarie Code for the Responsible Conduct of Research, internationally recognised standards and best practice will be embedded in the research lifecycle.

How?	Who?
<ul style="list-style-type: none"> <li>Clear Data Management policy and procedure in accordance with the Macquarie Code for the Responsible Conduct of Research.</li> </ul>	<ul style="list-style-type: none"> <li>University Library</li> <li>Research Office</li> <li>Information Technology</li> <li>Advisory Group</li> </ul>
<ul style="list-style-type: none"> <li>Secure and reliable data storage and clear mechanisms for access</li> </ul>	<ul style="list-style-type: none"> <li>Information Technology</li> </ul>
<ul style="list-style-type: none"> <li>Education, communications, support and tools (including Data Management Planning tool) to enable good data management practices including advice with regard to data management planning, storage, data description, curation, sharing, archiving</li> </ul>	<ul style="list-style-type: none"> <li>University Library</li> <li>Research Office</li> <li>Information Technology</li> <li>Data Science and eResearch Team</li> </ul>
<ul style="list-style-type: none"> <li>Identify and pursue opportunities for Macquarie University to lead and contribute to the development of research data excellence across the five future-shaping research priorities</li> </ul>	<ul style="list-style-type: none"> <li>Faculties</li> <li>Data Science and eResearch Team</li> <li>Advisory Group</li> </ul>

<sup>5</sup> Piwowar HA, Vision TJ (2013) <https://dx.doi.org/10.7717/peerj.175>

## 2. Underpinning capacity

### Why?

Data Science and eResearch *capacity* is the underpinning leadership, systems, processes, policy and support structures that enable researchers to ‘focus on what they alone can do’<sup>6</sup>. These provide the supporting platform upon which capabilities can be extended and accelerated and impactful research can be realised. Where organisational capacity and researcher capability align, the opportunity for a step-change in innovative, world-leading, world-changing research becomes a greater possibility.

### Where do we want to be?

Following a clear vision, and guided by strong and inclusive leadership, Macquarie will strategically use, develop and acquire institutional infrastructure (physical, technological, human resources, structural) and funds to accelerate performance in the five future-shaping research priorities.

How?	Who?
Institute effective governance, leadership and decision making via a <b>Data Science and eResearch Advisory Group</b> .	<ul style="list-style-type: none"> <li>• DVCR</li> <li>• PVCR (P&amp;I)</li> </ul>
Establish a <b>Data Science and eResearch Platform</b> by recruiting a Data Science and eResearch team consisting of a Director, an eResearch Analyst/Project Manager and an administrative support team member. Within the DVCR portfolio, the team will guide and coordinate the activities described within this strategy, provide advice and expertise, and contribute to building capacity through external funding bids. This team will work closely with current and future lead researchers in data science and its applications, and with the two eResearch staff proposed to be appointed as part of the current IT change process.	<ul style="list-style-type: none"> <li>• DVCR</li> <li>• PVCR (P&amp;I)</li> </ul>
Develop and implement <b>policy, procedure and guidelines</b> that support management of eResearch, for example data management, access to high performance computing facilities, internal and external funding.	<ul style="list-style-type: none"> <li>• University Library</li> <li>• Information Technology Research Office</li> <li>• Data Science and eResearch Team</li> </ul>
<b>Manage data storage and high performance computing solutions</b> via an external provider. Researchers will be made aware of additional or alternative data storage options for use at a discipline level.	<ul style="list-style-type: none"> <li>• Information Technology</li> <li>• University Library</li> <li>• Faculties</li> </ul>
<b>Investigate mechanisms for funding</b> <sup>7</sup> to support the development of research projects that rely on eResearch tools and practices, to ensure the longevity and sustainability of projects post the term of the project.	<ul style="list-style-type: none"> <li>• DVCR</li> <li>• PVCR (P&amp;I)</li> <li>• Faculties</li> <li>• Data Science and eResearch Team</li> </ul>

<sup>6</sup> Our University: A Framing of Futures

<sup>7</sup> Internal funding and leveraging external opportunities via funding bodies, partners, collaborators.

Build **stronger relationships with agencies** such as ANDS, government eResearch bodies, funding bodies etc.

- Data Science and eResearch Team
- Faculties

### 3. Building staff and student capability

#### Why?

By developing a strong data science learning culture Macquarie positions itself as a destination of choice for researchers and higher degree research candidates. Data science as a career is fast-growing in academia, industry and government. By building knowledge and skills we accelerate research results, enable collaboration, and enhance our competitive advantage to win government funding and industry partnerships.

#### Where do we want to be?

Macquarie researchers will possess the capabilities to maximise the potential of their research data to produce accelerated and world-leading outputs, higher degree research candidates will have the skills and knowledge sought after by the job market, and HDR supervisors will be equipped to support this aim.

Our culture will be one that fosters shared learning among researchers, research students, and professional staff within and across disciplines. Through communicating and celebrating the work of Macquarie University researchers and research students we will encourage excellence and motivate learning.

Both a University-wide and Faculty/Department-specific approach is required that enables opportunities for Higher Degree Researcher students and researcher staff. Programs must acknowledge and cater for different skill levels and different disciplinary needs.

How?	Who?
<ul style="list-style-type: none"> <li>• <b>University-wide</b> staff and HDR development:               <ul style="list-style-type: none"> <li>○ Research productivity/technology/tools workshops</li> <li>○ Software carpentry workshops</li> <li>○ Drop in sessions</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Data Science and eResearch Team</li> <li>• Advisory Group</li> <li>• University Library</li> <li>• Information Technology</li> </ul>
<ul style="list-style-type: none"> <li>• <b>Faculty and departmental</b> staff development:               <ul style="list-style-type: none"> <li>○ Research productivity/technology/tools workshops</li> <li>○ Software carpentry workshops</li> <li>○ Drop in sessions</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Data Science and eResearch Team</li> <li>• Advisory Group</li> <li>• University Library</li> <li>• Information Technology</li> </ul>
<ul style="list-style-type: none"> <li>• <b>Higher Degree Research</b> development:               <ul style="list-style-type: none"> <li>○ HDR coursework programs</li> <li>○ HDR Skills Workshops</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Dean, HDR</li> <li>• Data Science and eResearch Team</li> <li>• Advisory Group</li> <li>• University Library</li> <li>• Information Technology</li> </ul>
<ul style="list-style-type: none"> <li>• Information Technology will provide training for the products and services delivered by that Unit, including High Performance Computing and Data Management tools</li> </ul>	<ul style="list-style-type: none"> <li>• Information Technology</li> </ul>

While outside the scope of this strategy it is acknowledged that undergraduate and post-graduate coursework programs may also assess data science literacy programs and graduate attributes in line with the objectives of *Learning for the future; Learning and Teaching Strategic Framework: 2015 - 2020*.

#### 4. Community of practitioners

##### Why?

Cross-disciplinary research, and learning, will be enhanced through community. Good data management and open data practices open possibilities for novel and diverse approaches to problems. Community – learning through sharing, communication, collaboration – is the avenue through which these possibilities can be realised.

##### Where do we want to be?

A vibrant and active community of data science and eResearch practitioners, evidenced through University-wide communication, promotion and celebration; outreach and engagement with the wider community through fora, showcases, classes and citizen science; events and Faculty/Department-specific activities designed to foster a culture of shared learning; partnerships and collaborations. Through community we give recognition to and promote our strengths.

How?	Who?
<ul style="list-style-type: none"> <li>• Promote, raise visibility and awareness of data science and eResearch activity via a communications strategy.</li> </ul>	<ul style="list-style-type: none"> <li>• PVCR Coordinator, through consultation with Advisory Group and Data Science and eResearch Team</li> </ul>
<ul style="list-style-type: none"> <li>• Community of Practice Events:               <ul style="list-style-type: none"> <li>○ Informal groups</li> <li>○ Bi-monthly speaker/demonstration/panel and networking</li> <li>○ Hacky Hour</li> <li>○ Open Access Week</li> <li>○ ResBaz</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Data Science and eResearch Team</li> <li>• Faculties</li> </ul>
<ul style="list-style-type: none"> <li>• Research:               <ul style="list-style-type: none"> <li>○ Collaborations and partnerships</li> <li>○ Open data, Open access</li> <li>○ Citizen Science Projects</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Data Science and eResearch Team</li> <li>• Research Partnership Managers</li> <li>• University Library</li> <li>• Information Technology Researchers</li> <li>• Faculty</li> </ul>
<ul style="list-style-type: none"> <li>• Committees               <ul style="list-style-type: none"> <li>○ University Data Science and eResearch Advisory Group</li> <li>○ Faculty Data Science and eResearch sub-committees (of Faculty Research Committee)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• DVCR</li> <li>• PVCRs</li> <li>• ADRs</li> </ul>
<ul style="list-style-type: none"> <li>• Further develop relationships with agencies such as ANDS</li> </ul>	<ul style="list-style-type: none"> <li>• Data Science and eResearch Team</li> </ul>

## 5. Research careers

### Why?

The Data Scientist/eResearcher career path presents challenges for universities. Skills required in data science are highly sought after in academia, government and industry, making attraction and retention of Data Scientists a critical issue for research performance. Like other universities, Macquarie needs to consider how academic careers are changing as research crosses over into data science and programming. We need to consider strategies for integrating software development staff within the research team. Integral to this is how we attract and retain research staff with the required capabilities and the passion to contribute to world-leading research. In our consultations with the leading eResearch groups in Australian universities this was highlighted as a major issue and one for which there is yet a solution to be found

### Where do we want to be?

An employer of choice for researchers.

How?	Who?
Investigate alternative researcher career pathways and e-researcher careers, attraction and retention strategies.	<ul style="list-style-type: none"><li>• PVCR (P&amp;I)</li><li>• PVCR (I&amp;D)</li><li>• Human Resources</li><li>• Data Science and eResearch Team</li></ul>

## STAKEHOLDERS

---

### Internal

The success of this strategy relies on the ability of the following key stakeholders together:

- Pro Vice-Chancellor, Research Performance and Innovation
- Data Science and eResearch team (new)
- University Library
- Information Technology
- Faculty Champions
- Research Office
- Dean, HDR

It is recommended the Pro Vice-Chancellor (Research Performance and Innovation) have broad oversight of the five key elements of the strategy. The PVC (RP&I) will provide governance and leadership as Chair of a **Data Science and eResearch Advisory Group**, which will advise the DVC, Research with regard to strategic data science and eResearch resourcing. The group above will provide representation on the Advisory Group (diagram 1) as well as support the five key elements of the strategy through the work of each respective unit.

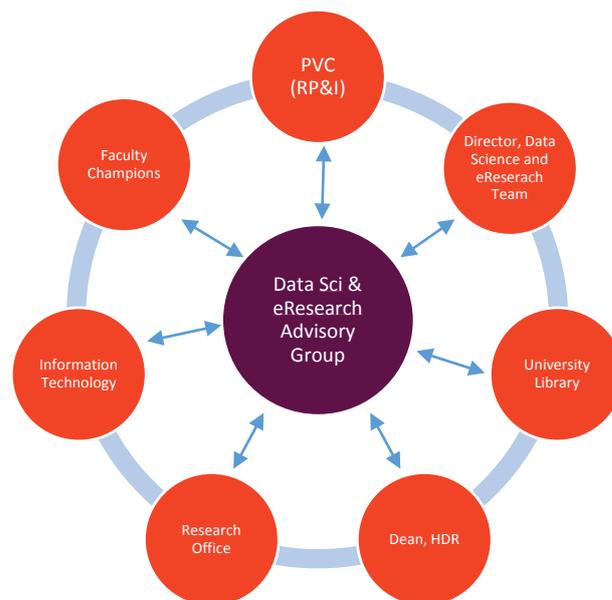


Diagram 1: Data Science and eResearch Advisory Group

### External

Macquarie will need to build on its existing relationships with NCRIS-funded organisations. Through the development of eResearch expertise Macquarie will be more strongly positioned to bid for funding and to influence the direction of eResearch capability within Australia.

Government, NGOs and industry partnerships (existing and potential) will be both contributors and beneficiaries of a more research data-intensive Macquarie.