



UNSW
SYDNEY

Australia's
Global
University

UNSW RESEARCH STORAGE INFRASTRUCTURE – CURRENT AND CONCEPTUAL TARGET STATE

PREPARED: 2025, BY: RIZWAN S AND MOHAMMAD ISLAM

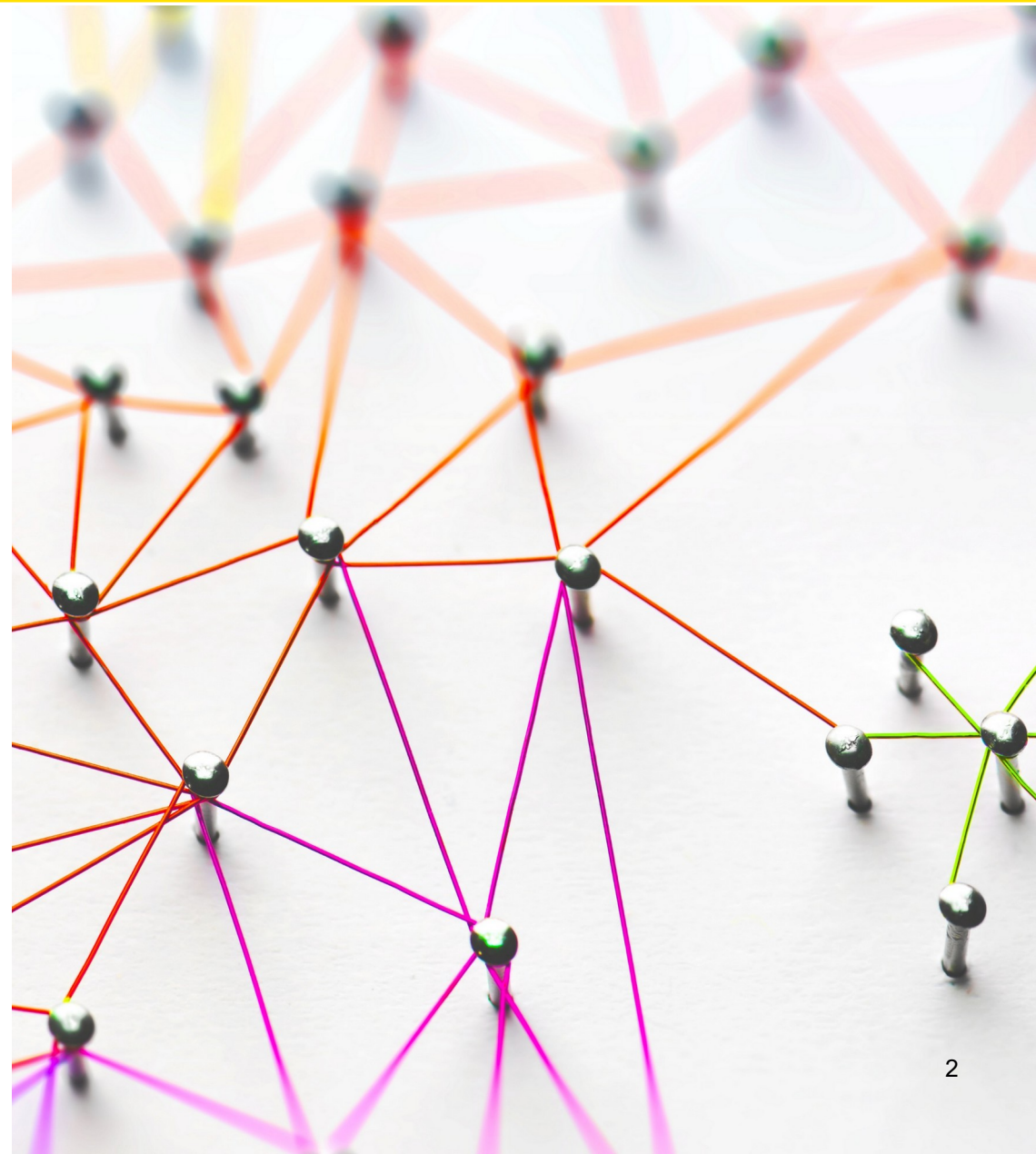
UNSW Research Storage Infrastructure

- | | | |
|----|-------------------------|-------------------|
| 1. | Background and Context | 3 |
| 2. | Case for Change | 4 |
| 3. | What Value RDE Delivers | 5 |

CURRENT STATE AND CONCEPTUAL ARCHITECTURE

- | | | |
|----|--------------------------------------|--------------------|
| 4. | Current State | 6 |
| 5. | Motivations | 11 |
| 6. | Solution Approach | 12 |
| 7. | Conceptual Target state - Platform | 13 |
| 8. | Conceptual Target state - Data Tools | 14 |
| 9. | Conceptual Target state - TRE | 15 |

- | | | |
|--|-----------------|--------------------|
| | APPENDIX | 17 |
|--|-----------------|--------------------|

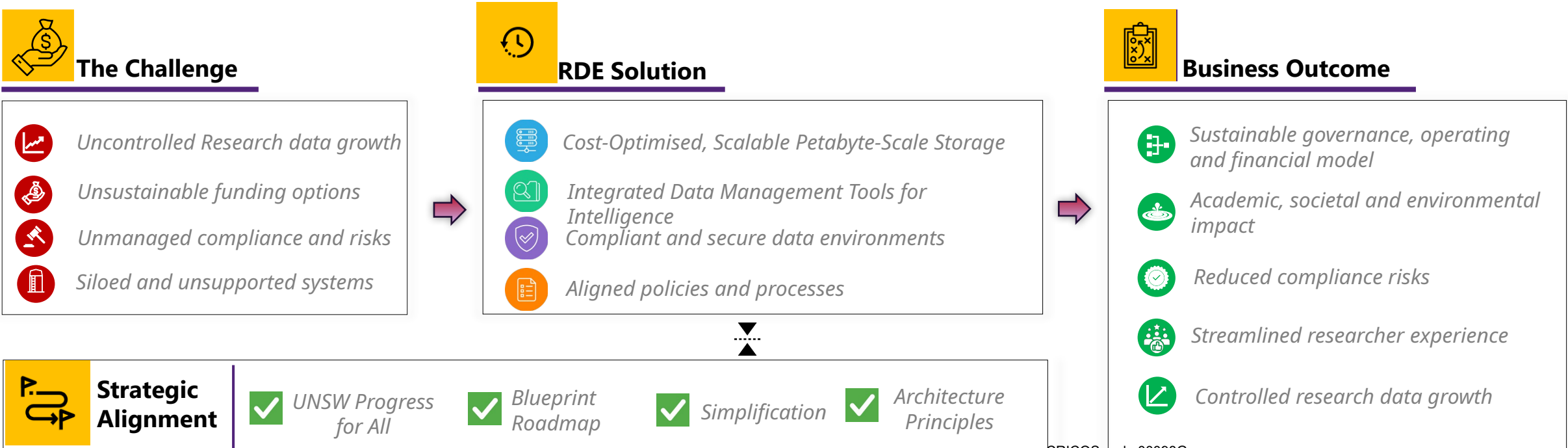


Program Background

UNSW's research data is growing rapidly around **30% annually** driven by advanced instruments and AI models. Current systems are **fragmented**, **compliance** is weak, and **funding** is unsustainable, creating rising costs and risks.

The RDE Program addresses the urgent need for a **scalable**, **cost-effective**, and **governed** research data environment by delivering a secure platform with integrated tools and policy-aligned governance, enhancing researcher experience and enabling future opportunities.

The program will be executed in three phases: **Plan** (2024-25), **Deliver** (2026–2028), and **Sustain** (2029 onward). These stages will enable scalable infrastructure, automation, and streamlined access, underpinned by high-speed networks, petabyte-scale tiered storage, and integrated metadata systems



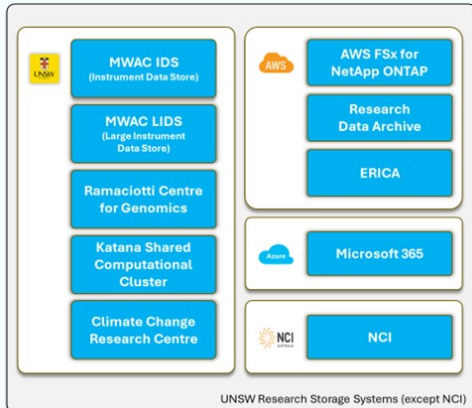
Why (Case for change)

The case for change is driven by fragmented and siloed research platforms that hinder collaboration and consistency. Internal assessments including a Gartner maturity review, Platform heatmap, and sector benchmarking revealed UNSW's position as low to medium across key capabilities. These findings highlight critical risks in scalability, resilience, and cyber readiness, reinforcing the need for urgent uplift.



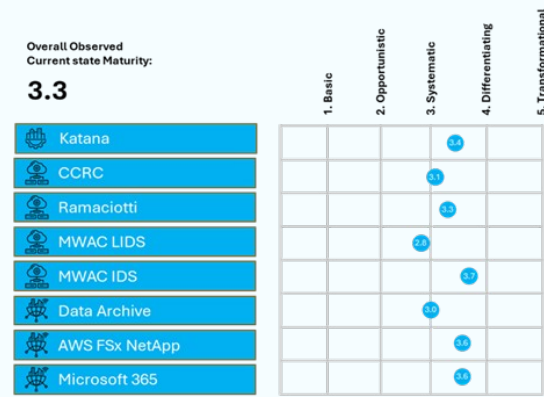
Silo Research Data Platforms:
A

Siloed platforms lead to fragmented data management, causing duplication of effort and inconsistent practices across research units. This hampers collaboration, increases compliance risk, and results in a poor researcher experience.




Gartner Maturity Assessment
B

The Gartner maturity assessment revealed low to medium scores across core capability, technology fit, business alignment, and operational efficiency. These results highlight significant risks and underscore the need for targeted improvements to meet future research demands.




Research Platform Heatmap
C

We conducted a heat map assessment of our research data platforms, uncovering key risks in areas such as ageing technology, limited scalability, resilience gaps, capacity constraints, and inadequate cyber security controls. These issues pose challenges to long-term sustainability and compliance.

Platform	Platform Age	Availability/Resilience to Data Corruption (RPO/RTO)	Scalability Capacity	Disaster Readiness	Business Continuity	Technical and Architectural	Storage Alignment	Security Maturity (by CIS)
MWAC IDS	10 years	High	Low	Low	Low	Low	Low	Low
MWAC LIDS	10 years	High	Low	Low	Low	Low	Low	Low
Ramaciotti	10 years	High	Low	Low	Low	Low	Low	Low
Katana	10 years	High	Low	Low	Low	Low	Low	Low
Data Archive	10 years	High	Low	Low	Low	Low	Low	Low
AWS FSx NetApp	10 years	High	Low	Low	Low	Low	Low	Low
Microsoft 365	10 years	High	Low	Low	Low	Low	Low	Low

CRICOS code 00098G



Sector Benchmarking
D

Our peers have made significant investments to uplift their research data management capabilities over the past decade. UNSW's research data management maturity lags behind comparable research-intensive Universities in Australia.



What Value RDE Delivers

This slide highlights the key value that the RDE Program brings to UNSW and its researchers enabling a scalable, secure, and efficient data ecosystem. It showcases how RDE supports institutional goals while enhancing the research experience through improved access, collaboration, and compliance.



For UNSW



SCALABLE INFRASTRUCTURE

Supports petabyte-scale data growth through tiered storage and hybrid architecture.



COMPLAICE & RISK MANAGEMENT

Aligns with UNSW, national data policies, improves data classification, retention, and auditability.



RESEARCH POSITIOING

Enhances competitiveness for funding and partnerships by providing world-class data infrastructure.



VISIBILITY & PLANNING

Enables better forecasting with data dashboards and centralised monitoring tools.



OPERATIONAL EFFICIENCY

Consolidates fragmented systems, reduces manual processes, and optimises support efforts.



FUNDING & GOVERNANCE

Enables transparent costing, accurate forecasting, and sustainable operating models.



STRATEGIC ALIGNMENT

Directly supports Blueprint, digital uplift, research excellence, and operational resilience goals.



ENTERPRISE RESEARCH PLATFORM

Breaks down silos across faculties, ensuring a consistent and integrated approach to research data.



For Researcher



SIMPLIFIED COLLOBORATION

Enables secure, real-time data sharing with colleagues and external partners.



USER EXPERIENCE

Provides a standardised, easy-to-navigate data environment across disciplines and platforms.



DATA CONTROL & TRANSPERENCY

Empowers researchers with insights into their own data usage and storage consumption.



REDUCED OVERHEAD

Automates time-consuming processes like storage requests, tiering, and archiving.

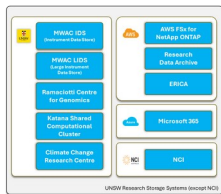


Current State & Conceptual Architecture

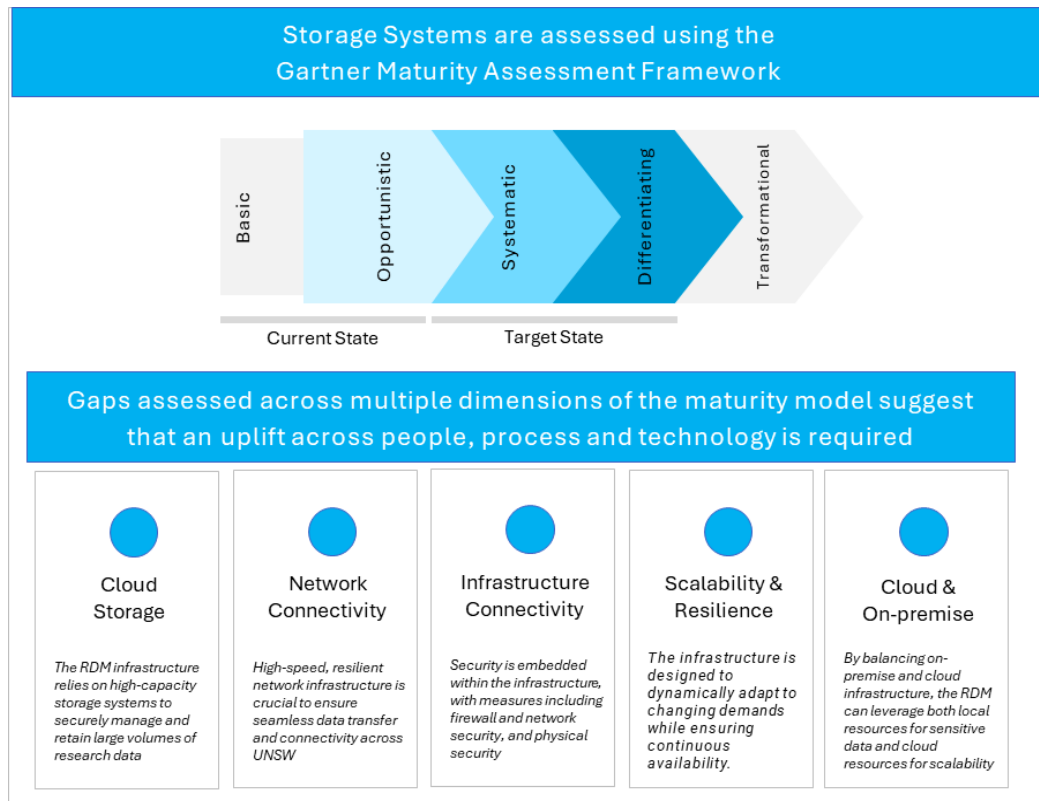


Research Storage Platform – Maturity Assessment

This slide describes the **Gartner Maturity Assessment Framework** and its use to determine the maturity of the UNSW Research Storage Platform. The goal is to identify each storage system's maturity, capability gaps against best practice, and align the systems with our strategic objectives.



UNSW

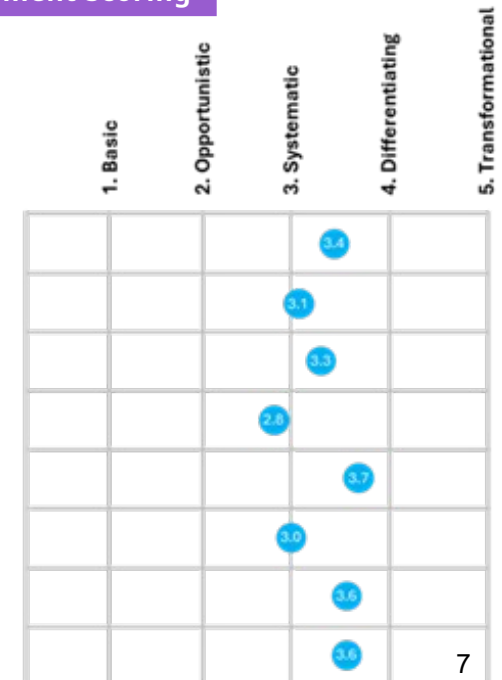


Assessment Scoring

Overall Observed Current state Maturity:

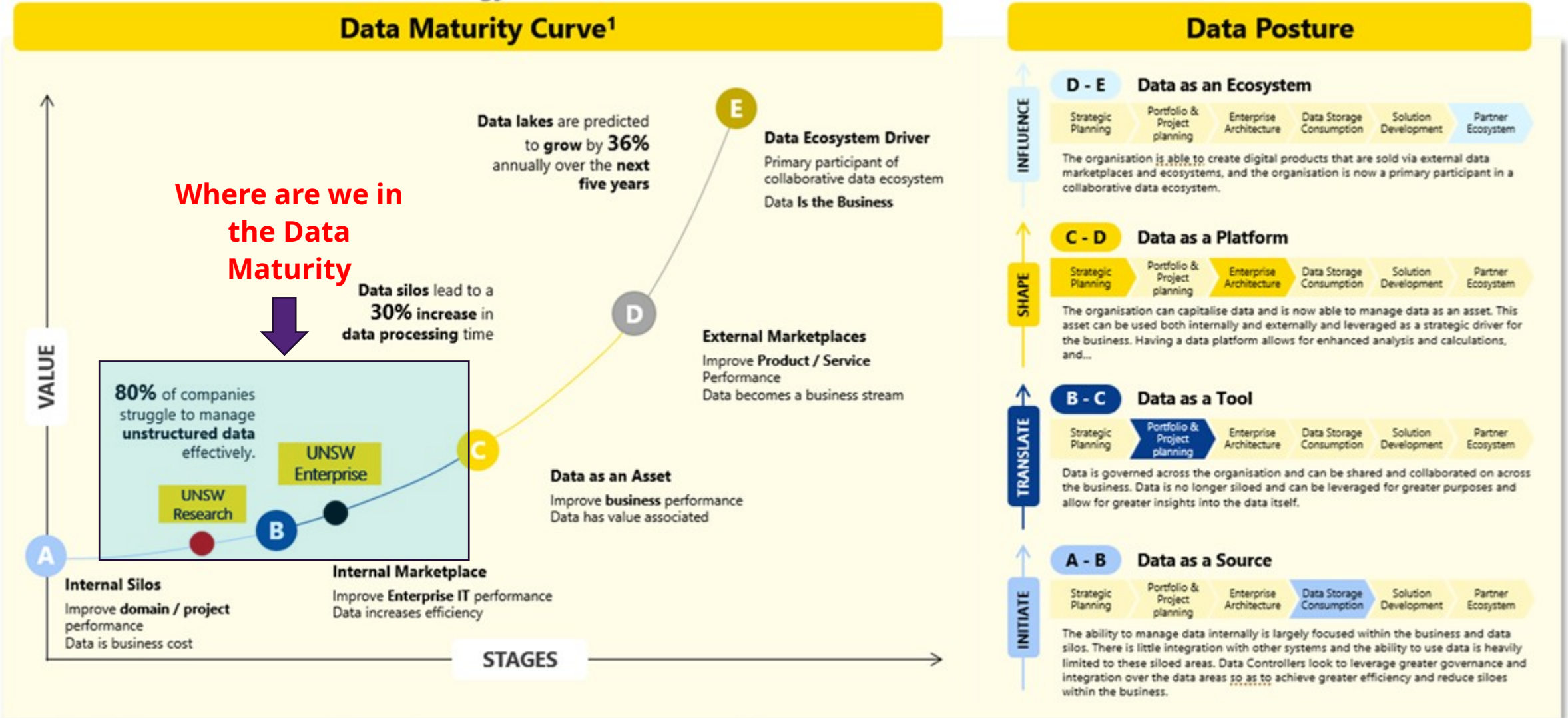
3.3

- Katana
- CCRC
- Ramaciotti
- MWAC LIDS
- MWAC IDS
- Data Archive
- AWS FSx NetApp
- Microsoft 365



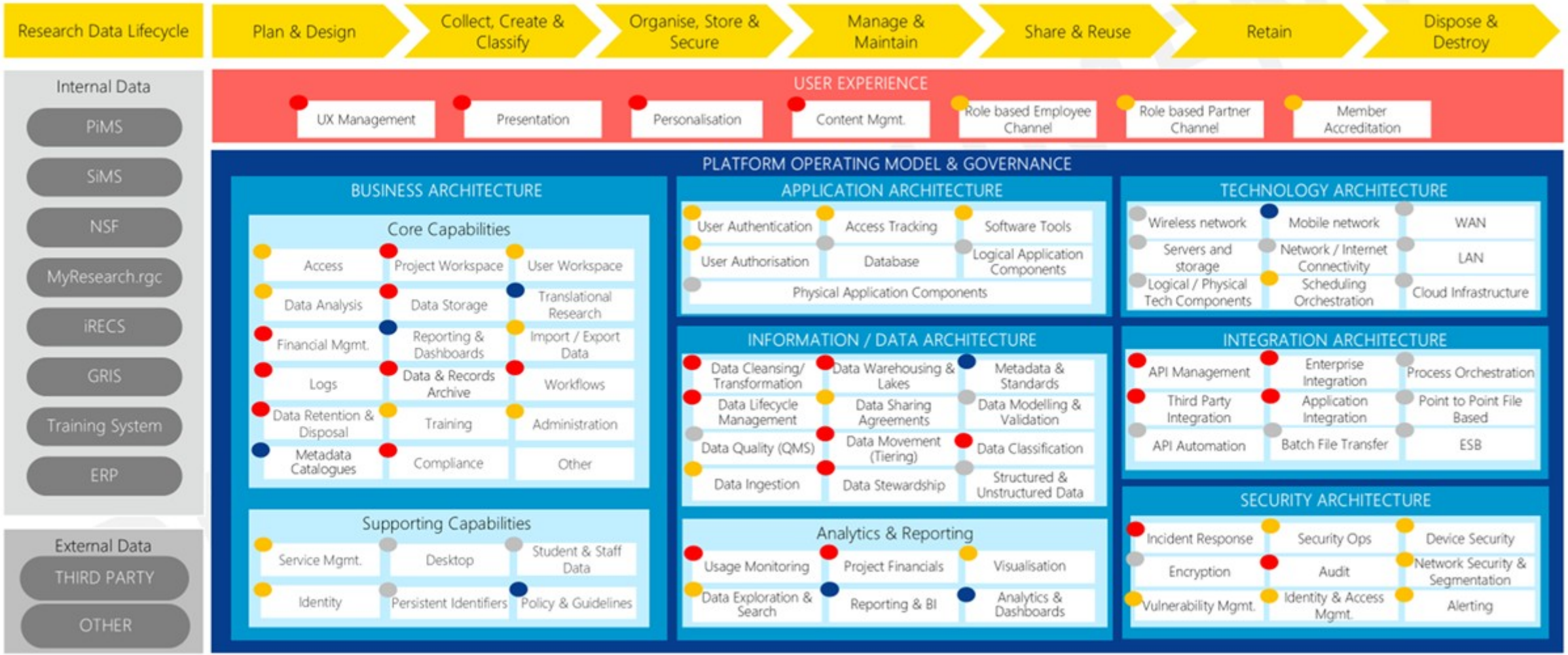
Data Maturity Curve

Stored research data poses a risk to UNSW without proper review processes. Currently, 16% of Research Data Management Plans (RDMPs) lack data classification, leaving the type of data unknown. Additionally, approximately 36% of the data is classified as sensitive or higher.



BDAT Capability Heatmap

We have used the Data Capability Framework as a baseline and aligned the specific elements across the 7 lenses-based categories, culminating in the 6 layers of architecture



● Does not Exist ● Needs Improvement ● Critical Gap ● No gaps identified

Research Journey Map – HCD Pain Points

DRAFT



Current State Research Data Journey

UNSW researchers experience numerous pain points when engaging along the research data lifecycle.

Characteristics of data that researchers are managing can drive specific pain points in their journey

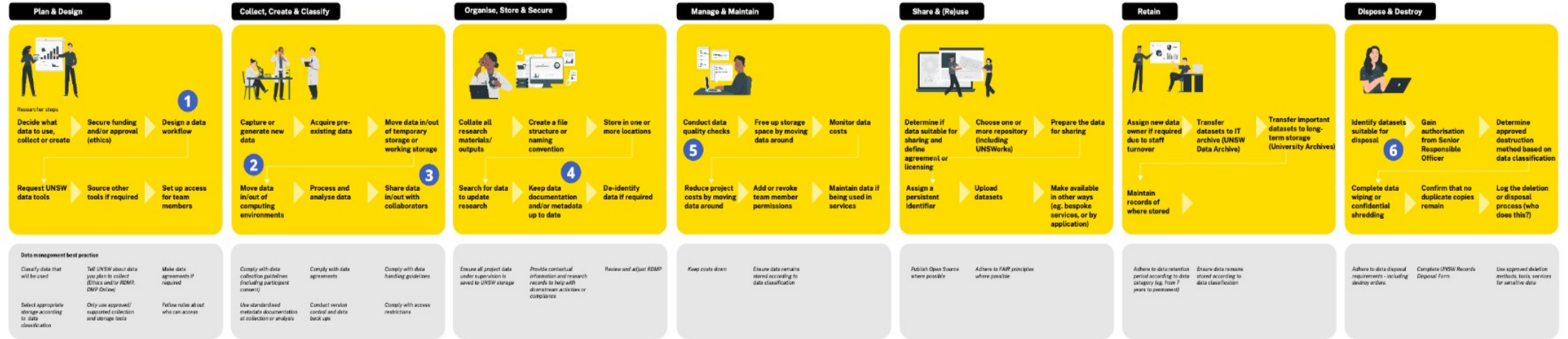
- Source of input data**
The origin of research data – whether openly available, generated by specific experiments, collected via sensors – often has the level of complexity in access, origin, and compliance
- Level of sensitivity**
Data containing sensitive or confidential information requires extra safeguards, adding layers of security, ethical considerations, and regulatory compliance to the research process.
- Size of data**
The size of data impacts how researchers store, process and transfer information, with large datasets posing logistical and technical challenges.
- Distribution**
How data is managed across teams, institutions or platforms influences collaboration, accessibility and governance.
- Output of data**
Research data can be transformed into reports, publications, presentations, shareable data sets or digital tools each with unique considerations for accessibility, longevity and reuse.
- Age of data**
The relevance, usability and management of data evolve over time, requiring strategies for archiving, updating or retiring datasets.

Research data journey

Pre-Research Phase

Active Research Phase

Post Research Phase



What's getting in the way of research data management best practice

- Policy processes are complex, time consuming, frustrating and not fit for purpose.
- Approvals and agreements have to be done every time.
- The same information is requested by multiple research support services.
- The data plan is not always documented because either no perceived value to the researcher, already considered as 'done' in the ethics approval or researcher standard protocol, or don't know about RDM.
- Best practice is hidden in long-winded policy documents. It's hard to find expert help or definitive answers.
- Researchers are nervous about doing the right thing with sensitive data.
- It can seem easier to 'do it my own way' (eg. just set up One Drive folders).
- Lack of awareness about what's available inside and outside UNSW, or preferred data tools not approved.
- Hard to source or innovate with new tools, due to approvals process.
- Collection tools have limited technical support (eg. OA on complex variables in RStudio, dealing with troubleshooting errors).
- Managing large data sets slows some research down, as researchers negotiate download, transfer times or jagged storage limits across devices.
- Can be hard to 'bolton' files due to connectivity issues during large data transfer.
- Available computing can be a research project limiter for some.
- Aligning storage & collaboration to data sharing agreement can be complex and hard to evidence.
- Standards may not align between collaboration parties (e.g. security).
- Credential clashes and usability issues dampen the goodwill of external data collaborators & partners.
- Collaborator access not consistent between tools. Some tools don't have any access, or are trickier to use (eg. Data Archive).
- Often forced to duplicate data to enable workflow across teams (eg. multiple students working on the same file).
- Some researchers are manually tracking data workflow, rather than doing it in a systematic or shareable way.
- It can be hard to know what data others in the same unit are working on, and this limits opportunities for collaboration.
- If researchers annotate data, often they do it for personal or project needs only (rather than considering future data users).
- Researchers are sometimes forced to re-run research when information about what former team members have done is lost or never captured.
- Even if a directory structure or naming conventions are agreed, it's often documented separately to data.
- Some projects require detailed traceability from data through to results/recommendations, and this is complex to achieve.
- It can be time-consuming to de-identify data. This is especially problematic at the end of project when the money is running out.
- Project data may be spread across multiple systems.
- Some projects suffer from 'versionitis' with many copies of the same data.
- Lack of awareness about data archive as a tool for managing storage needs only (rather than considering future data users).
- Difficult to assess data management best practice of others. It's currently a trust based system.
- From time to time, students or researchers walk away without updating data to shared drives – the research is lost.
- Some researchers are tracking usage across multiple resources, both internal and external.
- Time consuming to manage individual accounts & payments (no enterprise access).
- Staff turnover can lead to security risks as people forget to remove access permissions.
- It can be time consuming to prepare data for sharing, especially if this falls to someone outside of the original research team.
- Data sets are published in lots of different places, including both internal and external repositories, GitHub.
- Lack of long term storage for permanently published data products as a service.
- Writing until the end of a project to de-documenting and labelling often leads to incomplete or patchy records.
- Retention periods often longer than researcher tenure.
- Unclear on who's responsibility it is to adhere to data storage periods over the long term.
- Confusion about what goes into IT data archive vs University Archives.
- Lack of metadata can make the University Archives process more time consuming.
- There is no paired approach to data rationalisation, mostly leaving it up to the researcher.
- Terms such as 'community significances' are open to interpretation.
- There can be multiple copies or versions of research data.
- It's too hard/ time-consuming or expensive to figure out what to delete and get approval.
- It feels like a big deal to destroy any file, even WIP or copies.
- Concern that data might be useful to someone in the future.
- There can be a time and dollar investment to using approved processes and disposal services.
- The default becomes to 'keep everything'.

Priority pain points

- Getting on the right track**
Problem to solve: Helping researchers easily identify tools, processes and support to set up the research data effectively from the start.
- Managing large data**
Problem to solve: Reducing friction for researchers when moving or working with large data.
- Collaborating inside & outside UNSW**
Problem to solve: Making it easier for researchers to securely share and collaborate on data across teams, institutions, platforms.
- Keeping data tidy**
Problem to solve: Supporting researchers in maintaining organised, well-structured data throughout their research.
- Knowing who is using what, where, and when**
Problem to solve: Giving researchers clear oversight of their data, to simplify or enhance data management.
- Confidently rationalising data**
Problem to solve: Enabling researchers to confidently keep, archive or delete, ensuring compliance and long-term value.

We identified 12 Motivations for uplifting existing capabilities



Enhanced Performance & Speed

Significantly improve data processing speeds, enabling faster analysis and decision-making. This is crucial for research environments where timely insights are essential.



Scalable Platform

As data volumes grow, a scalable platform allows for handling of increased amounts of data without compromising performance and manage efficiency and bottlenecks.



Supporting Advanced Research

Provide a more powerful platform capable of using cutting-edge technologies and techniques to support advanced research and large-scale data analysis. Ensure secure access to sensitive data, such as personal research information and records.



Enhanced Data Management

Streamlining data handling, storage, and sharing processes will ensure efficient data management and compliance with UNSW guidelines and policies. A modernized system will ensure that the security of sensitive data is optimized.



Better User Experiences, Increased Attraction & Retention

An updated and modern system will allow for a better user experience and satisfaction, and a well-regarded data security system with more functionality will attract researchers to UNSW and ensure that they remain at UNSW.



Strengthened Cybersecurity Posture

Meet UNSW's strategic goal of becoming a top 50 university with strong data security infrastructure in place and processes in place to expand secure data architecture.



Improved Governance & Compliance

Aligned data governance and architecture plan which can be implemented across the university to enable a sustainable and efficient delivery of future data initiatives.



Reduce Potential Costs

Newer technologies can reduce operational costs through better resource management and automation, by reducing the cost of maintaining an outdated system and storage costs as well as optimising compute usage.



Enhanced Reporting capabilities

Implementing reporting capabilities will enable greater oversight and insight into how the system is running, and provide a better experience for the users



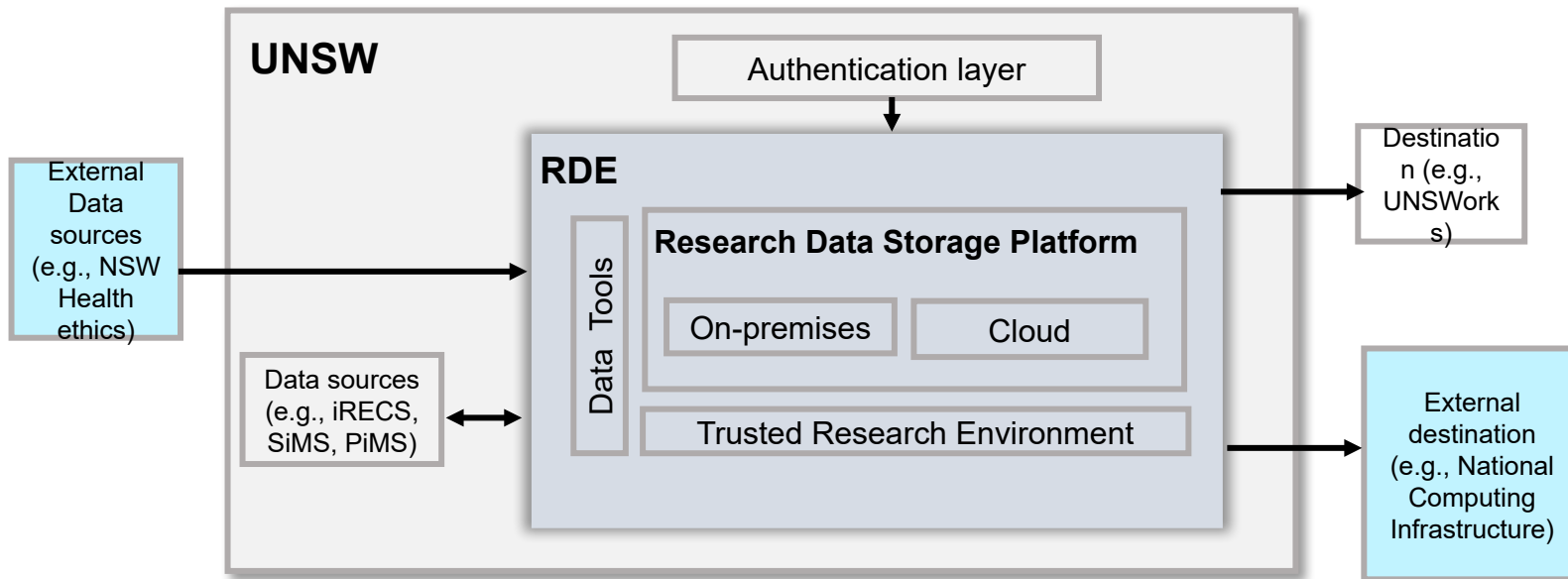
Improved Integration with other systems

Improving the system's ability to integrate with other software and systems will create a more streamlined and efficient research environment.

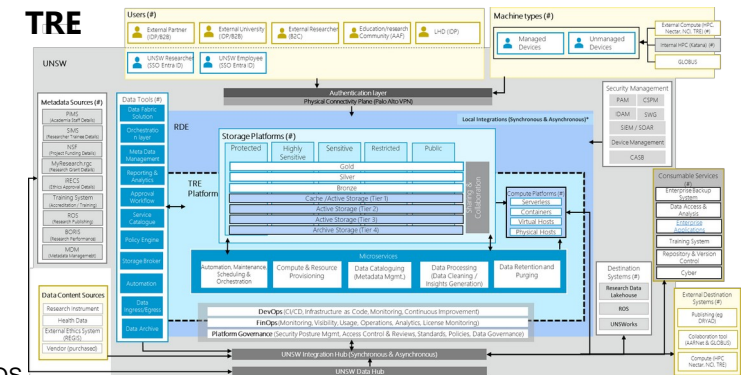
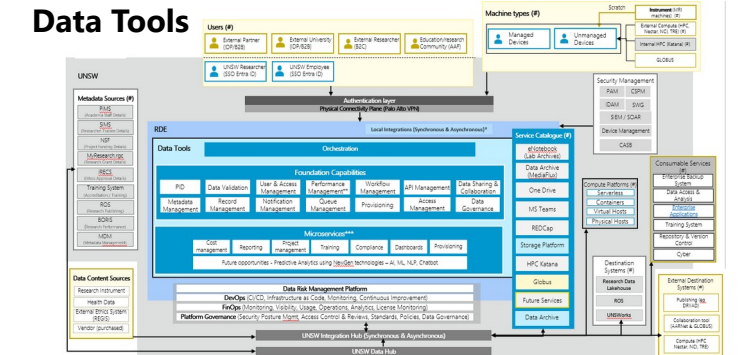
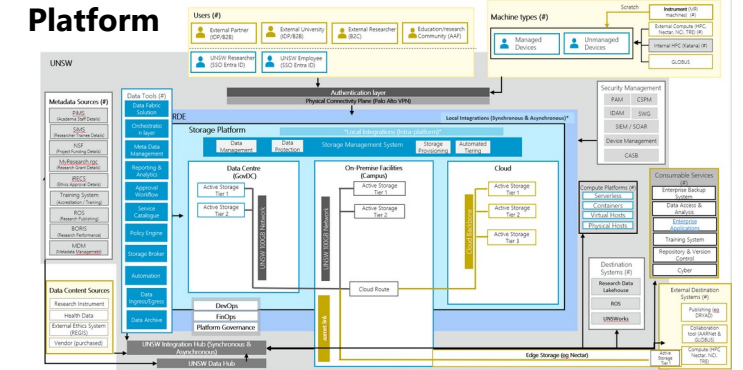
RDE Solution approach

The Research Data Platform's proposed design is a hybrid solution combining on-premises and cloud storage within a secure, trusted environment, supported by integrated data tools and unified authentication. This approach leverages the strengths of both environments on-premises for low-latency, high-performance workloads, and cloud for scalability and cost efficiency offering complementary capabilities to support diverse and evolving research needs.

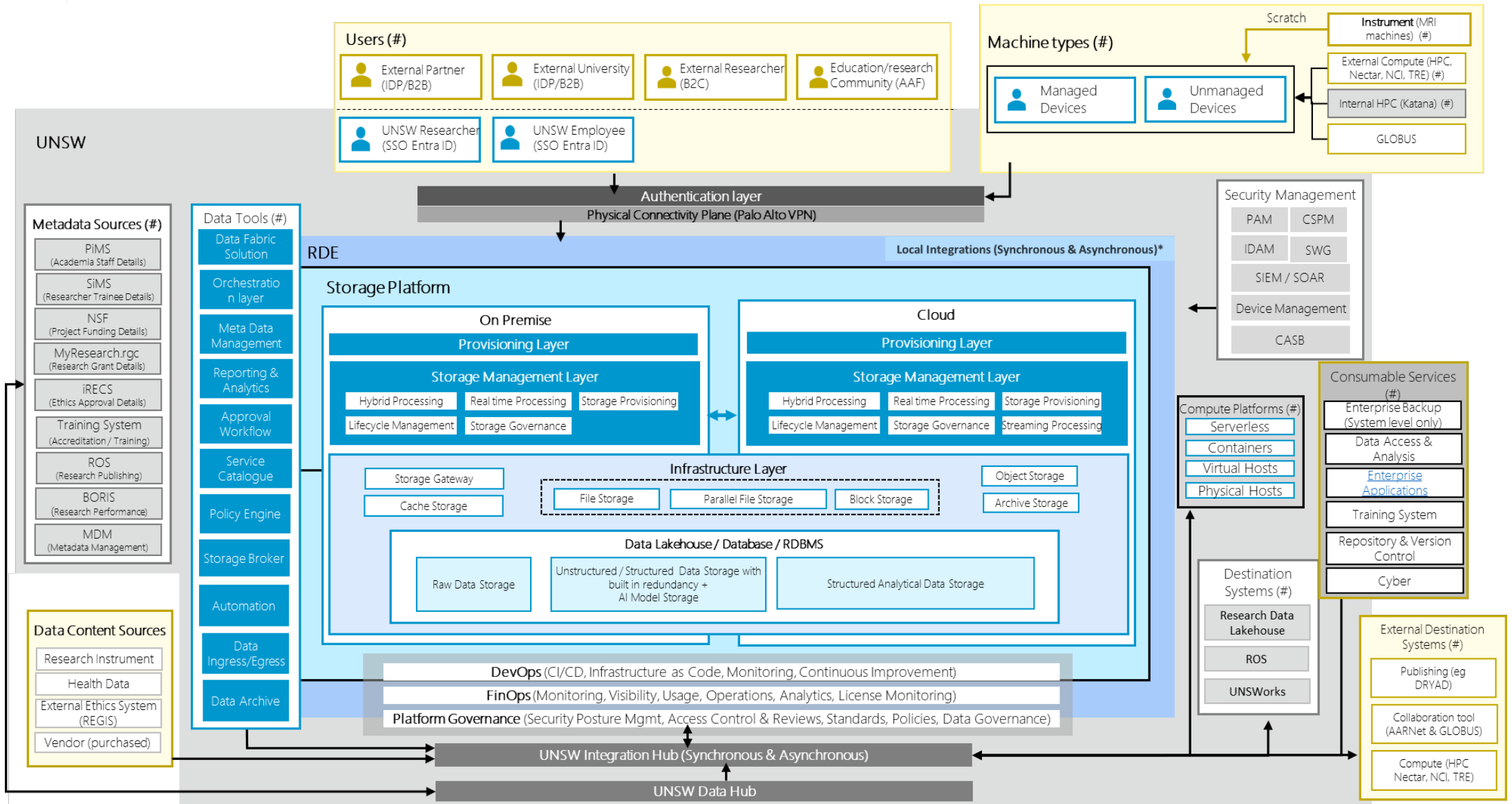
Context Diagram



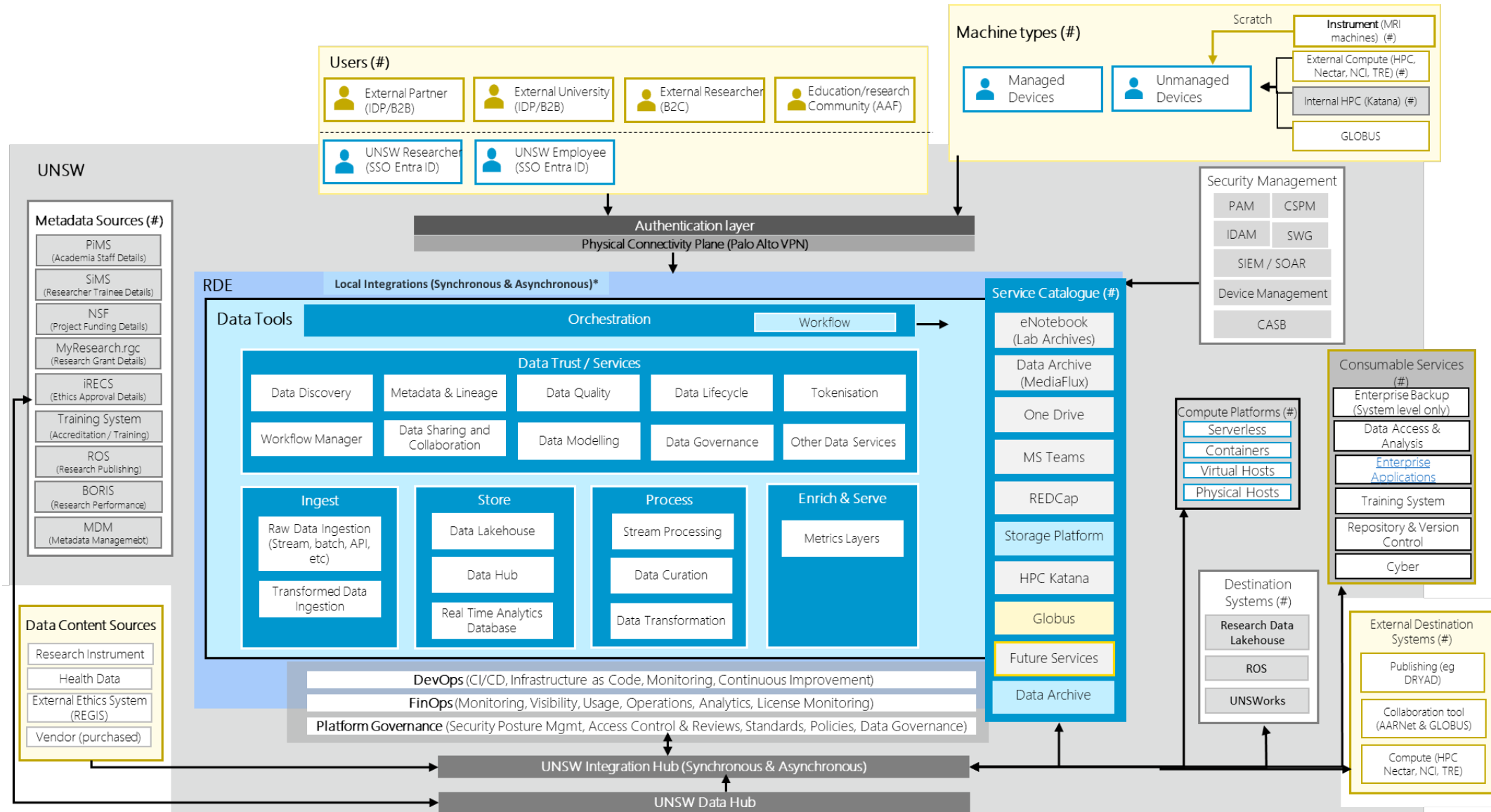
Conceptual Diagram



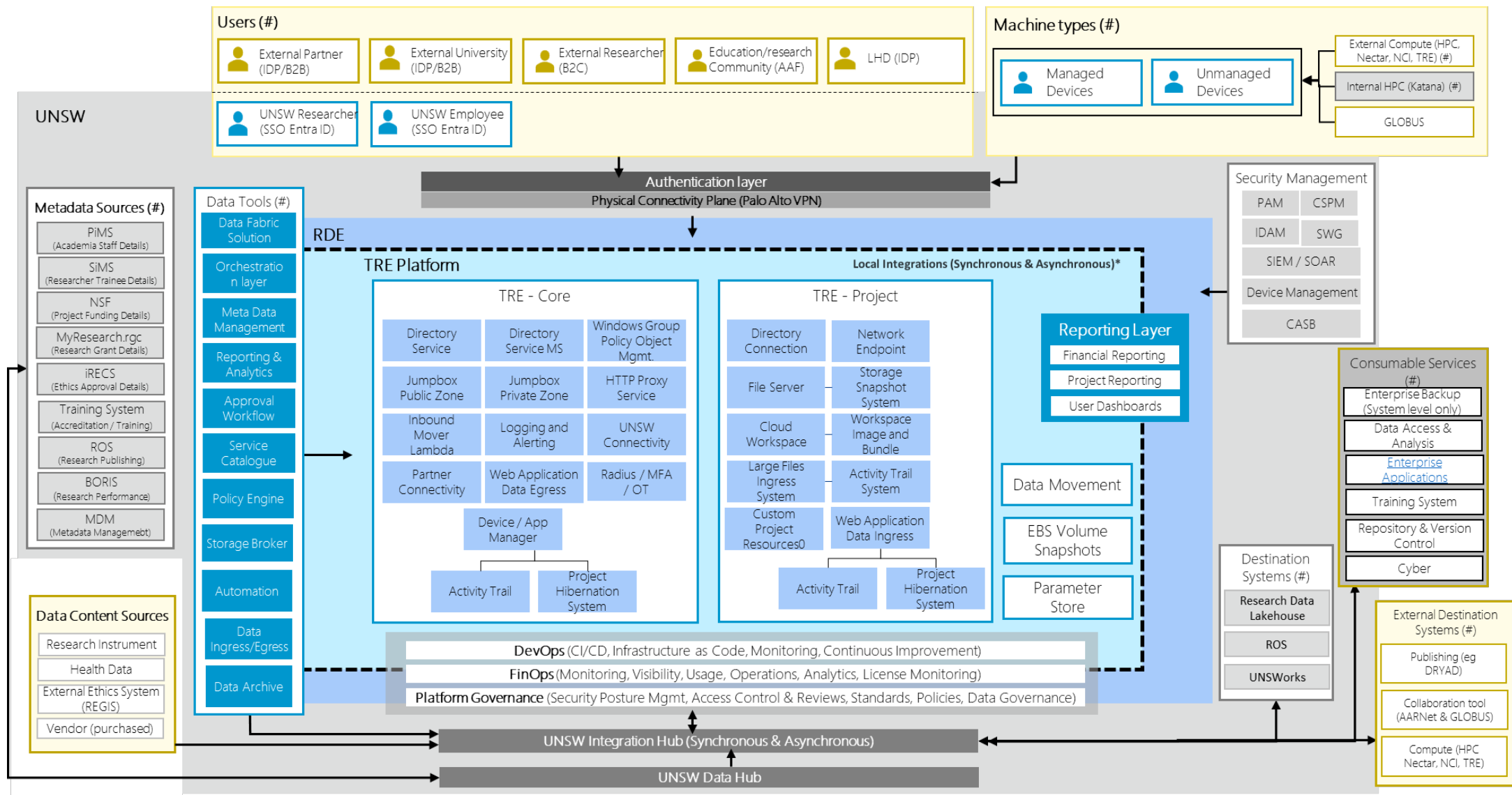
Conceptual Architecture – Storage Platform



Conceptual Architecture – Data Tools



Conceptual Architecture – TRE





Thank You



UNSW
SYDNEY

CRICOS code 00098G



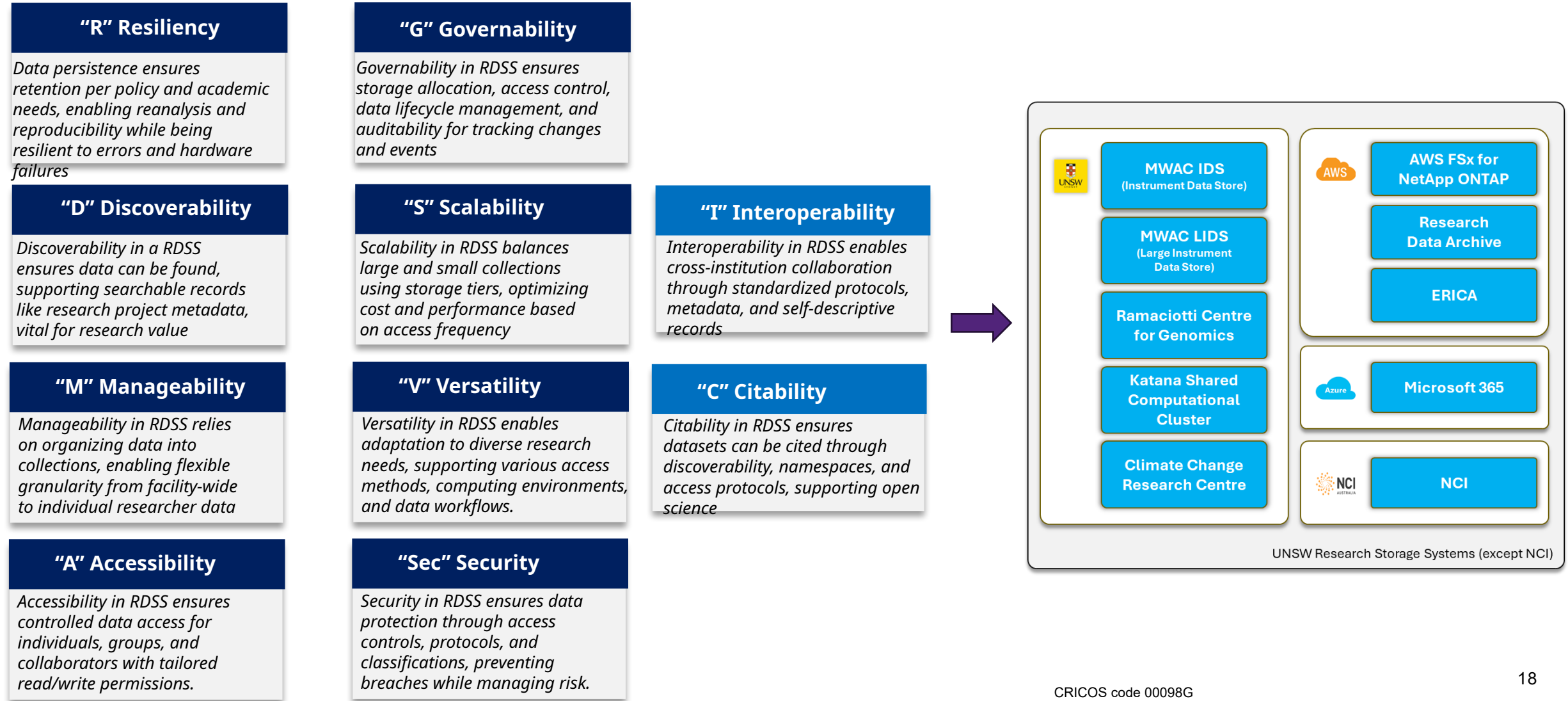
Appendix



UNSW
SYDNEY

UNSW Storage Assessment against RDRA Features

The RDRA is a high-level abstraction outlining eight RDSS features and two sub-features essential for sustainability, mapped against the UNSW storage current state.



Research Data Implementation Architecture (RDIA)

The RDIA components are mapped against UNSW's current state tools to assess alignment, identify gaps, and enhance research support infrastructure.

