Research Storage and Archiving (RSA) Project

Research Storage and Archiving Summary Report

Version Control

Version	Name	Role	Date	Summary of Changes
0.1	Anthony Normando	Author	17 Mar 2020	Initial document draft
0.1	Anthony Thomson	Reviewer	18 Mar 2020	Proposed minor amendments
0.2	Anthony Normando	Author	18 Mar 2020	Amendments incorporated
0.2	Ben Hinton Chris Gill	Quality Assurance	27 Mar 2020	Proposed minor amendments
0.3	Anthony Normando	Author	27 Mar 2020	Amendments incorporated
1.0	RSA Steering Committee	Noting	2 Apr 2020	For information.

Document Details

Document Name	Document Location
Research Storage and Archiving Summary Report	RSA Project SharePoint Site

Related Documents

Document Name	Document Location
Research Storage and Archiving User Stories	RSA Project SharePoint Site

Table of Contents

1. Introduction

1.1. Background

The ANU established the RSA Project in November 2019 to implement a storage and archiving solution that aligned to the ANU Research Community's needs. In December 2019, at the request of the RSA Steering Committee (RSA-SC), the RSA Project undertook a Discovery Phase to:

1. Validate existing Storage and Archiving User Stories (business requirements); and
2. Assess the University's overall RDM needs and recommend the highest-value capabilities the project should deliver.

1.2. Purpose

The purpose of this document is to provide a summary of the RSA User Stories developed for the project.

1.3. Intended Audience

The audience of this document is the RSA-SC.

2. RSA User Stories – Deliverable

2.1. Scope

The scope of the Summary Report is to:

- Document the University's business requirements (User Stories) for a new research storage and archiving solution – as defined by Research stakeholders (including Colleges, Research Schools, Researchers, and Executives)
- Document the high-level requirements of an enterprise storage and archiving solution (e.g. University Administrators and Corporate Services)
- Provide RSA stakeholders with a tangible record for validating the priority and broad specifications of their requirements
- Inform future deliverables associated with subsequent phases of the project, including Market Scan, Functional and Non-Functional requirements, and High Level Solution Architecture.

2.2. Out of Scope

Documenting research data management needs were considered out of scope for this report.

2.3. Engagement

The engagement approach taken to validate and understand the University's storage and archiving needs involved College and Schools representatives across all seven (7) Colleges, other Academic Units, Executives, and Information Technology Services (ITS). These activities were undertaken through workshops with Business Reference Group (BRG) and Technical Working Group (TWG), and additional representatives as provided by the representatives.

3. User Stories

A user story is a brief, plain-language explanation of a business need written from a user's point of view. The decision to leverage a User Story format means that the requirements are outcome focused and easily understood by stakeholders as the format uses business language. The table below describes the structure of a User Story accompanied by an example User Story.

Table 1: User Story Structure

Term	Description/ Definition
As a (Role), who is asking? (As a <*type of user*>)	Identifies the business users with the business need. Specifies roles such as Researcher, University Librarian, Customer Service Manager, and Administrator. Example – ***As a Researcher***
I want to <*perform some task*>	Describes the business need. Indicates what analysis is needed and what problem the business need to resolve. Example – ***I want to restrict IT Storage Administrators from viewing my data unless they have the necessary clearance or authorisation***
So that <*reason/ benefit*>	Indicates the business capabilities (analyses and actions) and business values (results) enabled by being able to meet the business need. Example – ***So that I can comply with the relevant project (e.g. ethics) requirements***
Acceptance Criteria	Acceptance criteria are statements of requirements that are described from the point of view of the user to determine when the user story is delivered and working as expected. Example – • ***Ability for a researcher to restrict IT Storage Administrators from viewing their data.*** • ***A process exists where IT Storage Administrator can gain clearance to view data (e.g. attain security clearance or ethics approval)***

4. User Roles

User Roles are a fundamental component of a User Story, as it provides a description of the *action/task* and the *reason/benefit* from the perspective of users that will use or be involved in the operation, management, and governance of the solution.

The User Roles identified in Table 1 were validated through engagement with stakeholder groups. Through our engagement, two common issues that prevented a further breakdown of roles were identified:

1. **Policy and procedure gaps.** Therefore, stakeholders were uncertain of who had responsibility, or if existing governance roles existed within the University (e.g. Data Custodians and Stewards).
2. **New technical capabilities.** These are being requested which impact people and processes that do not exist.

Given these issues, the approach taken to ensure all requirements are elicited was to create more encompassing user roles such as a “Super User”. “Super User” includes roles and responsibilities at a Research School/College level such as Chief Investigators, Data Managers and/or Owners and Storage Management and Allocation.

During subsequent project phases, roles will be further delineated and clarified through existing University governance frameworks and the development of further policies, processes and procedures.

Table 2: User Roles

User Role	Description
Researcher	A Researcher or PhD student that will utilise the Research Storage and Archiving solution for the conduct, management and / or oversight of research activities.
Super User	A College or Research School representative (e.g. Chief Investigator) / Data Manager or Owner who has additional privileges in relation to the management and allocation of storage as well as the management of data.
Research School Manager	Senior / executive-level University representatives responsible for the management of business and research within a particular College or Research School.
Director Research Services / Director Research Infrastructure	Senior / executive-level University representatives responsible for providing guidance and assistance to the Directors of University Colleges and Research Schools in the management of business and research.
Librarian	Senior / executive-level University representative/s responsible for the custodianship, curation and / or ownership of data.
IT Storage Administrator	An ITS representative or vendor service provider responsible for storage allocation, monitoring, reporting and management.
Director ITS	Senior / executive-level University representative responsible for the provision of IT policy, hardware, software and services to University users.
Chief Information Security Officer (CISO)	Senior / executive-level University representative responsible for developing and implementing IT security policies and hardware/software controls for protecting University IT systems, information and data.

5. Impact of Business Requirements

The business requirements captured in the Discovery Phase will inform the following activities throughout the next phases of the project:

1. **Market Scan and High-Level Solution Architecture (HLSA)** – The project must understand the fundamental functions and characteristics of the solution in order to conduct a market scan and develop a HLSA to provide further insight into the technical solution. These will be described as Functional and Non-Functional Requirements which will be informed by and aligned to the User Stories defined in this phase of the project.
2. **Procurement** – The development of the contracts (including schedules) will include the service and product requirements (i.e. Functional and Non-Functional Requirements) that the vendor must deliver. These requirements will be informed by and aligned to the User Stories developed in this phase.
3. **Contract & Schedules** – Development of contract schedules involves the definition of functional and non-functional requirements which will then define the goods/services that the service provider must provision and support for the life of the contract. The contract requirements defined in the next phase of the project should clearly trace back to the business requirements. Subsequently, the contract requirements will inform activities such as user acceptance testing of the solution prior to the go-live date for the solution.

6. Findings

This section describes the Findings from the Discovery Phase for the RSA Business Requirements:

- The original set of requirements (i.e. version 0.5 RSA User Stories) were not technology agnostic in a few circumstances (e.g. backup functionality)
- Requirements were ambiguous, albeit broadly accurate. Additional analysis was required to ensure the intent was clear

- Requirements were not appropriately categorised, but were largely representative of stakeholder needs.
- Phase 2 of the project will require a more encompassing and detailed perspective of roles with the University based on overarching policies and frameworks.

6.1. Capability

Access Control

College of Health and Medicine have identified a need to **ensure ITS staff are unable to view their data** stored on corporate storage in alignment with ACT Health requirements.

Access Storage

Researchers require **access to data storage from domestic and international locations**. This will be in line with the relevant Information Technology and Security policies, to ensure devices and access are managed.

Archiving

- **Retention periods can be applied** to data as there is an ongoing need to apply retention periods to data
- **Retrieval** of archived data can be **initiated by Super Users** and is supported by a process with the relevant reviews and approvals
- Archived data can be retrieved through Application Programming Interfaces (API's)
- **Metadata can be applied to data**, as well as be read and accessed by those with the necessary privileges.

Auditing

Auditing will leverage/integrate with the University's existing auditing capabilities.

Availability

Research School responses to the projects storage questionnaire has identified a broad range of **Recovery Time Objectives** (RTO)[1] ranging from hours to one month, or 5% of any given period. Research School responses encompass John Curtain School of Medical Research (JCSMR), Research School of Astronomy and Astrophysics (RSAA), Research School of Physics and Engineering (RSPE), Research School of Biology (RSB) and Research School of Psychology (RSP).

Cost Management

Development of necessary **policies, procedures, and cost models** are required to enable show back and chargeback.

Data Access

The storage ecosystem will have **various storage characteristics** such as, but not limited to, full/limited/no data protection measures, constant/intermittent data processing, and sequential read/write.

Data Destruction

Service providers will need to provide **documentary proof of data destruction** across all components of a storage solution and/or ecosystem. This is particularly important where research data that is not owned by the University is held on behalf of an external body.

[1] RTO - The tolerable time allowed to recover client systems after a disaster scenario has been declared. This is the time a business can afford to be without critical services before incurring significant losses.

Governance

- Numerous cases of **sensitive government data being created or received** were raised with the project team. Further discussions with researchers and the CISO will be required to **understand the level of security classification** to make the solution compliant. This will also further inform policy development around the University's security classification model.
- Data ownership has been discussed regularly as result of researchers having **limited clarity around data ownership**.

Data Movement

Data movement between solutions within the ecosystem and outside of the ecosystem will need to be supported by the necessary policies, processes, procedures and functionality.

Data Protection

Research School responses to the projects questionnaire has identified a broad range of **Recovery Point Objectives** (RPO)[2] ranging from less than one day to months. Research School responses encompass JCSMR, RSAA, RSPE, RSB and RSP.

Reporting

The solution will support reporting capabilities across a breadth of user groups, such as:

- Researchers – location of data, who has accessed it, storage used versus unused.
- Executives – Receive reports on storage allocation, usage (including archive), and cost by user/school/college.
- IT Storage Administrators – Produce customised reports.

Scalability

- The solution is **underpinned by a support model** that will enable it to scale up and down based on researcher needs as well as maintain agreed performance levels.
- The support model leverages existing or new processes to procure additional storage and/or storage services as required.

Storage Capacity

Derived from the Research Schools responses to the requirements survey, and historical data captured in the UICT RSA Project Proposal, the following **aggregated capacity sizes** (in Terabytes) have been identified:

- College of Law – 20* [3]
- College of Arts & Social Sciences - 1,000
- College of Asia & the Pacific – 140
- College of Business & Economics – 4
- College of Engineering & Computer Science – 600
- College of Health & Medicine - 1,465

[2] RPO – The age of a file/data that must be restored to end-user for continuation of normal business operations (e.g. Research). This may be a result of data corruption, deletion or loss.

[3] * - To be determined, noting requirements from similar college suggest this may be an upper limit need.

- College of Science - 1,970
- Other Schools - 250

The total aggregated capacity identified is **5,449 Terabytes.**

This represents the current amount of storage needed by colleges and schools that would otherwise be located on a solution provided by this project. As per Action Item 1.2, an assumed growth of 20% over the following 12 months would require the University to have approximately **6,520 Terabytes**.

Storage Deprovisioning

ITS and the relevant Super Users will be notified towards the end of a research project prompting them to assess and determine whether the storage is required for a longer period or can be deprovisioned. This will ensure **storage is not allocated forever.**

Storage Provisioning

College/School delegates with additional storage management privileges **(I.e. Super User) can manage storage allocations**. This will enable them to distribute out storage to research projects, for example – a School is allocated 10TB, which can be allocated across 6 different research projects by the delegate.

Storage Management

Capacity planning is supported by the necessary processes (e.g. integrated with investment and procurement processes) and capabilities (e.g. trend analysis).