ABC

# Jetstream - Archive Preservation Business Case

| | |
|---|---|
| **Project Title** | **Archive Preservation & Replacement** |
| **Executive Summary** | This project will replace software and infrastructure to deliver a self-service **search** interface to the content held in main cataloguing systems that make up the ABC Archive content managed by Content Services.<br><br>The project proposed will allow the ABC Archive content to be searchable by ABC content makers and will **preserve the media** that currently resides in MAS.<br><br>The system will also act as a **repository for all future digitised and digital born assets,** however the work to digitise tapes and film is explicitly excluded from this business case.<br><br>This business case requests 6 months of funding to build significant components of the new system.<br><br>Significantly, the new system will:<br>• **Preserve content from the MAS** Robotic tape system – 80% of all content will be preserved (except Sydney) in the 6-month timeframe. (Note: Sydney will be well underway and become a small operational expense after this project completes.)<br>• **Replace aging (1983) TARA and RADA databases** by extracting all metadata and allowing for content-makers and researchers to search the metadata in a new, modern system.<br>• **Allow for cataloguing** of new records to be entered straight into the new system.<br>• Move the hosting of the metadata/search system and build the media repository in the **ABC's AWS Cloud environment**.<br>• **Allow for all future digitisation** projects (e.g. tape and film) to add the media to the AWS cloud environment whereby it will immediately be available for download by content makers.<br>• **Be built on open standards** with no vendor lock-in and provide APIs for future integrations.<br>• All media will be stored in an **open format**.<br><br>Not Relevant [redacted] |

| | |
|---|---|
| **Background/ Reason for Project** | **Reason:** The current toolset used by Content Services to manage ABC Archive content is long past end of life and out of support, creating inefficiencies. Implementation of new technologies will bring about improved services for Content Makers (including self-service for users of the archives) and enable Content Services to be more efficient in their research.<br><br>**Background:**<br><br>Managing ABC archive collections involves the use of obsolete, disparate systems that are no longer fit for purpose. Content Makers have little opportunity to search for content themselves due to the complexities of navigating different databases and systems to find and retrieve ABC archive content. Furthermore, much of the current ABC archive is still largely tape based, with numerous collection and research workflows dependent on the handling of physical media, including physical carriers such as hard drives, and use of consumer grade cloud products for source/deliver/share of content.<br><br>The main archive metadata systems were built in 1983 and hardware refreshed in 1999. Content Services' digital archive 'MAS' was implemented in 2004 as a preservation system designed to store and manage file assets converted from obsolete analogue tape formats (primarily 1inch videotape and ¼ inch audiotape). MAS is a distributed system across 8 sites, with robotic tape libraries used for storage. MAS has neither the capacity or functionality to capture and archive the file-based content that the ABC produces, resulting in the utilisation of ABC Content Production systems to store archive assets. Additionally, lack of capacity in MAS has prevented the ongoing digitisation of remaining tape and film collections held in the archive, and the ingest of incoming material such has hard drives.<br><br>This project would replace the metadata systems with a single database and search system, allowing Content Makers and other ABC users to search for archive material in one place, regardless of where the content is held. This project will also see the MAS media assets migrated to cloud storage and made available from the single database/search system for download to all (or to a permission-based subset) of ABC users.<br><br>The digitisation of the remaining ABC archive content held on tape (video and audio) and film will be covered by a different business case, however the system described here will house the media and metadata for those assets. |

| | |
|---|---|
| **Objective(s) of overall project** | 1. Self-service media asset search for content makers.<br>2. Self-service media asset retrieval for content makers.<br>3. Advanced search capabilities for researchers.<br>4. Cataloguing of new archive items, the ability to edit and remove items. Upload of new archive media.<br>5. Retire TARA and RADA.<br>6. Move all content in the MAS robotic tape libraries to the new Archive.<br>7. Retire Ardome and MAS Robotic systems.<br>8. The ability for all of the other items that meet the archive criteria to be moved into the new archive. The media includes VTR tapes (approx 170,000 tapes), Production archives (DNL, Netia and TV Prod - approx 1000Tb), film and hard disks. Metadata system to be integrated include Stratus, Netia and OnAir.<br>9. Seamless integration with content production systems and automated workflow<br>10. Subject to rights, publish archival content to the public domain (e.g. YouTube).<br><br>The above objectives are listed in a draft sequence order but are subject to review by the stakeholders and the team.<br><br>**This Business Case**<br>Objectives 1-5 will form the first major deliverable of the project.<br>Objectives 6 and 7 will be well underway and become a relatively small operational piece of work once this project completes.<br><br>**Future Project Work (Not in scope)**<br>Objective 8 will create interfaces for other projects to be able to add content to the archive. A separate business case is being prepared to digitise the remaining media assets.<br>Objectives 9 and 10: The strategy for content management at the ABC is a work in progress and is a dependency to scoping this work. What is certain is that integration into our content production systems will be necessary and that we will be publishing some archive content to the public.<br>Importantly, the new system will be built to open standards utilising RESTful APIs for easier integration. |
| **Other Options Assessed** | **Do Nothing**<br>Content Services have been unable to accept new media files into the archive for the past 8 years. Content is instead being kept on temporary media (disk drives, tapes, fileshares) awaiting the archive or being kept on internal spinning disk as part of the IMS content production systems. This option will deliver none of the objectives of this business case and indeed will make a future project even harder. Without new archive storage, deteriorating film and tape cannot be digitised. |

| | |
|---|---|
| **Benefits** | Successful realisation of this project will lead to benefits that align with the "Investing in Audiences" strategy, with increased efficiency for large parts of the organisation.<br><br>– A single, unified search interface with the ability to view and select content for use will democratise access to the archive and promote content sharing across the organisation<br><br>– Greater discoverability of archival material by content makers, providing opportunities for the use of previously untapped sources of unique and distinctive content in audience facing products<br><br>– Minimises the need for archive researchers to spend time searching in multiple places for content, shifting the focus to delivery of specialist, high value services for content makers<br><br>– A cloud hosted, simplified support environment with the potential to substantially reduce maintenance costs and eliminating the need for a hardware replacement cycle.<br><br>– On the completion of Objective 8, and the separate digitisation business case, the support of video tape technologies can also be retired.<br><br>– Potential for the development of new data flows designed to simplify business processes and reduce the number of tape based & manual workflows across the ABC.<br><br>– Delivery of a centrally accessible repository will reduce content duplication throughout the ABC and potentially provide a reduced cost storage solution for production systems e.g. Digital News Libraries.<br><br>– Mitigation of the risks surrounding loss of ABC content through the migration of MAS content and data and eliminating the use of ad hoc storage solutions (hard drives)<br><br>– Open source solutions on AWS – no software vendor lock-in.<br><br>– Usable, trusted, collaborative search results<br><br>– Efficiencies will be driven as a result of this system. This platform will create the repository for future digitised content where Content Services will more effectively be able to provide content ready to be used in content production. Note the quantification of these efficiencies will require further review and be included in subsequent digitisation business cases.<br><br>– At the end of the project outlined in this business case, due to the shutdown of Tara and Rada, it is expected that there will be savings of approximately an efficiency gain o47C er annum. This will be achieved due to the termination of technology support (currently in Technology team under47F 47F |

| | |
|---|---|
| **Technical Benefits** | • Reduced support requirements<br>• MAS Robotic Systems and the use of the Ardome application deprecated<br>• TARA and RADA decommissioned<br>• No proprietary formats, no vendor lock-in<br>• AWS hosted archive search (within Cloud Shared Services)<br>• AWS hosted media files (within Cloud Shared Services) |
| **Proof of Concept Findings** | A proof of concept has been executed to demonstrate technical viability of this project. Several important outcomes have been proven:<br>• An Enterprise Data System (often referred to as a "Data Lake") can merge content from our aging metadata systems into single records.<br>• MAS files can be retrieved and restored into Cloud Storage<br>• Cloud storage of proxy files provides faster retrieval times than currently available.<br>• A web application can be built to search to service simple and advanced search.<br>• The proprietary format of video files within the MAS can be transcoded by more than one partner without the ability to transcode these files, all of the content would have to played out at 1x speed and re-encoded into a file. |

**Project Methodology**

The funding requested in this document is to support a 6-month iteration of the archive project.

The team will use agile methodologies to regularly report progress and ensure that at all times, the priority from the business can be prioritised. The following cycle is proposed:

| NAME | ATTENDEES | REASON |
|---|---|---|
| **DAILY STAND UPS** | All team members | Inform each other on the work expected that day and to report blockers. |
| **FORTNIGHTLY SPRINT REVIEW AND PLANNING** | All team members | Review work from previous fortnight and commit to work that will be carried out during the next fortnight. |
| **MONTHLY SHOWCASE** | Squad, sponsor, all interested stakeholders | Demonstrate progress to the business, allow for feedback. |
| **FORTNIGHTLY STEERING COMMITTEE** | Sponsor, senior team members. | Project progress, escalations, risks and actions. |

Not Relevant

[redacted]

| | |
|---|---|
| **Project Funding** | This project is part of Jetstream.<br>Funding is requested for the first 6 months and costs for this project are listed below.<br>Funding for period to 30 June 2018 47C was approved by the ABC Board on 29 March 2018.<br>Funding for 2018/19 will be included in Jetstream proposal for 2018/19. This will be submitted for Board approval in June 2018.<br>Estimated ongoing costs have been identified below for transparency. These will be subject to a separate business case in 2018/19 as they will be subject to strategic direction for archive digitisation as part of Jetstream. |

| | |
|---|---|
| **Assumptions** | • The consolidated data can provide meaningful search results (this has been proven in the Proof of Concept) |
| **Dependencies** | • IT infrastructure support for extracting content from the MAS<br>• IT Shared Services to provide network, shared tools and AWS usage principles<br>• IT Shared Services for access to Active Directory authentication |
| **Constraints** | • Time to extract content from MAS due to the aging hardware and networking limitations |

| | | |
|---|---|---|
| **Stakeholders Internal & External** | **SPONSOR** | 47F |
| | **PLATFORM OWNER** | |
| | **FINANCE** | |
| | **SOLUTION DESIGN AND PROJECT OWNER** | |
| | **PROJECT LEAD** | |
| | **USER REPRESENTATION** | |
| | **Project Steering Committee** | |
| | **NEWS CONTENT RESPRESENTATIVE** | 47F |
| | **ENTERTAINMENT CONTENT RESPRESENTATIVE** | Content Maker Manager (TBD) |
| | **FINANCE** | 47F |
| | **CONTENT SERVICES** | |
| | **TECHNOLOGY** | |
| | **SPONSOR** | |

## ESTIMATED COSTS/BUDGET - PROJECT

| **Financial Year Estimate** | **17/18** | **18/19** |
|---|---|---|
| Archive Preservation & Restoration | 47C | |
| **TOTAL** | | |

2018/19 Costs will be included in Jetstream budget request to Board in June 2018.

| | | |
|---|---|---|
| **FUNDING SOURCE** | **JETSTREAM** | |
| **ESTIMATED PROJECT COSTS** | | |
| **Costs for 6 month project April - Sept** | Resources (internal & external) | 47C |
| | Temporary storage necessary for MAS file extraction. | |
| | One-off Transcoding of all MAS Content | |
| | Decommissioning Costs | |
| | Travel & on costs (laptops etc) | |
| | **TOTAL Project Costs** | |

| | | |
|---|---|---|
| **FUNDING SOURCE** | **FUNDING SOURCE: JETSTREAM – NOT REQUESTED IN THIS BUSINESS CASE** | |
| **ESTIMATED ONGOING COSTS** | | |
| **Estimated annual costs from Sept/Oct 2018** | Software licences | 47C |
| | Technology Support (reduced from current) | |
| | AWS Costs<br>*(not including future digitisation storage – separate project)* | |
| | Machine Learning for all new content<br>*(15,000 hours per annum)* | |
| | Transcoding of all new video content<br>*(7,500 hours per annum)* | |
| | Transcoding of all new audio content<br>*(7,500 hours per annum)* | |
| | **ESTIMATED TOTAL Ongoing Costs (per annum)** | |

| **Risks** | **Risk** | **Impact** |
|---|---|---|
| | Current media deteriorates before this project makes a system available. | Low increasing risk, high impact. Mitigated by funding this project. |
| | Estimates are found to be too optimistic. | Medium risk, medium impact. Mitigated by regular steering committee updates. |
| | MAS Robot (particularly in Sydney) breaks during transfer. | Low chance, high impact. Cannot mitigate. If risk eventuates, cost of replacement hardware is large. |

| | |
|---|---|
| **Delivery Timeframe** | This business case requests budget for 6 months execution where the first outcome is for finished product delivery as detailed in Appendix A. |

## ENDORSED

| | Name | | Date |
|---|---|---|---|
| a/Head of Technology | [redacted] | [redacted] | 9/5/18 |
| Head of Content Services | | | 10/5/18 |
| Head of Digital Network | | | 9/5/18 |
| Finance Business Partner (Technology) | | | 9/5/18 |

## APPROVAL

| | Name | Signature | Date |
|---|---|---|---|
| Chief Digital & Information Officer | [redacted] | | 10.5.18 |

# Appendix A - Product delivery within 6 months

The first phase of this project will have 3 main components which can be started in parallel.

1. **Export and transcode of all media assets within the MAS**

   The export process has already commenced and is estimated to take a further **12** months to complete. The final destination of the media assets is Amazon S3 storage.

   The transcoding process has not been finalised but is not likely to incur any significant overhead to the delivery time. It is thought that transcoding in the cloud will provide the best value. Three vendors have demonstrated an ability to perform the custom transcode of MAS proprietary encoding and the project would find the best value for money whilst ensuring the quality we need.

2. **Create a data lake for the ABC Archive**

   To achieve this a standardised Archive Asset Record structure will need to be formalised. All existing data from the TARA, RADA, Trace and MAS Ardendo databases will be consolidated into the data lake using the new structure, informed by ABC Core.

   This process will result in a single record per asset from the identified systems along with links back to original data records. Links to the original data are only for providence purposes.

   Each record will provide a link to the associated media files, or physical assets.

   All the original data from the original databases will be kept within the data lake however they will not be directly accessed, they will be kept for preservation purposes, available as needed.

   This component of the work is to be undertaken by the Project Team in conjunction with subject matter experts from content services.

3. **Development of a web-based application to oversee all aspects of the archive.**

   This component is the tool which will realise the core deliverables from the business case, namely:

   Self-service media asset search for content makers.
   Self-service media asset retrieval for content makers.
   Advanced search capabilities for researchers.
   Cataloguing of new archive items.
   Management of existing archive items.
   Upload of new archive media.

   During this phase of the project, this system will be demonstrable, however it will not yet be in production use.

   When this component is completed and the data lake are completed the TARA and RADA databases can be decommissioned and Trace can be depreciated in function. Trace will still be used for managing tape assets until the digitisation project finishes.

   When the asset migration and transcode is completed the MAS Ardendo system can be decommissioned and the LTO storage disposed of.

   This component of work is to be undertaken by the Project Team.

| | |
|---|---|
| **Stakeholders**<br>**Internal & External** | **SPONSOR**<br>**PLATFORM OWNER**<br>**FINANCE**<br>**SOLUTION DESIGN AND PROJECT OWNER**<br>**PROJECT LEAD**<br>**USER REPRESENTATION**<br>47F |
| | **Project Steering Committee** |
| | **NEWS CONTENT RESPRESENTATIVE**<br>47F |
| | **ENTERTAINMENT CONTENT RESPRESENTATIVE**<br>Content Maker Manager (TBD) |
| | **FINANCE**<br>**CONTENT SERVICES**<br>**TECHNOLOGY**<br>**SPONSOR**<br>47F |

**ESTIMATED COSTS/BUDGET - PROJECT**

| **Financial Year Estimate** | **17/18** | **18/19** |
|---|---|---|
| Archive Preservation & Restoration | 47C | |
| **TOTAL** | | |

2018/19 Costs will be included in Jetstream budget request to Board in June 2018.

| ESTIMATED PROJECT COSTS: | **FUNDING SOURCE: JETSTREAM** | |
|---|---|---|
| **COSTS FOR 6 MONTH PROJECT APRIL - SEPT** | Resources (internal & external) | 47C |
| | Temporary storage necessary for MAS file extraction. | |
| | One-off Transcoding of all MAS Content | |
| | Decommissioning Costs | |
| | Travel & on costs (laptops etc) | |
| | **TOTAL Project Costs:** | |

| ESTIMATED ONGOING COSTS: | FUNDING SOURCE: JETSTREAM – NOT REQUESTED IN THIS BUSINESS CASE | |
|---|---|---|
| ESTIMATED ANNUAL COSTS FROM SEPT/OCT 2018 | Software licences<br>Technology Support (reduced from current)<br>AWS Costs (not including future digitisation storage – separate project)<br><br>Machine Learning for all new content (15,000 hours per annum)<br>Transcoding of all new video content (7,500 hours per annum)<br>Transcoding of all new audio content (7,500 hours per annum)<br>**ESTIMATED TOTAL Ongoing Costs (per annum):** | 47C [redacted] |

| Risks | Risk | Impact |
|---|---|---|
| | Current media deteriorates before this project makes a system available. | Low increasing risk, high impact. Mitigated by funding this project. |
| | Estimates are found to be too optimistic. | Medium risk, medium impact. Mitigated by regular steering committee updates. |
| | MAS Robot (particularly in Sydney) breaks during transfer. | Low chance, high impact. Cannot mitigate. If risk eventuates, cost of replacement hardware is large. |

| | |
|---|---|
| **Delivery Timeframe** | This business case requests budget for 6 months execution where the first outcome is for finished product delivery as detailed in Appendix A. |