



FINAL REPORT DMQ3 WORK PACKAGE

1. Data Quality from Instruments and Sensors

2. Gugu Badhun Digital History Project

Version 1.0, May 2007

Investigators:

Dianna Hardy, Matthew Morgan

Prepared by:

Dianna Hardy dianna.hardy@jcu.edu.au

Matthew Morgan matthew.morgan@jcu.edu.au

Franz Eilert frank.eilert@jcu.edu.au

School of Maths, Physics and Information Technology
Faculty of Science, Engineering and Information Technology
James Cook University

Executive Summary

This report presents the findings of research undertaken as a part of the DART (Dataset Acquisition, Accessibility and Annotation e-Research Technologies) project, specifically for Work Package DMQ3, ensuring data quality from instruments and sensors and developing a digital history demonstrator. The DMQ3 Work Package was modified, with board approval, following delays in obtaining live data from the Davies Reef deployment of sensor networks and the loss of our primary research resource from Monash University to Sydney University.

Despite the challenges presented during the work package, this work package is has produced two clearly important deliverables.

Firstly, a thorough review of leading edge data quality and associated software algorithms by Pyramid Intelligence systems resulted in the development of pilot software (java). The software was developed to implement quality assurance algorithms and test their efficiency and execution times against synthetic data streams. The models can be easily incorporated into Kelper actors, have been used in the JAINIS workflow environment to provide real-time data integrity.

Secondly, the creation of a simple digital video annotation tool that enabled the annotation of indigenous history videos via the internet. Videos detailing the history of the Gugu Badhun had been recorded with the help of the JCU School of Indigenous Australian Studies (SIAS) staff. The goal was to transfer the videos to an appropriate storage facility, providing annotations with additional information regarding the context of the videos, and making the videos available for display via the Internet. This provided the ideal scenario for a digital history demonstrator.

This work package has demonstrated that through agile management of the project at JCU, the work package delivered two significant outcomes.

Table of Contents

1	INTRODUCTION	4
1.1	ENSURING DATA QUALITY FROM INSTRUMENTS AND SENSORS	4
1.2	THE GUGU BADHUN DIGITAL HISTORY DEMONSTRATOR	4
1.3	ORGANISATION OF THIS REPORT	5
2	PROJECT MILESTONES	6
2.1	ENSURING DATA QUALITY FROM INSTRUMENTS AND SENSORS	6
2.2	DIGITAL HISTORY DEMONSTRATOR	6
2.3	PREPARE VIDEOS FOR UPLOAD TO A STORAGE CENTRE	7
2.4	UPLOAD VIDEOS TO STORAGE CENTRE	7
2.5	ASSIST GUGU BADHUN STAFF TO PREPARE ANNOTATIONS	7
2.6	DEVELOP A VIDEO ANNOTATION TOOL	7
2.7	PROVIDE USER DOCUMENTATION AND TRAINING ON THE NEW SYSTEM	8
2.8	USE THE TOOL TO ADD ANNOTATIONS TO SRB	8
2.9	CREATE A WEBSITE TO HOST THE GUGU BADHUN DIGITAL HISTORY VIDEOS	8
3	PROJECT OUTCOMES	9
3.1	ARCHITECTURE	9
4	ARCHIVAL STORAGE OF PROJECT DELIVERABLES	10
5	RECOMMENDATIONS	10
5.1	ENSURING DATA QUALITY FROM INSTRUMENTS AND SENSORS	10
5.2	DIGITAL HISTORY DEMONSTRATOR	10
6	PUBLICATIONS	10
7	TERMS OF REFERENCE	10
7.1	GLOSSARY	10
8	REPORT SIGNOFF	11

1 Introduction

The original DMQ3 work package project goal was to ‘Ensure that data from instruments and sensors is of sufficient quality to warrant curation.’ The work package suffered a number of delays due to the lack of data and our researcher Mohammed Medhat Gerber moving to Sydney University. During this time a requirement for digital history demonstrator as part of the DART deliverables was highlighted. JCU staff had previously worked with the School of Indigenous Australian Studies (SIAS) who had a requirement for a lightweight, simple, web based annotation system.

The DMQ3 project goal was modified, with board approval, to reflect an overall commitment to quality in the JCU work packages and the development of a simple open source video annotation system:

DMQ3 - Ensure project deliverables provide the level of quality to meet client expectations, including:

- Data from instruments and sensors is of sufficient quality for curation, and
- Tools and demonstrators provide useful, intuitive and reliable interfaces.

This is reflected in the rest of this report.

1.1 Ensuring Data Quality from Instruments and Sensors

The original primary objective of DMQ3 was to ensure data quality of instruments and specifically sensor networks utilising statistical tools. Delays in deploying sensor networks to Davies Reef (DMQ1) resulting from the failure of microwave hardware and associated weather delays (noting that 2006 was the windiest year on record in the Coral Sea), meant that live data was not available for testing. Also the loss of our research resource at Monash University provided additional challenges. For these reasons a set of simulations were created to test the outcomes of this work package.

1.2 The Gugu Badhun Digital History Demonstrator

As part of the DART project, three demonstrators were created to provide ‘proof of concept’. The demonstrators utilise the tools created in the research. The demonstrators related to digital history were:

- Western Cape demonstrator – University of Queensland
- Women on Farms demonstrator – Monash University
- Gugu Badhun Digital History demonstrator – James Cook University

This project report describes the results obtained from the creation of the Gugu Badhun Digital History demonstrator. The other two demonstrators were not attempted or completed due to the complexity or legal issues regarding the use of the software.

The reason the Gugu Badhun project was chosen was because prior to the establishment of the DART project, members of the School of Indigenous Australia Studies (SIAS) at JCU commenced a digital history project. The stated goal of the project was to ‘record and present the life histories of elders of the Gugu Badhun people and those of non-Indigenous people for whom the upper Burdekin region is a land of long-standing emotional and cultural significance.’ The digital videos detailing the history of the Gugu Badhun had been recorded, with the goal of transferring them to an appropriate storage facility, providing annotations with additional information regarding the context of the videos, and making the videos available for display via the Internet. This provided the ideal scenario for a digital history demonstrator. Initially there was some initial resistance due to a previous experience with a University developed software. However this was overcome by providing an understanding of the support requirements.

1.3 Organisation of this report

The purpose of this report is to detail the work undertaken in relation to the development of the tools for ensuring data quality from instruments and sensors, development of the digital history demonstrator and to present recommendations developed on the basis of that work.

The report covers two distinct areas:

1. Ensuring Data Quality from Instruments and Sensors.

- Requirements for development of data quality tools for instruments and sensor networks; and
- Discussion of the investigative process for data quality.

2. Digital History Demonstrator

- The background, motivation and Use Cases;
- Architecture, development process and deployment of software; diagrams of the software; and
- Feedback from the Gugu Badhun project staff.

Achievement of the project milestones is discussed in the next section.

2 Project Milestones

2.1 Ensuring Data Quality from Instruments and Sensors

The ensuring data quality component of this work package was delayed following the loss of the Monash resource and the delays in receiving live data from Davies Reef. Simulations were created and Pyramid Intelligence Systems was engaged to develop a set of algorithms for testing the simulation data and subsequent live data streams. These statistical models were based on theoretical statistical models utilised in engineering applications. It included tools similar to control charts, and pre-prediction of sensor failure, calibration drift etc. Pilot software (java) was developed to implement these algorithms and test their efficiency and execution times against synthetic data streams. The models are easily incorporated into Kelper actors and hence used in the JAINIS workflow environment and provide real-time data integrity.

The real time analysis and quality assurance of sensor data is of particular interest to the Great Barrier Reef Ocean Observing System (GBR-OOS) which is a component investment of NCRIS 5.12 (IMOS). An AIMS instrument technician has been co-located with the JCU-DART group from Q4 2006, and has been providing input into the sensor QA requirements of GBR-OOS to the DMQ3 group. GBR-OOS are keen to implement this software when they deploy sensors on the Great Barrier Reef.

Copies of the quality control investigation paper is attached.

2.2 Digital History Demonstrator

In late 2006, the Gugu Badhun project was chosen as a suitable case study for a digital history demonstrator for DART. The initial objectives defined for the project were as follows:

- Prepare videos for upload to a storage centre (SRB)
- Upload videos to storage centre
- Assist Gugu Badhun staff to prepare annotations
- Develop a video annotation tool (Mattotea)
- Use the tool to add annotations to SRB
- Create a website to host the Gugu Badhun digital history videos
- Provide user documentation and training on the new system

The Chief Investigator for the project was Ian Atkinson, with project management provided by Frank Eilert. The software was developed by Matthew Morgan, and assistance with interface design and user training was provided by Dianna Hardy. Work on the demonstrator began in December 2006 and was completed in May 2007. In addition to this Final Report, staff working on this demonstrator contributed information on interim findings to a number of other DART activities such as the Portal Workshop held at Monash University in August 2006. The objectives are detailed below:

2.3 Prepare videos for upload to a storage centre

The Gugu Badhun staff digitally recorded 20 separate interviews with Indigenous and non-indigenous people associated with the Valley of Lagoons cattle station near Greenvale, Queensland. Most of the interviews were of 2-3 hours in length. These videos were captured on a digital video camera which stores the film in DV (digital video) format. Upon return to Townsville, these videos were provided to the DART project for upload into SRB.

After further discussions with the GBT (Gugu Badhun Team), due to the length of these videos, the team decided that they would like to create smaller 10 minute clips which could be uploaded rather than the entire interview. Seventeen sample clips were chosen, and extracted from the hour-long video tapes.

2.4 Upload videos to storage centre

Seventeen sample clips were formatted in FLV in preparation of their upload into SRB. The digital demonstrator requires that the videos be in formatted as FLV in order to use a Flash video player. Open source software *Super* (<http://www.erightssoft.com/SUPER.html>) and *FLV MetaData Injector* (<http://www.buraks.com/flvmdi/>) were used to format the videos from DV into FLV.

Once the videos were formatted correctly, the files were uploaded to SRB. At this point the original file name was added to the metadata for identification purposes.

2.5 Assist Gugu Badhun staff to prepare annotations

During this stage of the research, assistance was provided to the GBT in order for them to determine the annotations that should be added to each clip. For each clip, users can add metadata and annotations.

The metadata fields available are:

- Title
- Description
- Date of interview
- Interviewer
- Interviewee
- Location of interview

In addition, multiple annotations can be added at any location in the video clip. The DART team assisted the GBT to determine what sort of annotations would be needed for each clip.

2.6 Develop a video annotation tool

During the time that the annotations were being developed by the GBT, the DART team created the video annotation tool. An iterative, agile software development process was followed, based on a set of use cases established after in-depth discussion with the GBT.

Approval of a web mockup of the system was given by the GBT in January 2007, and final development was completed in May 2007.

2.7 Provide user documentation and training on the new system

User documentation on the new system was created both in an on-line (context sensitive help, embedded in the system) and off-line (Word document) formats. The DART team provided training for a member of the GBT to enable him to enter the annotations.

2.8 Use the tool to add annotations to SRB

As part of the user testing process, the GBT entered metadata and annotations for each video clip using the digital demonstrator.

2.9 Create a website to host the Gugu Badhun digital history videos

In the final hand-over of the system to the GBT, the DART team created a website for the GBT which includes the software for the digital demonstrator. The site allows the GBT to upload new videos, add metadata and annotations for existing videos, and has space for additional documentation such as transcriptions of the video clips, and background information on the Gugu Badhun people.

This site is being hosted by the HPC (High Performance Computing) unit of JCU, and this group will offer ongoing storage and technical support for this application and content.

3 Project Outcomes

The DART project has created a digital demonstrator with the following functionality:

- Allow digital videos to be uploaded to a secure SRB storage centre
- Allow metadata to be attached to video files
- Allow annotations to be attached to videos at multiple time locations in the file
- Provide a search engine which is capable of searching both metadata and annotations stored in SRB for each video
- Provide a set of links describing videos which meet the search criteria

The Gugu Badhun project was chosen as a test case for this demonstrator. A production website has been set up for the GBT, and the users have been trained in its use and maintenance.

3.1 Architecture

The accepted web page mock-up was transferred to a JSR168 portlet to be run under the Gridsphere Version 2 environment. Functionality was then added to allow the portlet to access the videos on the SRB instance. AJAX was also used so that the web page does not reload each time the user makes changes to the metadata of the video.

Gridsphere is used as the base for delivery of the video annotation tool. Gridsphere provides two levels of access to the video annotation tool's portlet. Prior to login users can view the video, description metadata and annotations. Once logged in the user is able to update the description metadata and annotations.

The system was designed so that the view of the data is separate from the management of the data. This allows the view of the portlet to be easily updated should the need arise.

Software managers were created within the application to interact with the view and the SRB instance. The manager is able to pass information to the view through the use of session variables.

The view accesses 5 managers:

1. Menu Manager – returns the file structure of the SRB instance (containing the videos) to create the menu of video clips. Each video file is also accompanied by it's file size.
2. Video Manager – caches the required video on the server for fast access
3. Description Metadata Manager – accesses and edits the metadata of the loaded video file.
4. Annotation View Manager – accesses the annotation metadata of the loaded video file.
5. Annotation Edit Manager – edits the annotation metadata of the loaded video file.

4 Archival Storage of Project Deliverables

The software produced by this work package, the initial version of Mattotea will be created on a Virtual Machine (VM), zipped up and shipped to Monash University for archival storage with the other software artefacts from the DART Project.

5 Recommendations

Both components of this work package have provided successful outcomes. The recommendations for further development of each of the components is discussed below.

5.1 Ensuring Data Quality from Instruments and Sensors

The results from the investigation into data quality have been implemented in the DMQ5 work package in a very simple form. The components will be implemented in a more comprehensive form in the ARCHER project in support of data captured both on the Great Barrier Reef and from various scientific instruments.

5.2 Digital History Demonstrator

The digital history demonstrator will be available for use of the Gugu Badhun staff and may be utilised in other projects. There are no funding or plans to further develop this product. There are a number of similar products that are available either commercially or being developed for specific business models. The tool provides a solid foundation for video annotation via the web and may be further developed if a business case and funding becomes available.

6 Publications

Nil publications were created from this work.

7 Terms of Reference

7.1 Glossary

Acronym	Definition
AJAX	Asynchronous JavaScript and XML
AIMS	Australian Institute for Marine Science
FLV	Flash Video Format (File Extension)
GBR-OOS	Great Barrier Reef Ocean Observing System
GBT	Gugu Badhun Team
HPC	High Performance Computing
IMOS	Integrated Marine Observing Systems
JSR	Java Specification Request
SIAS	School of Indigenous Australian Studies
SRB	Storage Resource Broker

8 Report Signoff

It is agreed between

Franz Eilert, Matthew Morgan and Dianna Hardy

and

Associate Professor Ian Atkinson

and

Dr Andrew Treloar

That the **Final Report Document** for the DART DMQ3 – ‘Ensure that data from instruments and sensors is of sufficient quality to warrant curation’, gives a full account of the work undertaken for the DART Project.

Franz Eilert	0401 710 272
Matthew Morgan	07 4781 6730
Dianna Hardy	07 4781 6921

- has been read and reviewed by all parties,
- shows that the DART DMQ3 – ‘Ensure that data from instruments and sensors is of sufficient quality to warrant curation’ has been completed satisfactorily,
- clearly outlines the deliverables stated in the DMQ3 requirements documentation have been met.

Dated this 25th day of May 2007

Signed
Chief Investigator
Associate Professor
Ian Atkinson

Signed
For and on behalf of DART
Project Director
Andrew Treloar