

# A Data Laboratory for Health Services and Population Health Research

ROBYN REBOLLO<sup>1</sup>, KIM KEOGH<sup>2</sup>, MALCOLM WOLSKI<sup>3</sup>,

<sup>1</sup>Griffith University, Brisbane, Australia, [r.rebollo@griffith.edu.au](mailto:r.rebollo@griffith.edu.au)

<sup>2</sup>Griffith University, Brisbane, Australia, [k.keogh@griffith.edu.au](mailto:k.keogh@griffith.edu.au)

<sup>3</sup>Griffith University, Brisbane, Australia, [m.wolski@griffith.edu.au](mailto:m.wolski@griffith.edu.au)

## INTRODUCTION

In October 2008, during an expansion phase staff from the Research Centre for Clinical and Community Practice Innovation (RCCCPI) came to the conclusion they needed to develop better data management practices for long term preservation and re-use of research data. The Centre has over fifty researchers involved in Centre activities spread over three Griffith campuses. In addition, there are ongoing collaborations with staff from a number of other Universities in Australia and overseas as well as with staff in the QLD Department of Health.

As a result of a previous activity involving the eResearch Support Services unit in the Division of Information Services, they asked for some assistance in submitting a grant application for infrastructure funding. This resulted in a successful infrastructure grant to develop the Health Outcomes Laboratory for Data (HOLD): A Data Laboratory for Health Services and Population Health Research. This presentation describes the processes, tools and technologies used to build this solution.

## PROJECT OBJECTIVE

The purpose of the HOLD data warehouse is to:

- Link datasets in a meaningful and useable way to increase research capacity
- Provide quality assurance through feedback to data providers and researchers
- Provide a coordinated single point of access to health-related data
- Facilitate population and health services research based on sound data
- Facilitate collaboration between health, social and urban planning researchers
- Facilitate collaboration with partner investigators and policy-makers

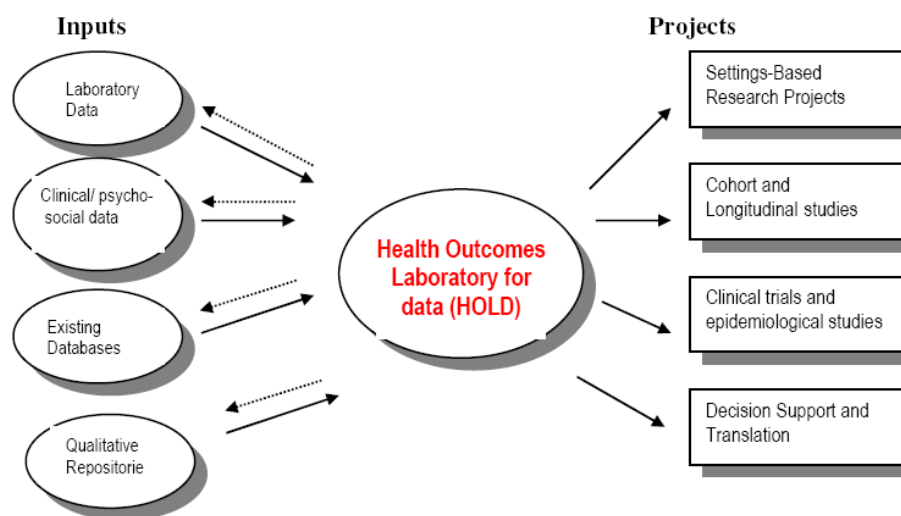


Figure 1 Conceptual Model of HOLD

## UNDERLYING STANDARDS AND ACCESS MANAGEMENT

In early 2009, eResearch Support Services staff met with the centre staff to start developing the infrastructure. To build the infrastructure and processes required, much effort was put into developing and getting agreement on underlying information workflows, taxonomies and controlled vocabularies and metadata for information lifecycle management. The Australian Standard Research Classifications & National Health and Medical Research Council Keyword / Phrases (<http://www.nhmrc.gov.au/grants/policy/keywords.htm>) were used for a taxonomy and controlled vocabulary. The metadata collected for published output in particular will enable later retrieval and cross referencing between object assets within the repository. Objects in the repository are stored in agreed open standards formats (e.g. documents, audio, video, GIS data sets etc.) suitable for later re-use.

Access management has been designed for the whole research lifecycle to ensure it is consistent and appropriate for the purpose considering the sensitivity of the data collected and to meet any legislative requirements.

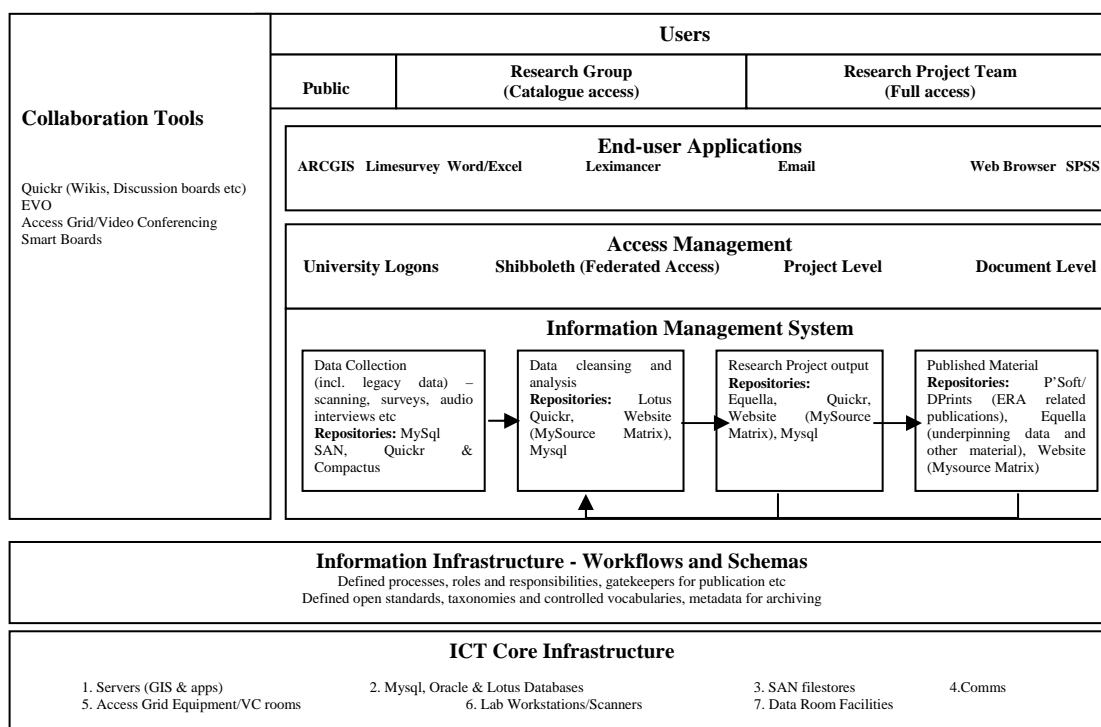
## WORKFLOW ESTABLISHMENT

Formal recommendations for facilitating the workflow of pre-published and published electronic content produced by RCCCPI and affiliated researchers were established. The pre published workflow enables researchers to collaborate more effectively, re-use research data and migrate their research data to a published domain environment. The standard workflow for published research data maintains the integrity of content and core metadata associated with the RCCCPI digital repository.

## DATA CAPTURE

All audio recording equipment used for interviews was replaced with equipment that could store mp3 format audio files, record higher quality audio and longer sessions. The hand recorders also came with a phone adapter so they could be used for telephone recorded interviews.

An open source software product (Limesurvey) is used to deliver online surveys as well as producing paper surveys. Online and hand recorded results are uploaded using the same process and software. The author of an online survey is required to enter appropriate metadata and other relevant information (e.g. Project ID, ethics approval number etc) to create the survey. The survey itself goes through a controlled publication process. Paper based surveys can also be scanned into a repository if ethic approval requires a copy to be kept as well as the data.



**Figure 2: Core Infrastructure network for the Griffith University HOLD data laboratory**

## COLLABORATION AND PRE-PUBLICATION

A Lotus product, Quickr, was used to develop an online collaboration and sharing environment for RCCCPI researchers which has: a) discussion board for posting announcements on recent research updates, training options, program opportunities, etc b) shared document libraries that provide document management features such as check-in and checkout, drafts and access control, c) a wiki space for collaborating on project lists, checklists, event planning and team goals, d) ability to add external collaborators, and d) access points to databases and current awareness resources (e-journals) for research.

An online ARCGIS environment was developed to host GIS models and appropriate hardware and software was purchased for each campus laboratory. Video conference equipment and the EVO desktop solution has been acquired to assist in cross campus and inter-University collaboration. Interactive Whiteboards have been acquired and set up with videoconferencing equipment to allow for more productive collaboration.

## PUBLICATION OF OUTPUTS

Research outputs consist of journal articles, data sets, documents, content for web page (restricted and public sites). All published content goes through a controlled process to ensure quality and other requirements are met e.g. copyright or legal requirements and then stored in a long term digital repository. A software product, called EQUELLA, is used for this purpose.