

A Web 3.0 Approach to Building an Online Digital Archive of Architectural Practice in Post-War Queensland

Jane Hunter¹, John Macarthur¹, Deborah van der Plaats¹, Andrew Wilson¹, Andrae Muys¹,
Gavin Bannerman²

¹The University of Queensland, ²The State Library of Queensland
Brisbane, Queensland, Australia

ABSTRACT

The “Architectural Practice in Post-War Queensland: Building and Interpreting an Oral History Archive” project is a collaboration between the University of Queensland, the State Library of Queensland and four of the longest-standing architectural firms in Queensland. The aim of this project is to build a comprehensive multimedia digital archive that documents architectural practice in post-war Queensland (1945-1975) – a period of architectural practice that was highly significant but is largely undocumented. Currently most of the knowledge of this period is in individual’s memories, in private hands, or highly dispersed across architectural firms and collecting institutions (State Library of Queensland, John Oxley Library and Fryer Library).

This paper describes our innovative approach (and the Semantic Web and Web 3.0 technologies that we are adopting) to building a comprehensive online and evolving knowledge-base for research, teaching and practice within the disciplines of history, architecture and design. We also describe the challenges that this project faces including ensuring the archive’s sustainability, resolving issues of identity and implementing quality control over the community-generated content.

1. INTRODUCTION

The “Architectural Practice in Post-War Queensland: Building and Interpreting an Oral History Archive” project is a collaboration between the University of Queensland, the State Library of Queensland and four of the longest-standing architectural firms in Queensland (BVN Architecture, Conrad and Gargett, Wilson Architects and Robert Riddel Architecture). In 2011, this project received three years funding from the ARC Linkage program to record the stories of Queensland’s post-war architects and to use them as the foundation for a digital archive. The online repository is designed to expand over time through community contributions - to generate a comprehensive online resource documenting the individual, firms and buildings that contributed to Queensland’s architectural identity in the post-World War II period (1945-1975).

In this paper we describe the project’s overall objectives and the methodology that we are using to collaboratively build and populate the archive. We describe the key technical components and the existing tools/services that we are planning to integrate, extend and tailor for our user community. We also provide some examples of user interfaces that we are developing to enable scholars and researchers to analyse the content and contribute new derived and aggregated content that they generate through our e-research tools. Finally we discuss the key challenges that this archival project faces and suggest approaches for overcoming these challenges.

2. OBJECTIVES

The aims of this project are to:

- document the oral histories of a generation of architects who studied and worked in Queensland from 1945-1975 (and who are now aged in their eighties and nineties),
- gather these into a single online multimedia archive producing a new knowledgebase of Queensland architecture and design,
- supplement the oral histories with additional related content retrieved from related cultural collections (libraries and archives), architectural firms or generated by researchers in the field (e.g., Wilson and Macarthur’s study of the architect James Birrell [5], Wilson’s edited catalogue of the postwar houses of Hayes and Scott [6], Ian Sinnamon’s essay on the work of architect Karl Langer [7], Watson & McKay’s *Directory of Queensland Architects to 1940* [8] and the respective reports by Fiona Gardiner [9] and Robert Riddel [10] on *Significant Queensland 20th Century Architecture*)

Brisbane, Australia

20-24 August 2012
International Council on Archives Congress

- use innovative semantic web technologies to link tacit knowledge extracted from individual oral histories to tangible knowledge (drawings, books, photographs, manuscripts) that exist within personal and firm archives as well as State and institutional archives and libraries (such as the State Library of Queensland and the Fryer Library).

The combined use of ontologies (describing the people, places, buildings, events, firms of post-war Qld architecture) and semantic tagging and inferencing tools enable us to document and visualize relationships and connections between individual oral histories and tangible collections. Unifying currently dispersed knowledge and resources, both tangible and intangible, will hopefully reveal deeper understandings of the era, its people and the architecture produced. The use of Semantic Web technologies will also offer new opportunities to query large volumes of data (the employment and client records of Queensland's architectural firms) for the first time, and in doing so, to capture for the social networks that have informed Queensland architecture throughout the post-war period. Finally, by documenting a period in Queensland's history that significantly informed the current identity of Queensland architecture, the project will also generate new knowledge and lead to a better understanding of regional design and inform current debates concerned with sustainability and climate change.

3. METHODOLOGY

The first stage of the project involves conducting and recording a series of oral history interviews and public forums with the key architects from this period. These events will comprise both **private interviews**, one-on-one conversations between the project co-ordinators and architect/s as well as a number of larger **public forums** to be held at the State Library of Queensland. The public forums will be dedicated to a specific theme (education, style, climate, region, etc.) and the conversation will be directed at a selected panel of guests but chaired by eminent local architectural historians (for example, Don Watson [8] and Rob Riddel [10]).

Both the oral history interviews and the public forums are filmed, captured as digital files (.wav and .avi) and transcribed. These digital stories form the foundation of the archive. In addition, downloadable webcasts of the filmed interviews and/or transcripts are available from the Website¹.

In parallel with producing the oral history interviews, we will be collecting digitized content (from architectural firms, Australian Institute of Architects (AIA) and researchers in this field), uploading it into a local content management system and indexing it using semantically rich metadata (based on an ontology and controlled vocabularies).

Next we apply semantic tagging software to the transcripts of the interviews to identify key entities (people names, places, dates, events, firms, buildings) that are mentioned in the interviews. The semantic tags will be drawn from an *ontology of Queensland architecture* (that defines key entities (people, firms, buildings, documents, stories, events), relationships and instances) that underpins the knowledgebase and enables the coherent integration and organization of the heterogeneous and distributed content.

The semantic tags attached to interview/oral history segments, can be used both to expedite the search and discovery of relevant oral history segments but can also trigger the search and retrieval of related resources (such as images, drawings or manuscripts) from both the local Content Management System (CMS) as well as from related online collections such as the NLA's Trove²/Picture Australia or Queensland Heritage Register³ – thus allowing us to build compound objects that aggregate all of the distributed but related resources about a particular architect or building.

For example, the interview on Karl Langer will be linked to the collection of his architectural drawings and letters held by the State Library, photographs of his buildings in Picture Queensland/Picture Australia, full-text scans of his books or excerpts from his books (e.g., Sub-tropical Housing [13]) and contemporary critiques of his work.

Related resources (or segments) will be encapsulated in compound objects or resource maps based on the OAI-ORE protocol [11]. Users will be able to search, discover, display, edit and extend these compound objects through novel graphical web interfaces that enable them to document new connections as they discover or infer them.

To generate community interest in the project, an online survey on the project Web site will invite architects who studied and worked in Queensland to submit their own details including a chronology of practice and to provide feedback to the existing content. An additional blog monitored by the project leaders will encourage the broader community (those outside the profession) to comment on aspects of post-war architecture (e.g., nominate their favourite building) and to upload related materials such as photographs or plans.

¹ <http://qldarch.net>

² <http://trove.nla.gov.au>

³ <https://www.derm.qld.gov.au/chimsi/basicSearch.html>

4. TECHNICAL ARCHITECTURE

Figure 1 provides an overview of the architectural components of the archive. The system itself comprises four main components:

- Web-based User Interface
- Semantic and Presentation Services
- Ontology and Controlled Vocabularies
- Data Management.

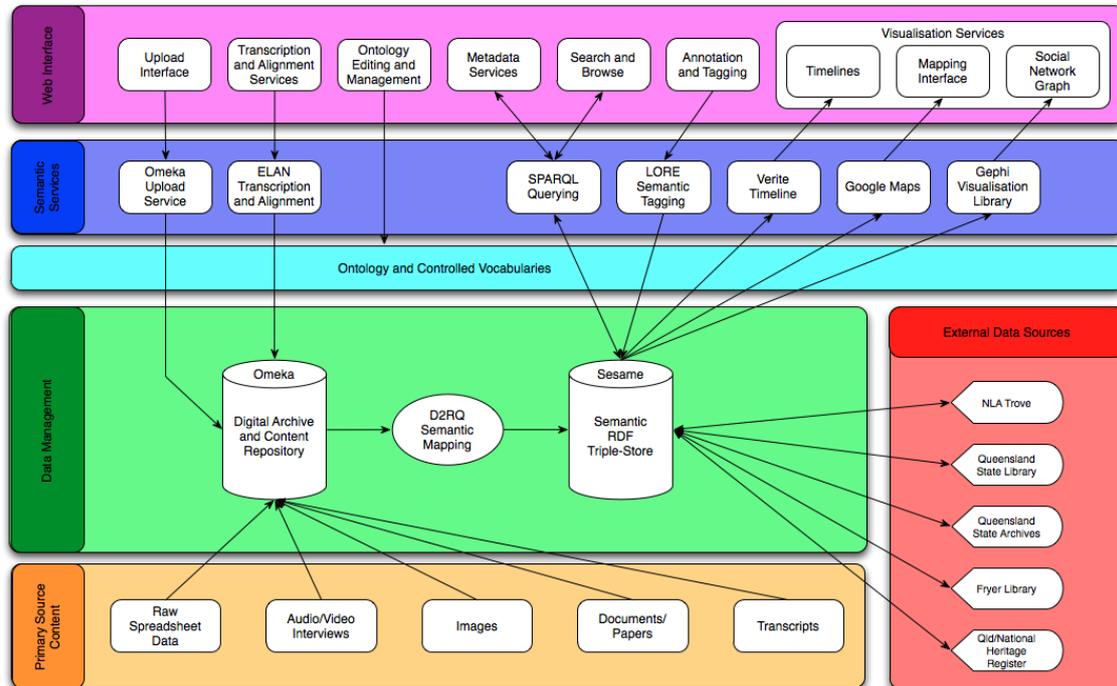


Figure 1: Overall Technical Architecture for the Post-War Qld Architecture Digital Archive

The Web Portal will provide a menu that includes general information (background, collaborators, contacts etc.) about the project but also provides access to the underlying content and services. More specifically it will provide authenticated users (project participants) with interfaces for: uploading content into the content management system; replaying interviews and aligning transcripts with the audio/video files; reviewing and editing the ontology and controlled vocabularies; attaching tags and metadata to content in the repository; searching and browsing the content; displaying interactive visualizations that enable users to explore search results or particular themes via timelines, maps or graphical interfaces.

For the Semantic Services we have chosen to adopt the Omeka⁴ open source Web content management system developed by the Roy Rosenzweig Center for History and New Media at George Mason University. Project participants upload content to Omeka as it is acquired and attach metadata on upload. Content includes: interviews (audiovisual files), transcripts, photos, scanned books, scanned articles, scanned manuscripts, CAD drawings etc. The ELAN⁵ open source transcription tool developed by the Language Archive will be used for annotating the audio-visual interviews and for aligning transcripts with the audio/video files. The LORE semantic annotation and compound object authoring tool [11] will be used for defining semantic relationships between individual resources (both within the local Omeka repository and in external online collections). Semantic tags/annotations and RDF resource maps will be stored in a Sesame RDF triple store with a SPARQL query interface.

Location metadata associated with buildings, firms and events will be used to associate resources with map locations and to display content on mapping interfaces (Google Maps). The open source Verite Timeline JS tool⁶ will be used for displaying chronologies (biographies, building histories, firm histories) on timelines. Finally we plan to employ the Gephi⁷ open source visualization software for generating graphical views of social and professional relationships between architects and the dynamics of Queensland architectural firms over time.

⁴ <http://omeka.org/>

⁵ <http://tla.mpi.nl/tools/tla-tools/elan/>

⁶ <http://timeline.verite.co/>

⁷ <http://gephi.org/>

4.1 Upper Ontology

Underpinning many of the services described above is the OWL ontology that we have developed – which provides a common data model to support the integration of disparate datasets acquired from multiple distributed sources and organizations.

Figure 2 provides a graphical overview of the upper ontology. The core classes and sub-classes are:

- Persons
 - Architects
- Organizations
 - Universities, Firms
- Places
 - Buildings
- Documents
- Events
 - Interviews, Conferences, Forums, Panels

The ontology also defines specific attributes/properties associated with each of these classes as well as relationships between classes.

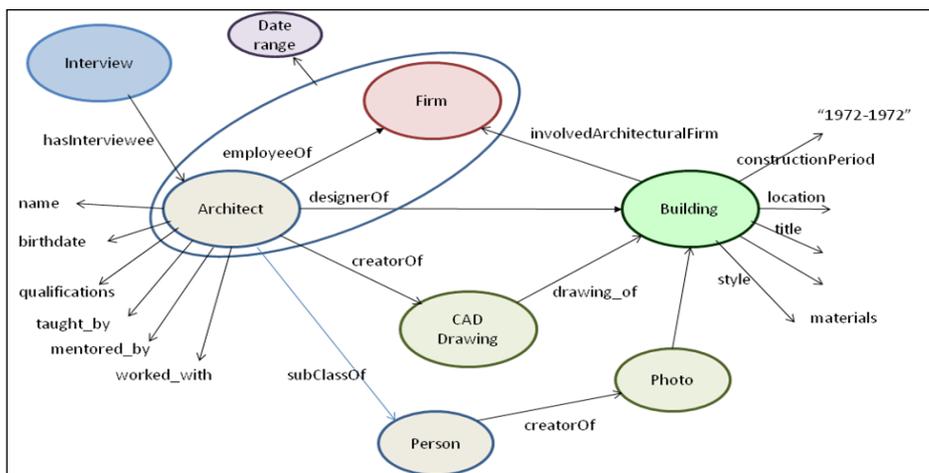


Figure 2: Upper Level Ontology for the Post-War Qld Architecture Digital Archive

5. USER INTERFACES

Users access the archive through a Web Portal [1] that provides access to the underlying Omeka archive (that stores the interviews, associated content (transcripts, images, articles)), annotation and tagging tools and interfaces to enable search, browsing and visualization of the underlying content through timelines and mapping interfaces. Figure 3 shows the current Web interface to one of the video interviews in the collection.

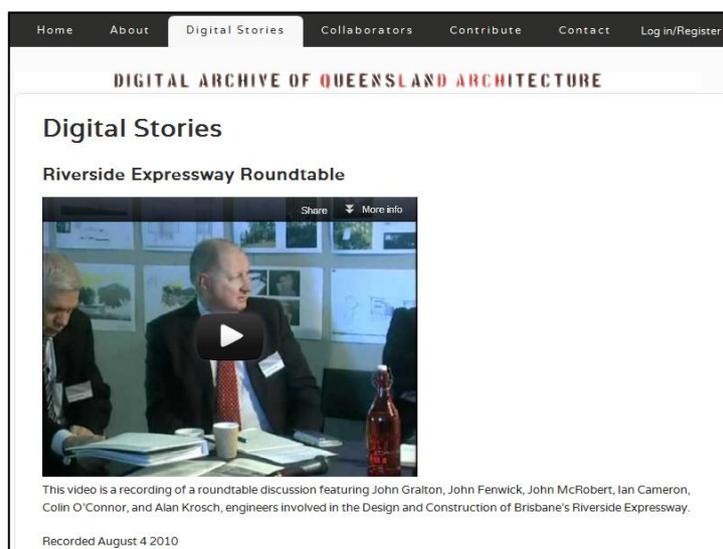


Figure 3: Web Interface to a Video Recording representing one of the Repository's Digital Stories

Figure 4 below illustrates the user interface developed to support the annotation of architectural images extracted from Picture Australia (as part of the HarvANA project) [12]. This approach is being refined for this project by plugging in the ontology described in Section 4.1 above.

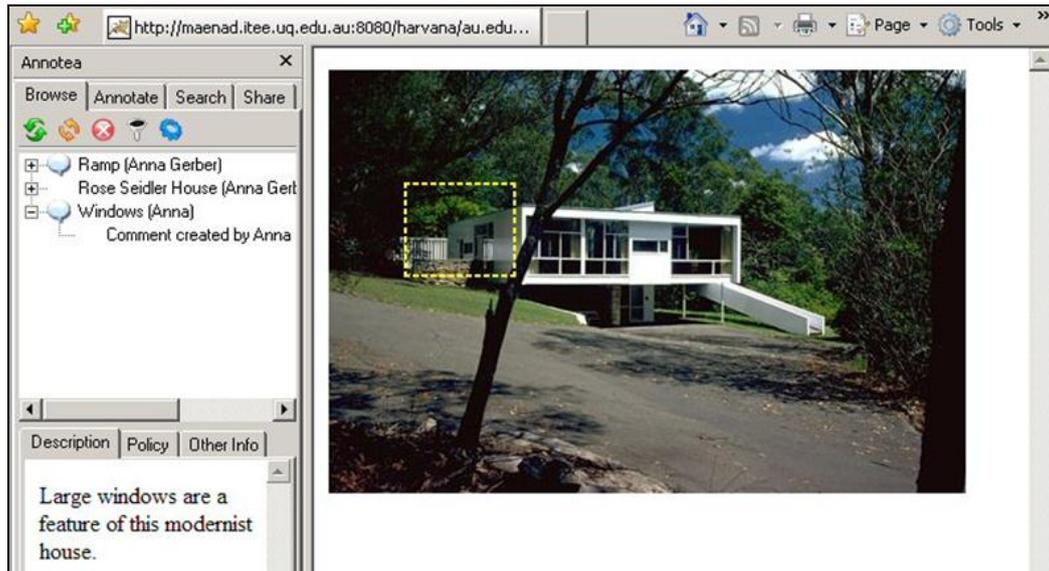


Figure 4: Illustration of Annotation Tool Interface Applied to Architectural Images

A prototype interface has been developed to illustrate the use of timelines to aggregate and visualize related resources through an interactive interface. Figure 5 below illustrates the Web interface to a timeline that presents resources and events describing the history of the old Queensland Museum Building. This timeline has been generated using the SIMILE Timeline widget⁸. Users can scroll along the timeline and click on particular events to display the content (photos, text, scanned newspaper articles, audio/video) associated with that event.

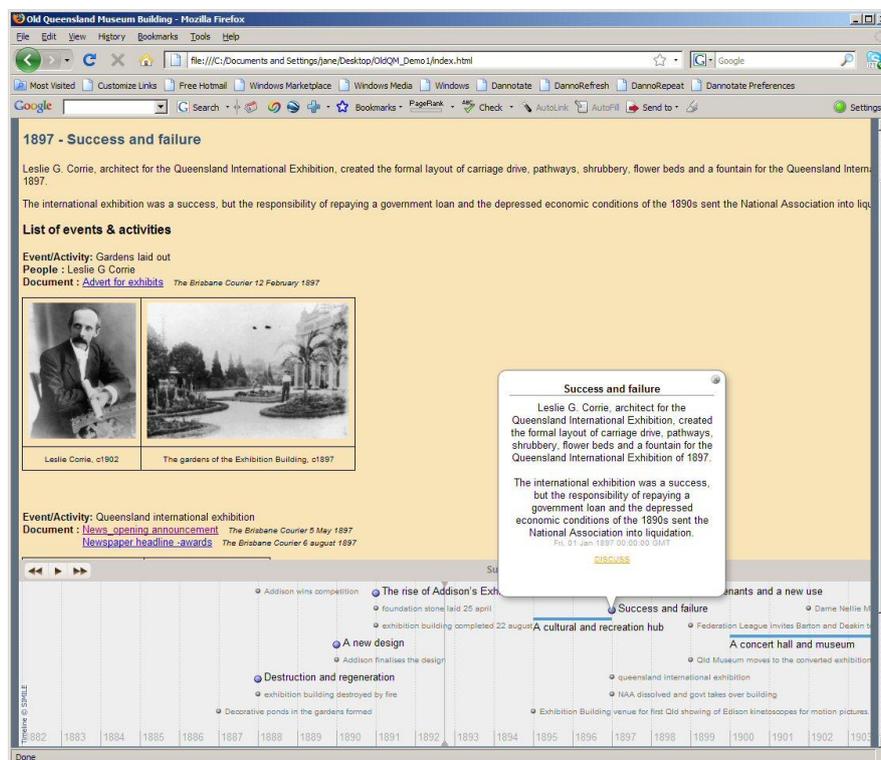


Figure 5: Example of an Interactive Timeline about the Old Queensland Museum Building

⁸ <http://www.simile-widgets.org/timeline/>

6. CHALLENGES

There are a number of significant challenges that this project will face both during the development phase as well as during the deployment and finalization phases. Below we list some of the more critical technical, social, cultural and financial challenges:

- Enabling high quality named-entity recognition and semantic tagging of the audiovisual interviews - an evaluation of the state-of-the-art speech-recognition service (Sphinx Speech Recognition Toolkit) found that even the best of such services generate poor quality results when applied in our context. The automatic speech recognition results produced by Sphinx were very poor quality (50-60% precision) due to the Australian accents, quality of the audio recordings, multiple speakers and overspeaking. Alternative approaches include outsourcing the transcription to a service such as Amazon Mechanical Turk [2], whereby anonymous online workers undertake the transcription for a very low cost. Our current preferred approach is to customize an existing tagging tool (ELAN) by plugging in our ontology – and enabling members of the team to quickly and easily tag significant terms using a pull-down menu (of people names, place names, buildings, projects, firms, dates etc).
- Building a community of contributors – it is relatively easy to build an online digital archive but establishing an online community of enthusiastic researchers, scholars and architects who can easily and frequently contribute content, knowledge and ideas to the archive is a much greater challenge. By collaborating with the State Library of Queensland who will be hosting a blog as well as forums, exhibitions and activities associated with the Asia Pacific Design Library [3], we can tap into an existing and growing local, national and global community.
- Identity resolution – *Person* name and *Firm* name disambiguation across documents is a complex problem – particularly in the context of distributed multimedia projects such as this. People’s names can be represented in many different forms, they are not unique and they can change over time (women’s married names). Firm names are equally complex as firms merge and new partners are added. One of the aims of this project is to document the evolution of architectural firms both during the post-WWII period and afterwards and provide visualizations that enable users to review this chronology over time and provide feedback/corrections.
- Quality Control over the community contributions – the increasing adoption of crowd-sourcing approaches in many fields has harnessed community intelligence to solve many problems and to spread the workload and reduce costs associated with collecting data and metadata. However it has also introduced the problem associated with the quality and trustworthiness of community-generated content. Our approach to quality control involves multiple mechanisms. Firstly we are using RDF schemas, controlled vocabularies and ontologies to validate contributor input, prior to upload. Contributors need to register prior to upload, so the provenance of all contributions is recorded. Finally the system will support a mediation step carried out by project team members – who review external contributions before they are published on the public Web site.
- Ensuring the Sustainability of the Archive – during the project development, both the content and services will be hosted on a server located at the University of Queensland. However at the end of the project, we anticipate that the content (and possibly also the services) will be migrated across to the State Library of Queensland collections. This approach will help ensure the long term preservation and accessibility of the content to the widest audience.

7. CONCLUSION

The “Architectural Practice in Post-War Queensland: Building and Interpreting an Oral History Archive” project has just completed its first year (out of three years) but during this time it has established a robust foundation for the ongoing project. The anticipated robustness and longevity of the infrastructure are based on the decision to adopt and integrate a number of open source tools (including Omeka, ELAN and Sesame) and Web 3.0 open standards (including OWL, RDF and the Open Annotation (OA) model). Moreover by building the archive through a collaboration that involves the local architectural research community, the State Library and industry partners as well as the general public, we anticipate that the archive will be sustained beyond the life of this specific project.

REFERENCES

- [1] The Digital Archive for Queensland Architecture <http://qldarch.net/>
- [2] Amazon Mechanical Turk <http://aws.amazon.com/mturk/>
- [3] Asia Pacific Design Library <http://apdl.slq.qld.gov.au/>
- [4] CMU Sphinx, Open Source Toolkit for Speech Recognition, Carnegie Mellon University <http://cmusphinx.sourceforge.net/>
- [5] Wilson, A. C. & Macarthur, J. (Eds.) (1997) *Birrell : work from the office of James Birrell*, Melbourne, NMBW Publications.
- [6] Wilson, A. C. (Ed.) (2005) *Hayes & Scott. Post-war houses*, Brisbane, University of Queensland Press.
- [7] Sinnamon, I. (1985) An Educated Eye: Karl Langer in Australia. *Landscape Australia*, 1, 48-56.
- [8] Watson, D. A. McKay, J. (1984) *A Directory of Queensland Architects to 1940*, St Lucia, University of Queensland Library.
- [9] Gardiner, F. (1988) Register of Significant 20th century architecture Qld. Report to RAIA (Qld).
- [10] Riddel, R. (2005) Significant Queensland 20th century architecture: A Report for RAIA (Queensland).
- [11] Gerber, A & Hunter, J (2009) "*LORE: A Compound Object Authoring and Publishing Tool for Literary Scholars*", Digital Humanities 09, University of Maryland, USA, June 22 - 25, 2009.
- [12] Hunter, J, Khan, I & Gerber, A (2008) "*HarvANA - Harvesting Community Tags to Enrich Collection Metadata*", ACM IEEE Joint Conference on Digital Libraries, JCDL 2008. Pittsburgh, June 16 – 20.
- [13] Langer, K. (1944) *Sub-tropical housing*, Brisbane, University of Queensland.