INTRODUCTION
Public Record Office Victoria (PROV), and the Victorian eResearch Strategic Initiative (VeRSI), are collaborating in a pilot software development of a crowdsourcing online transcription platform, through which members of the public can access images of public records, transcribe, tag and geo-locate them.
The collaboration represents a melding of the “Gov 2.0” and “eResearch” strategies of the two organisations, and a utilisation of the emerging cultural activity of crowdsourcing that has the potential to create a valuable public information resource and a rewarding experience for participants.

PROV
PROV is the archival authority for the State of Victoria, holding records from the 1830s to the 2000s, which researchers can access physically at PROV reading rooms. Through its digitisation programme, PROV makes an increasing number of these records available over the internet, as digital images. Online access facilitates access to records, particularly for people who find it difficult to access PROV’s reading rooms in Melbourne or Ballarat.

For the last twenty years, teams of volunteers at PROV have transcribed millions of names and other details from public records to assist their own and others’ research into family, place and community.
As part of Public Records 2.0, a key initiative of the Victorian Government 2.0 Action Plan, PROV is committed to building functionality for an online volunteering program.

VERSI
Online access to digital records can also lead to a qualitative change in the use of historical records for research. Traditionally researchers have focused on finding, copying, and studying particular records, and on extracting summary information from the records to create indexes and other finding aids, whereas if the records can be fully transcribed, in a structured way, researchers may make use of the emerging eResearch methodologies such as visualisation, statistical analysis, pattern recognition, and deal effectively with records at a large scale; as datasets rather than as individual records. The crowdsourcing transcription service being developed in this pilot project can therefore be seen as infrastructure for innovative eResearch in the Digital Humanities and Social Sciences; an important part of VeRSI's mission.

THE CROWDSOURCING TRANSCRIPTION PILOT PROJECT
The goal of the project is to develop and publish an open source online application that will enable members of the public to find, transcribe, index and geo-locate images of records held by Public Record Office Victoria.
The application will enable a new generation of volunteers and researchers to work remotely, using digitised copies of records. The extraction of structured data from hand-written nineteenth century documents has presented particular challenges, and this project will allow a broad range of members of the community to build up skills in using these records and share the results of their work.
The initial records used for the application will be council rate records, which contain various forms of data that are of value to a wide range of researchers in the community including family and local historians, heritage managers, archaeologists, historians and social scientists. The types of data
contained in council rate records include: personal names; places of residence; dates of occupancy; types of property and property valuations.

The functionality of the application, however, will allow for the indexing and transcription of many different types and formats of records. As well as enabling a new method of working with records, this tool will enable new types of information to be captured and shared with new audiences, particularly the academic research community, who will be able to compile economic and social data from transcribed records. It will be easier for all researchers to search for records, and to extract the transcribed data.

Outcomes of the project include:

- Delivery of a model for collaborative humanities eResearch infrastructure design.
- The deployment of the TEI (Text Encoding for Interchange¹) mark-up language, which has traditionally been used in academic settings, by a public archive.
- Support for the growth of PROV’s regional and metropolitan volunteering program by enabling online participation through a crowdsourcing tool.
- Support for community-based research by providing a tool for indexing and transcribing images of public records in community collections, or relevant to local area communities, such as the 150 members of the PROV Places of Deposit program².
- Facilitation of the emergence of the semantic web and increasing discoverability of information via transcriptions and indexed (linked) data, including semantic mark-up and geo-location of transcribed text.
- Enabling the creation and management of data and images in a standard, open format that allows them to be stored in PROV’s Digital Archive for long-term preservation and access. This format will permit these images and data to be re-usable by other systems and users. Current examples of this approach are the National Library of Australia’s Trove site³ and the National Museum of Australia’s History Wall.⁴
- Meeting an existing demand from government and the community for a more sophisticated access model to permanent public records via more comprehensive descriptions of records complemented by a richer metadata set.
- The consolidation of a range of research activities and data creation into a public resource.

REFERENCES


ABOUT THE AUTHORS

Abigail Belfrage is an historian working in Online Engagement at Public Record Office Victoria (PROV), the archives of the State of Victoria. In partnership with the Victorian eResearch Strategic Initiative she is working on a pilot project to create online transcription and text encoding software for public records held at PROV. As well as her own research interests in landscapes and places, she is passionate about and how social media and other emerging web technologies can provide new ways for people to create and share knowledge about places, people, technology and processes that they care about.

Conal Tuohy is an eResearch Business Analyst and software developer at the Victorian eResearch Strategic Initiative (VeRSI), working in the Digital Humanities area.

Between 2002 and 2008, he was lead developer at the New Zealand Electronic Text Centre; a platform for online publishing of the Centre’s collection of digitised New Zealand and Pacific books. From 2008 to 2010 he worked for the eScholarship Research Centre at the University of Melbourne, particularly in the area of archival metadata.
His interests include the digitisation of cultural heritage, text encoding and knowledge representation.