

User Interaction Ideas:

What if there were no boundaries - how would users want to interact with newspaper data, other users, and add value to the data and the service?

Author: Rose Holley (Manager - ANDP)

Version: 1.1

Date: 3 December 2007

Background

A national digital program on the intended scale of the Australian Newspapers Digitisation Program will have a significant impact in the digital space, particularly for genealogists and local historians. Key requirements for these users are not only to be able to access and search the data but also to be able to interact with it and with other users of it in order to generate and create new information and resource in new ways. Newspapers users will in particular want to be able to share data, share research, confirm accuracy of data, and make linkages around data particularly in respect of people's names. Newspapers hold the history of the nation in 3 broad areas - peoples, places and events. The Library has the potential to transform the newspapers from a static historical resource into a living history that can be shared, developed and re-purposed in new and exciting ways by use of web 2.0 applications and technology. It is essential that the users can mine the data in old and new ways, add value to the data, and mash (share) the data in online communities using latest web 2.0 technologies. The Library is unaware of any other newspaper digitisation project that is freely and publicly available and has enabled user interaction to this extent. The Library believes it has the capability to offer such a service and meet the needs and expectations of users in new and exciting ways. Some ideas are outlined below of possible user interactions in the digital world of Australian newspapers.

Accuracy of data, correcting OCR, addition of notes, tagging

Newspapers are renowned for their inaccuracy of data particularly in feature articles and particularly with the misspelling of people's names. Users could have the ability to add notes or amendments to the data by use of annotations so that others can also see and verify/dispute the value added information. The annotations could be searched across and viewable on demand.

The quality of the OCR text may be poor, but the actual page image readable to the human eye. Users may wish to correct the OCR data. If a word is wrong in the OCR it will not be retrieved in the search results but if it is corrected it will. Making corrections to people's names in particular would enhance accuracy and usefulness of name searching for genealogists. Users may wish to monitor and let other people view how many corrections they make and become 'OCR Correction Kings'. OCR corrections could be moderated in a similar way to Wikipedia or the Library could trust the community and let OCR corrections be a freely flowing process, monitored by the community itself.

Users may wish to tag articles with subject keywords e.g. World War 2, rabbit proof fence, and view their own and other peoples tags in tag clouds. They may want to integrate searching across annotations and tags with article text.

Sharing (mashing/repurposing) data and personal research

For genealogists it is really important that they have some means to share the fact that they are searching for peoples names or that particular data has been useful to them. Unlike Flickr

where users know exactly who they want to open their photo albums to genealogists in most cases won't know who else is also researching information on these names so it is important they can have some vehicle preferably inside the ANDP to inform or alert people to this. This could be by use of 'blogs' or forums set up as part of the ANDP. If people want to share the data, cite the data, and re-purpose the data they should be able to in ways that will be useful to them. All the data will be freely available. So long as it is within copyright regulations users may want to do different things with data. For example creating a newspaper e-card for your birthday with the date you were born, or a hard copy reprint of the whole newspaper.

Telling their story and what happened next?

The facility to add annotations (notes) may prove hugely successful and take on a life of its own. It could be expanded so that rather than adding brief notes users can add lengthy information detailing their own views on articles/events, talk about similar events, explain what happened years after the events and to provide links to other items they have created by use of the newspaper data e.g. family histories. This would change historic newspapers into a living history resource about and for Australians. The newly generated value added data would be of cultural heritage value. If a means for adding value in this way is not achieved within the ANDP itself users will find their own ways to share the information via blogs etc. This additional and important information would then be out of context from the original resource and 'lost'. It is highly likely it would not be digitally preserved as a significant cultural heritage resource.

Subject Trails.

Some people searching newspapers are looking for a specific event through time. It would seem logical that the users, rather than Library staff have the ability to build up public 'trails' and 'timelines' of events that interest them, that can then be viewed and edited by others. This may be particularly useful in the education sector where educators create topics/features/trails for their students to access.

Location of events. Geospatial and temporal searching

Users have demonstrated that sometimes they find it useful to tag places on a map and add notes about events that happened here (e.g. the tagging of streets on digital maps that happened after Hurricane Katrina). This is useful if searching is enabled on a geographical map interface to find articles, rather than traditional text based methods. There is potential to use Google technologies in the interface. Changes in historical names and border changes etc will probably have taken place, so the authorized national gazetteer could be built in. It is advantageous that Australia has a single recognized gazetteer cross referencing all old and new names. There is potential to develop geospatial searching over temporal time zones (especially useful for school children). Ian Johnson at the University of Sydney archaeology department has done much work on historical map searching and the Timemap software is available open source: <http://www.timemap.net/> It would be challenging to meld Google open source software with Timemap software and instead of using data sets use tags, and also to make it simple enough for the average user. This could be a real ground breaking delivery technique that strongly relies on visual searching techniques rather than text based methods. This is something users are unlikely to request since there is little understanding of visual search techniques as opposed to traditional text methods and very few geospatial search interfaces are aimed at the general public rather than geographers. Using visual 'timelines' effectively for large corpuses of data has not been implemented widely, but it is a useful way of searching the corpus.

Alerting, 'saved search' profiles, what's coming up.

Users will want to know what new content is coming up and what the forthcoming scanning schedule is. RSS feeds could inform users not only of new content added, but pages in which they are interested. Because local historians will have ongoing research it is better that they have an auto-search alert e-mailed to them or incorporated into the RSS, rather than manually having to keep repeating searches.

Linkages with other data/future developments

There is potential in the future for the ANDP to become the authoritative 'national history site' if some of the above user interactions are taken on by users. History = PEOPLE, EVENTS,

PLACES. The newspapers cover all 3. It would add value to the user experience if other resources could be searched simultaneously e.g. historic video and audio clips to match events, maps to match places, and integration of historic images (Picture Australia) to match peoples, places and events. The retrieval of other resources for people will be particularly important. One of the most important sources for looking at peoples names in early 19th century are shipping registers. The other key source is historic maps.

Personalisation

For newspaper users it would be useful to have a 'my newspaper' interface where they set up their own interface preferences and store their own useful articles. The newspaper archive www.newspaperarchive.com (a subscription service) has some really useful functionality such as 'my archive', 'my filing cabinet', and forums which could be applied.

Indexes

Integration of indexes will be an important aspect of the service at a later stage. Many of the existing hard copy and online newspaper indexes have been created by use of volunteers and it is highly likely that users may want some interaction in the application and integration of indexes. It is uncertain at this stage what this may be.

Newspaper Biographies

Newspaper historians have created useful data and research on specific newspaper titles. This is largely inaccessible in the online world. It would be useful to be able to digitally access and display newspaper biographies in the service, but it is not desirable that data is stored in the database itself. The preference would be to link to an open source resource to show the biography data (e.g. Wikipedia) and have users create, modify and edit that data externally to the service.

Summary

The Library is well placed to demonstrate how the data in a full text database can be made accessible to users in ways that meet and possibly exceed their needs and expectations. The Library will draw on its experience and expertise, and work closely with users and stakeholders to develop a service that meets the needs of the 21st Century user, whilst also enabling the Library to further expose and improve access to the rest of its collections by developing new models for access.