

OLCAP: Online Language Community Access Pilot

Jason Lee

AIATSIS <Jason.Lee@aiatsis.gov.au>

Patrick McConvell

AIATSIS <Patrick.McConvell@aiatsis.gov.au>

OLCAP, the Online Language Community Access Pilot, is a trial community-focused approach to accessing language documentation online. This project aims to provide language documentation in the form of audio-visual and text material online via the internet. Audio or video is linked with time-aligned transcription, translation and other annotation. OLCAP currently focuses on three general areas: Cape York Peninsula, centred around Lockhart River; The Victoria River District; and, the Iwaidja language community. In this presentation, I will give an overview of the project and the progress made so far.

1 Introduction

OLCAP, the Online Language Community Access Pilot, is a trial community-focused approach to accessing language documentation online. This project aims to provide language documentation in the form of audio-visual and text material online via the internet. Audio or video is linked with time-aligned transcription, translation and other annotation. OLCAP currently focuses on three general areas: Cape York Peninsula, centred around Lockhart River; The Victoria River District; and, the Iwaidja language community. In this presentation, I will give an overview of the project and the progress made so far.

OLCAP is a collaboration among several bodies. The Australian Government Department of Communications, Information Technology and the Arts (DCITA) is the funding body. The Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) is providing staffing and infrastructure. The Max Planck Institute for Psycholinguistics (MPI) in Nijmegen in the Netherlands is providing technical expertise. Various individuals have been enlisted at the ground level. These include linguists from Diwurruwurru-Jaru, the Katherine Regional Aboriginal Language Centre.

The original aims of the project were to implement and test ways to enable local people to access and use the material in up to six steps:

1. finding, downloading and watching and listening to recordings and annotation;
2. copying the material to CD/DVD where appropriate;
3. adding information about the material, including who should access it;

4. adding more annotation or noting where annotation is lacking;
5. editing the material to make new products, e.g. for education, where appropriate;
6. uploading new information and material to a repository.

One of the main rationales for the project was to see if we could improve the time frame and efficiency of repatriation of materials.

2 Overview

OLCAP centres around a website, located at: <http://olcap.aiatsis.gov.au/>. It utilises software developed by MPI. The main archive metadata structure and software is called IMDI (ISLE Meta Data Initiative). One key feature of the software is ANNEX, which allows the viewing of linguistic annotations time-aligned with audio and/or video. MPI are also in the process of making a tool for viewing dictionaries online. (Other tools for publishing dictionaries online have been available for a while, such as that inbuilt in Shoebox/Toolbox, Lexique Pro or Kirrkirr, but one aim for the proposed MPI lexicon tool, LEXUS, is interactivity with ANNEX).

Currently, there is very little material online. This is, however, a pilot. Therefore, quantity is not of the highest importance. A locally produced VRD (Victoria River District) video from the language centre in the Gurindji language was put into ELAN (MPI transcribing tool) then uploaded to the AIATSIS server and can be streamed using ANNEX.

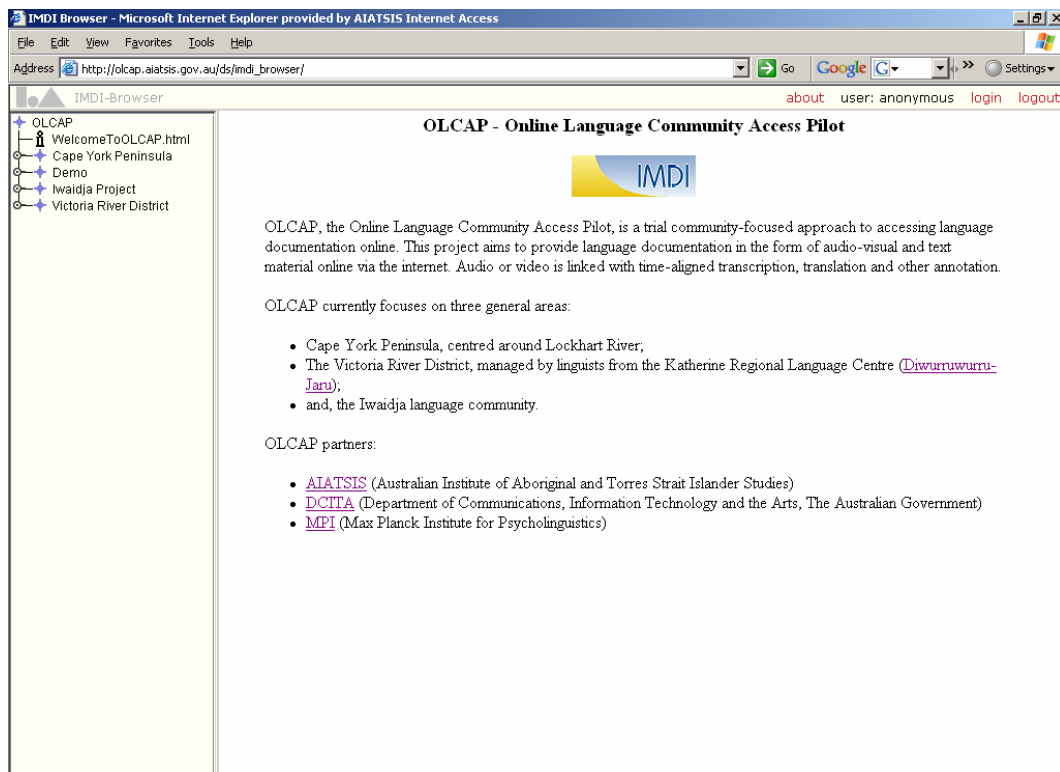


Figure 1: Screen shot of the OLCAP home page



Figure 2: Screen shot of annotated video showing annotations as block text

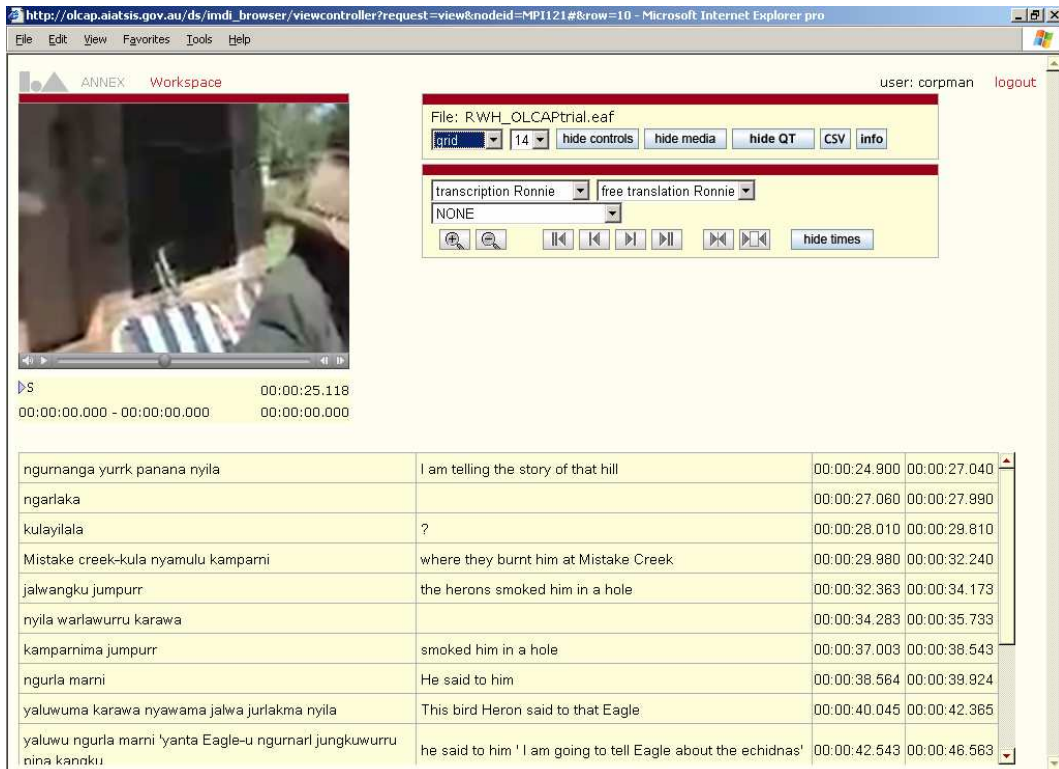


Figure 3: Screen shot of annotated video showing annotations in a grid

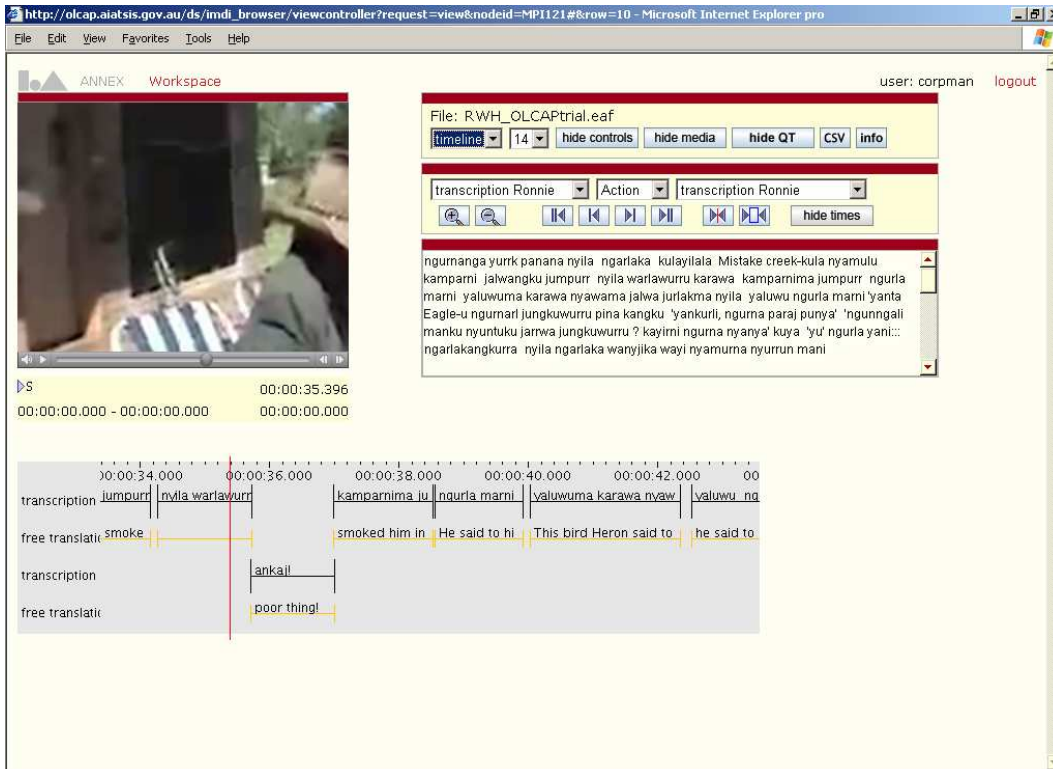


Figure 4: Screen shot of annotated video showing annotations in a timeline

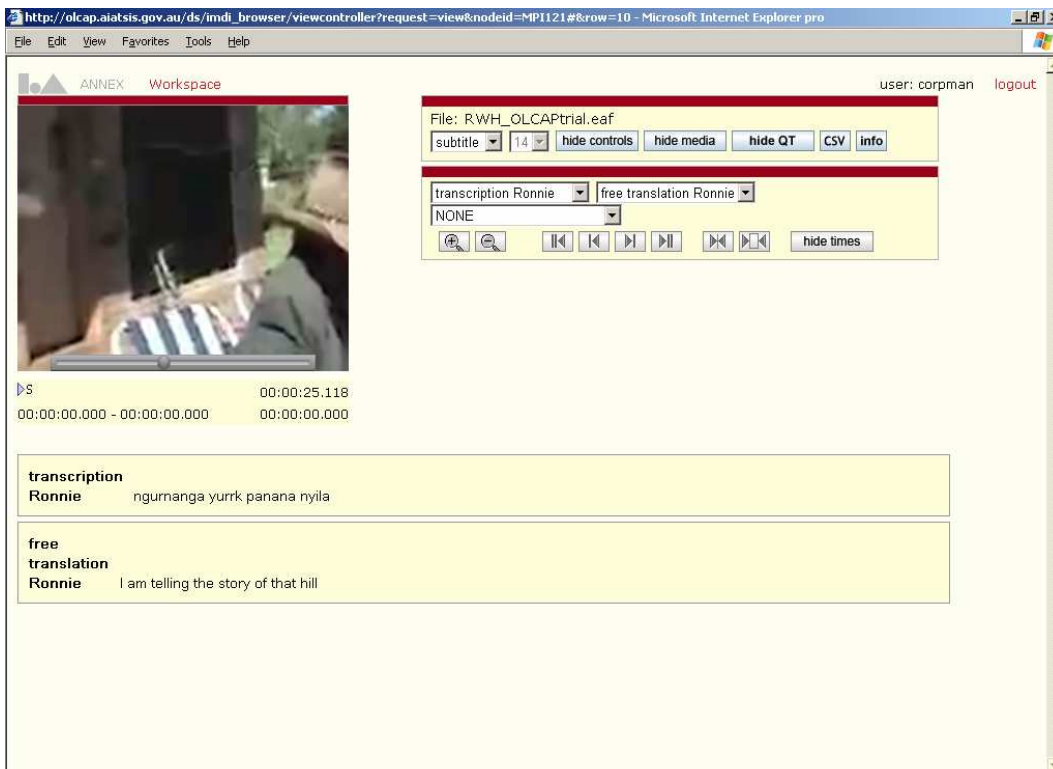


Figure 5: Screen shot of annotated video showing annotations as subtitles

3 Progress so far

The OLCAP server is physically at AIATSIS. It became active from the beginning of September 2007. Some sample data has since been put on the server.

Several workshops and meetings have been run in Canberra and Katherine to get the project moving.

A receiving computer was installed and tested at the Katherine Language Centre. Accessing the archive worked from various computers on site, but a point of concern was that uploading was particularly slow. Slow connections should be easily solved by the centre purchasing a better broadband plan. However at other locations including Lockhart River there may be more significant problems due to the quality of broadband access.



Figure 6: The newly installed OLCAP workstation in the Katherine Language Centre library

4 Similar systems

There are several systems of digital archiving of similar material in progress or in existence. Here I shall review some of them.

4.1 Other IMDI archives

There are other archives online that utilise IMDI software. Including Canberra, these are their locations:

- Iquitos, Peru - <http://iquito.iiap.org.pe/>

- Kiel, Germany - <http://lingcorp.linguistik.uni-kiel.de/>
- Lund, Sweden - <http://dam-lr.sol.lu.se/>
- London, UK - <http://imdi.soas.ac.uk/>
- Rio de Janeiro, Brazil - <http://app02.museudoindio.gov.br/>
- Buenos Aires, Argentina - <http://dobes.caicyt.gov.ar/>
- Canberra, Australia - <http://olcap.aiatsis.gov.au/>
- Leiden, the Netherlands - <http://imdi.inl.nl/>

4.2 *Other digital archives*

Other large digital archives exist that do not utilise IMDI software. Some of them can not be accessed online, such as the AIATSIS audiovisual archive. However, a small portion of its metadata is searchable and browsable via the AIATSIS library online catalogue MURA.

PARADISEC, the Pacific and Regional Archive for Digital Sources in Endangered Cultures, is similar to IMDI archives in many ways. However, the metadata layers in PARADISEC are not required to be as rich as that promoted by IMDI structure.

4.3 *Ara Irititja / Our Story / Memory Place*

Aboriginal communities around Australia have started archiving digital audio visual materials themselves. The focus of these ‘archives’ has so far been on standalone solutions of media kept on location rather than remotely accessed via the internet. However, some solutions have been moving toward an online solution.

The first piece of software that I will mention is the Ara Irititja software. It forms the basis of the ‘Our Story’ databases of the Northern Territory Library:

The Our Story database is a key part of the new LKC model. It uses the Ara Irititja software developed for Anangu Pitjantjatjara communities in Central Australia. It stores photographs, videos, sound recordings and other documentary items. An important feature of the database is the ability to restrict access to individual items to cater for cultural sensitivities. Further information about the Ara Irititja software can be obtained from the website: <http://www.irititja.com/>.

(http://www.ntl.nt.gov.au/about_us/knowledgecentres)

One thing to note is that the Our Story database is stored on individual local hard drives, manually backed up by copying the hard drive.

The successor to Our Story, ‘Memory Place’, is currently still in development. The NT library is still planning to keep these archives offline. The software is likely to have online capabilities though. Details can be found at: <http://sourceforge.net/projects/memoryplace/>.

One of the archives utilising the Ara Irititja software is Kanyirninpa Jukurrpa, the Martu History and Archive Project in the Western Desert of Australia.

4.4 Mukurtu Wumpurrarni-kari

Mukurtu Wumpurrarni-kari is a digital archiving solution for Warumungu communities around Tennant Creek that is similar to Ara Irititja developed by Kimberly Christen and her associates. It is potentially online and is XML based but currently runs on a local network in the Culture Centre and there are no plans for using the internet. Their goal is to create an open source software package for use by any Indigenous community.

4.5 Bidwern

Bidwern is another digital archiving solution for a specific region. Bidwern was created for the western Arnhem land communities. The main feature is a GIS interface. It is online via restricted access. It is supposed to include digital dictionaries but we don't know how it might interface with MPI software.

4.6 Open repositories and tools

Other than these customised solutions for digital archiving of audio visual material, there are many popular internet repositories and tools for miscellaneous media that individuals might possibly be utilising, such as Wikipedia, YouTube, ImageShack, Photobucket, etc.

5 IMDI

One issue that we are facing is the difficulty for users dealing directly with IMDI editors. The way IMDI archives are designed 'strongly encourages' the input of rich metadata. However, it is not very user friendly and is designed more for researchers, who even so face a steep learning curve. If one of our goals is to allow naive users to input metadata, a better solution has to be found. There is also considerable repetitious data entry required, although IMDI developers have tried to reduce this with each new version of the IMDI editor. We hope that even more streamlined solutions will eventuate.

Another issue is that of the user interface. The IMDI archive user interface is not very friendly, compared with the Ara Irititja interface, for example. With the limited resources for the current project, the best we can aim for is to use a CMS (Content Management System) to produce a simpler HTML interface. In the longer term, if this pilot project is extended, a simpler interface along the lines of Ara Irititja should be sought. Wouldn't it be nice if IMDI developers and Memory Place developers could get together and collaborate?

Online	Local
Economies of scale managing an archive online. Management also possible remotely.	Lack of management / funding of archives in small under-funded communities.
Upload / download of large files slow and expensive.	Not limited by internet availability / quality.
Basic structure that can be taken up by any community.	Some localised solutions are expensive to build and expensive to transfer to other communities.
One size does not always fit all.	Localised solutions can be custom fit to communities.
MPI software targets linguistic data (e.g. time-aligned interlinear text and lexicons) at the expense of user friendliness.	Some solutions are more user-friendly but do not target linguistic data.

Figure 7: A comparison of online versus stand-alone archive solutions

6 Conclusion

Error! Reference source not found. is a rough comparison of online versus stand-alone local solutions for digital archives.

There is perhaps no ideal solution. However, by working through this pilot project and careful consultation, we can hopefully move forward.

7 References

- Ara Irititja Project, The*, 2007, (Ara Irititja homepage). Retrieved 2 November 2007, <<http://www.irititja.com/>>.
- Bidwern Project*, 2006, (Bidwern homepage). Retrieved 2 November 2007, <<http://www.anu.edu.au/bidwern/>>.
- Christen, Kimberly, 2007, *The Politics of Search: Archival Accountability in Aboriginal Australia*. Work in progress. Paper presented at MIT5: Media in Transition 5 Conference, Technological Translations and Digital Dilemmas Panel. Retrieved 2 November 2007, <http://cooneycreative.com/kimberlychristen/wp-content/uploads/2007/04/politics_of_search_christen.pdf>.
- Christen, Kimberly, 2007, *Warumungu Community Digital Archive Project (version 1.2 July 2007)*, Long Road. Retrieved 2 November 2007, <http://www.kimberlychristen.com/?page_id=3>.
- Christen, Kimberly, 2007, *Mukurtu - 1200+ pieces of content and growing*, Long Road. Retrieved 2 November 2007, <<http://www.kimberlychristen.com/?p=201>>.
- ISLE MetaData Initiative (IMDI)*, 2007, (IMDI homepage). Retrieved 2 November 2007, <<http://www.mpi.nl/IMDI/>>.
- Memory Place*, 2007, SourceForge.net: Memory Place. Retrieved 2 November 2007,

<<http://sourceforge.net/projects/memoryplace/>>.

Libraries and Knowledge Centres, 2007, Northern Territory Government of Australia, Department of Local Government, Housing and Sport. Retrieved 2 November 2007, <http://www.ntl.nt.gov.au/about_us/knowledgecentres>.

OLCAP, 2007, (OLCAP homepage). Australian Institute of Aboriginal and Torres Strait Islander Studies. Retrieved 2 November 2007, <<http://olcap.aiatsis.gov.au/>>.

PARADISEC, 2007, (PARADISEC homepage). Retrieved 2 November 2007, <<http://www.paradisec.org.au/>>.