

Digital posters: Publishing Gurindji plant and animal poster content as websites using an open-source template-based RO-Crate preview tool

Ben Foley, Abigail Davis, Felicity Meakins

The University of Queensland, Australia

b.foley@uq.edu.au, abigail.davis@student.uq.edu.au, f.meakins@uq.edu.au

Abstract

Technology can play an important role scaffolding the return to language vitality. This paper outlines the repurposing of language material from previously created Gurindji ecological posters to create websites, using an easily implementable workflow which conforms to the RO-Crate standard, ensuring the longevity of the websites and underlying data. Through this work, four websites have been published, connecting the old with the new, bringing the voices and knowledge of Kajijirri and Marlaluka (Old People) to the ngumayijang (next generations).

1 Introduction

Bringing together Gurindji language material from an award-winning poster series and an existing website tool, our work demonstrates the benefits arising from packaging existing language material according to the RO-Crate standard. We describe a relatively fast, low-cost, low-maintenance and long-lasting method of publishing language content online with data in RO-Crate format. The production leverages the prior work done in collating content, requiring minimal further work to reformat and republish for online publication. Four websites were built using this method (for a list of website addresses, see Appendix A).

2 Background

2.1 About the posters

The Gurindji plant and animal websites began their life as analogue posters (see Figure 2). The Gurindji plant and animal posters project ran between 2014–2018 as a collaboration between non-Indigenous linguist Felicity Meakins (University of Queensland), Gurindji project facilitator Cassandra Algy (Karungkarni Art and Culture Aboriginal Corporation), Gurindji rangers (Murnkurrumurnkurru Central Land Council ranger group) and Gurindji

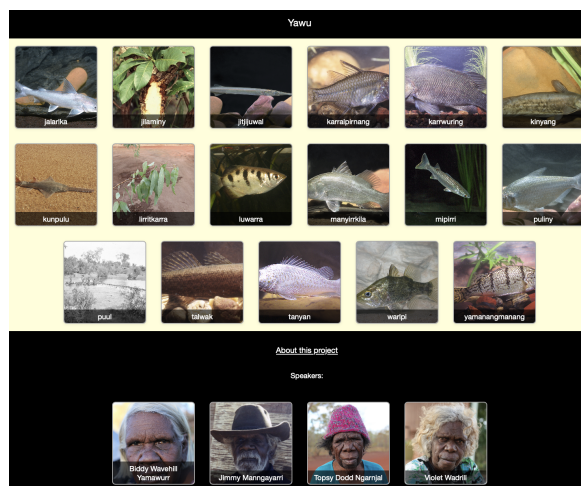


Figure 1: Yawu fish site homepage

Elders and Cultural Custodians, in particular Violet Wadrill.

The posters were made in four one week workshops (one per year) on Gurindji Country. We shortlisted plants and animals for the different themed posters by deciding on the most culturally-important biota to the Gurindji community. As well as discussion with Gurindji Elders and Cultural Custodians, we relied on the ‘Bilinarra, Gurindji and Malngin Plants and Animals’ book (Hector et al., 2012) which was created by the same team with non-Indigenous biologist Glenn Wightman from the Northern Territory Department of Land Resource Management. Cultural and ecological information about the plants and animals were recorded with Gurindji Elders, and web hosted¹ with QR codes created to easily access the audio from mobile phones. Bird and fish photos were sourced from various outside photographers. Plant photos were taken on Gurindji Country by then Karungkarni Art manager Penny Smith. Graphic designer Max Addinsall created the four Gurindji poster series which included the QR codes.

¹<https://ngumpin.org.au/gurindji>



Figure 2: Gurindji fish poster with QR codes created in 2015³

The posters were published by Indigenous publisher Batchelor Press.²

The team also created learning resources and activities for Gurindji children at Kalkaringi School and delivered lessons to upper primary classes.

In 2018, the posters won the Northern Territory Land Resource Management (LRM) Environment & Conservation Award. The award recognises the importance of First Nation’s knowledge to understanding the ecology of Australia.

2.2 Technology background

The tool used to create the four websites was developed by the Language Data Commons of Australia⁴ (LDA), an Australian research and infrastructure project which aims to improve the ways people work and conduct research with recordings, manuscripts and other language material. LDA partners with institutions and communities to keep at-risk language collections safe for the future. Some of the work that LDA is doing includes storing language material and metadata in ways

²<https://batchelorpress.com>

³<https://batchelorpress.com/product/gurindji-fish-poster>

⁴<https://ldaca.edu.au>



Figure 3: Ranger Kenny Ricky teaches Gurindji children about local birds with Elders Ronnie Wavehill† and Paddy Doolak†. (Photo: Cassandra Algy 2014)

that are good for archiving, by using the RO-Crate data standard⁵.

2.2.1 RO-Crate

LDA is involved in maintaining the RO-Crate standard, an approach to packaging data along with rich metadata (Stian Soiland-Reyes et al., 2022). An RO-Crate is a file-based method of storing data, in which data files (e.g. audio recordings or images) are stored alongside a metadata file containing descriptions of the data, and licence information detailing conditions of access. The metadata file is in JSON-LD⁶ format, making the format highly machine-readable.

Using RO-Crates for storing data and metadata leads to opportunities (where it has been deemed appropriate according to access protocols) for easy sharing, archiving, reusability and reproducibility. RO-Crate can be used to package language collections for archiving and long-term safe-keeping, connecting datasets, publishing research data along with any analysis and the tools used, amongst other uses.

A benefit of using RO-Crate format is that the metadata schemas that are used to describe the material in collections can be adapted and customised to describe data richly, which enables people to use culturally specific schemas to describe collections.

2.2.2 RO-Crate preview

The RO-Crate specification includes requirements about what information is mandatory or optional in a description of material—such as identifiers, item types and date published metadata. Along with

⁵<https://www.researchobject.org/ro-crate>

⁶<https://json-ld.org>

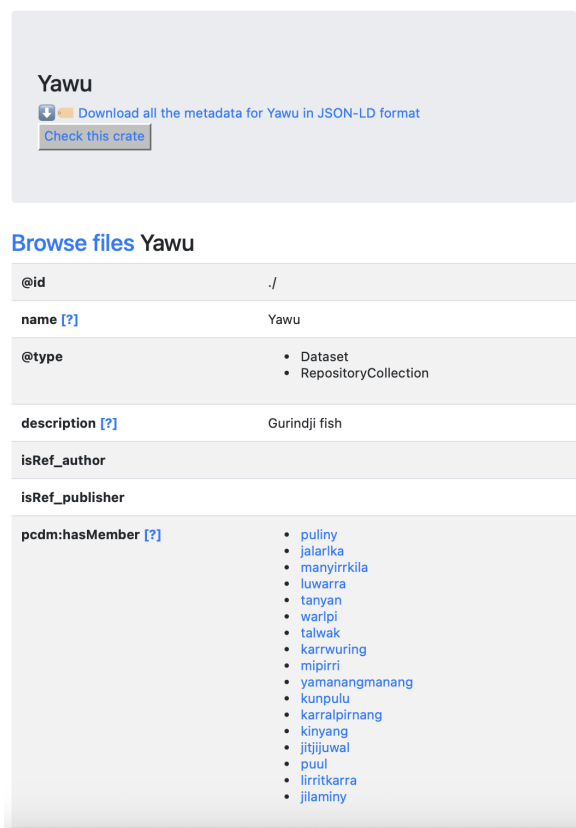


Figure 4: RO-Crate default HTML preview

these information requirements, the specification provides for the inclusion of a standalone, accessible HTML version of the data and metadata in the package, intended to make the RO-Crate more human-readable.

The preview HTML provides a way to explore a collection without requiring specialised software or technical knowledge, ensuring that anyone with a web browser can navigate and understand a collection's data and context (e.g., licensing and access information which are described in the metadata). A generic HTML preview (see Figure 4) can be built from any RO-Crate, using the `ro-crate-html-js`⁷ command-line tool or by using the `Crate-O`⁸ browser-based tool. When using `Crate-O`, there is an option to choose from two versions of HTML preview styles. Future work with `Crate-O` includes an option to select from a variety of HTML templates and uploading custom templates.

LDaCA has developed `ro-crate-html-lite`⁹, a tool

⁷<https://github.com/Arkisto-Platform/ro-crate-html-js>

⁸<https://language-research-technology.github.io/crate-o>

⁹<https://github.com/Language-Research-Technology/ro-crate-html-lite>

to build HTML previews that are more specifically designed to suit the needs of individual collections. The `ro-crate-html-lite` tool uses HTML templates to enable developers to customise the look and feel of the generated website using Nunjucks¹⁰ templates, a commonplace website templating language. Templates can be shared and customised, leading to opportunities such as the work described in this paper. Our work makes use of templates that were developed this year to publish website versions of a series of Indigenous language bird apps built between 2015-2021 that are no longer available.

The `ro-crate-html-lite` tool can also be adapted to build multi-page sites based on objects in the collection, instead of the generic `ro-crate-html-js` tool's approach of a single-page HTML for the entire collection.

2.2.3 PILARS

LDaCA also maintains PILARS¹¹, a set of principles for the design and implementation of archival repositories. Relevant to the work described in this paper, one of these protocols is that "data is portable and not locked into a particular storage system" (Peter Sefton et al., 2024).

The approach used in our work to publish the poster content online follows this protocol in that the output of the RO-Crate preview tool is in plain HTML format. There are no long-term dependencies on particular website content management systems or proprietary/open-source databases.

3 Approach

The approach taken in this work followed stages of collating content, packaging the material in RO-Crate format, adjusting the HTML templates and rebuilding the HTML preview, and publishing the HTML preview pages.

3.1 Preparing spreadsheets

We created four spreadsheets, one for each poster, based on an RO-Crate spreadsheet template available from the LDaCA website resources page¹². The spreadsheets have a worksheet/tab for Entries which follows a format of one row per entry or "object", e.g. one row for each fish, bird or plant shown on the posters. Metadata about the entry/object is entered in columns, including an identifier, and

¹⁰<https://mozilla.github.io/nunjucks>

¹¹<http://w3id.org/ldac/pilars>

¹²<https://www.ldaca.edu.au/resources/user-guides/crate-o/convert-spreadsheet>

properties for name, speaker, sentence, translations, photo credits etc. Another worksheet, People, holds information about the speaker names and descriptions. In another worksheet, Files, the file paths to media assets are entered, primarily making a relationship as to which entry or person the file is related to or part of. Filling in the spreadsheets involved copying content from a document which was the content source of the original poster publication, and copying some content from PDF versions of the posters. Text was cleaned in a plain text file to remove formatting artifacts prior to pasting into the spreadsheet.

3.2 Organising media

Media assets were collated into folders/directories, with one top-level directory for each poster. Within these, audio and images directories were created for the respective media types, and inside *audio*, separate folders were used for call sounds, name and sentence recordings. This structure is arbitrary; media can be arranged in any folder structure. The path to the media file, relative to the metadata file, is used in the metadata as the file identifier. All photos were batch compressed in Adobe Photoshop so they loaded faster online and required less data by Gurindji users. All audio was normalised using Audacity.

3.3 Metadata conversion

Data can be packaged in RO-Crate format with Crate-O, a browser-based tool which can be used to enter metadata about files and to convert spreadsheets of metadata into RO-Crates. Metadata spreadsheets can also be converted to RO-Crates using a command line tool (`ro-crate-excel`¹³). In our case, we used Crate-O to convert the spreadsheets to RO-Crates. Once the spreadsheets were complete, a new RO-Crate was created in Crate-O, the metadata was then added by uploading the spreadsheet, and the RO-Crate was saved. This process generates an `ro-crate-metadata.json` file and the default HTML preview file.

3.4 HTML preview build

To build the deluxe HTML pages, the `ro-crate-html-lite` repository was downloaded and installed. The `RO-crate-metadata` JSON file and the directory of media assets were moved into the repository for

processing. A configuration file was edited to specify which HTML templates are used for which object type. The tool was then run without adapting the templates to confirm that everything was working, and the first draft site was built. In our excitement about how efficient and effective the production process was, we created a subdomain of an existing site and uploaded the HTML and asset files to a web server. We had a site!

We then iterated the build process, making changes to the templates to suit the content, adding a template for People and including the speakers on the home page. We do not have audio recordings of the fish calls, so the call audio player was removed for that site. After adapting the templates, the sites were rebuilt and republished.

3.5 Production timeframe

The activity occurred over a period of two weeks, with actual time taken approximately 12 hours. This time included an initial training session of two hours involving the three authors, one of whom had used the tool before. This session covered an introduction on how to use the tool, involving authors one and two, and how to complete the content spreadsheets. During the initial session, author three completed the content spreadsheets for the four posters. Author two then adapted the templates. Building the last two sites took approximately 20 minutes each due to the templates being complete. Much of the 20 minutes was spent on fixing file path errors in the spreadsheet and re-running the process, uploading files to the hosting server, and checking the site.

4 Results and Discussion

Packaging the Gurindji poster content in RO-Crate format provided the opportunity to use an existing tool to build websites to make the content accessible online. Four websites were built and published online, one for each of the Gurindji plant and animal posters. The websites are mobile-friendly and have no dependencies on code libraries or other software packages, making them extremely low-maintenance and likely to survive for many years without further work. However, to continue to be accessible, we must attend to paying for the domain name and hosting accounts. We chose to use an existing, paid hosting service that we use to publish related language material.

¹³<https://github.com/Language-Research-Technology/ro-crate-excel>

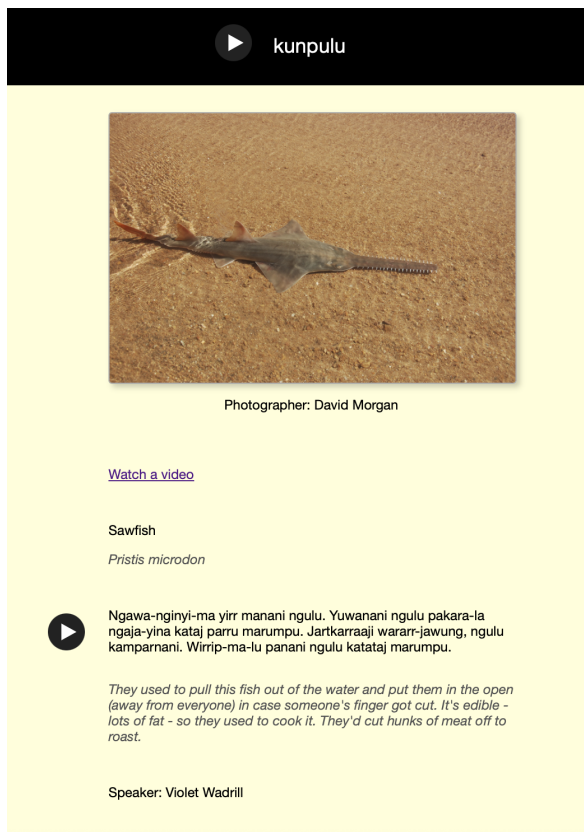


Figure 5: Yawu detail page for kunpulu (sawfish)

4.1 Archiving

Now in RO-Crate format, the content is also able to be accessioned to an archive such as PARADISEC (Barwick and Thieberger, 2012) for long-term safe-keeping. While the HTML sites that are now published online are in a format that will endure, there is value in having the data accessible in an archival repository with governance around storage and access, that doesn't rely on an individual to maintain domain name and hosting accounts.

4.2 Documentation

Prior to our work, the ro-crate-html-lite tool had little documentation. During the process we enriched the documentation, developing a step-by-step guide on how to populate the spreadsheet, download and install the tool, and customise the templates (see Appendix B). The documentation is intended to make the process more accessible for people to publish their own content.

4.3 Recommendations

4.3.1 Optimisation

At the time of publication, the images and audio in the four collections are low-resolution, compressed,

presentation quality files (JPEG and MP3). Future work for this project includes adding the higher-resolution versions (TIFF and WAV). The metadata schema we used has properties to denote files as being PrimaryMaterial and DerivedMaterial, which will be used to describe the high-resolution and presentation versions respectively. Templates could then be updated with additional functionality to download high-resolution files where available.

4.3.2 Design resourcing

Further benefits could be had with more design resourcing allocated to improve the site styling. The templates that we worked from display well on mobile and desktop. However, we noted some unkind cropping of images at varying screen sizes which occasionally caused the head of a bird to not be included in the cropped view.

5 Conclusion

First Nations Peoples in Australia are spearheading the maintenance and renewal of languages alongside Indigenous Ecological Knowledge (Tudor-Smith et al., 2024). Technology has an important adjacent role to play in assisting communities to renew languages. Leah Leaman, the Director of Karungkarni Art and granddaughter of Violet Wadrill sees the value in connecting the old with the new, where new technology can uphold old knowledge systems.

We are proud to see our Elders Violet Wadrill, Bidy Wavehill and Topsy Dodd Ngarnjal on the main pages. These web-pages will be increasingly important as we are losing our Kajijirri and Marlarluka (Old People). It means they can continue speaking to the ngumayijang (next generations).

Without interventions such as these websites to increase language transmission to younger generations, by the end of the century there could be a nearly five-fold increase in sleeping languages, with at least 1,500 languages ceasing to be spoken (Bromham et al., 2022)¹⁴. The importance of this work is underscored by the current UNESCO International Decade of Indigenous Languages (2022-2032).

¹⁴<https://www.bbc.com/storyworks/specials/unlocking-science/giving-new-life-to-old-languages-in-australia>

Limitations

Inherent in data-driven projects such as this are limitations around 1) the formats and structure of the data and tools, 2) skills and literacies required to prepare content and operate tools, and 3) resourcing to produce and maintain.

The templates currently available in the tool are limited in number and variety. Currently, the available templates all present data in lists and are not optimised for the presentation of tabular or network data. Work is underway to develop other templates including tabular, network and geospatial data display, to reduce the limitation of providing only list-based templates.

The task of adapting templates requires design and coding skills, potentially limiting people benefiting from having easy access to custom HTML previews when their data is in RO-Crate format. When we began this work, there was little documentation of how to use the tool, and no documentation about how to adapt templates. Our contribution of a "recipe" describing how to use and adapt the tool aims to reduce barriers to others, however design and coding skills are still required to modify templates.

Resourcing limitations can be significant for people to engage with digital publishing. In our work we were fortunate to be able to publish the four sites using an existing domain name and hosting. The cost of domain registration and hosting may be a limitation for people to publish their collections.

Ethics Statement

The original Gurindji plant and animal posters were co-designed with Gurindji people and organisations (see Section 2). Formal ethics was granted through the University of Queensland Human Ethics Committee. It was important that, as well as accurately reflecting Gurindji Life Ways, all of the information on the posters and websites was publicly open information. No secret, sacred or sensitive material was included. For example, we did not include the curlew bird call on the bird website because hearing the call can cause pregnant women to miscarry. Acknowledgments The original posters and collection of plant and animal information and resources was funded by the Central Land Council, Indigenous Language and Arts (ILA) through Karungkarni Arts, and Australian Research Council (ARC) DECRA (DE140100854, Meakins, UQ). The creation of the websites was funded by ARC

Laureate Fellowship (FL250100115, Meakins, UQ) and the Language Data Commons of Australia (LDaCA)¹⁵.

LDaCA is a co-investment partnership with the Australian Research Data Commons (ARDC) through the HASS and Indigenous Research Data Commons. The ARDC is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS).

References

- Linda Barwick and Nicholas Thieberger. 2012. *Keeping Records of Language Diversity in Melanesia: The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)*. University of Hawai'i Press.
- Lindell Bromham, Russell Dinnage, Hedvig Skirgård, Andrew Ritchie, Marcel Cardillo, Felicity Meakins, Simon Greenhill, and Xia Hua. 2022. *Global predictors of language endangerment and the future of linguistic diversity*. *Nature Ecology & Evolution*, 6(2):163–173.
- Ivy Kulngari Hector, George Jungurra Kalabidi, Spider Banjo, Topsy Nangari Ngarnjal Dodd, Ronnie Jangala Wirrba Wavehill, Dandy Danbayarri, Violet Nanaku Wadrill, Bernard Puntiyarri, Ida Bernard Malyik, Bidy Wavehill, Helen Morris, Lauren Campbell, Felicity Meakins, and Glenn Wightmann. 2012. *Bilinarra, Gurindji and Malngin Plants and Animals: Aboriginal Knowledge of Flora and Fauna from Judbarra/Gregory National Park, Nijburru, Kalkarindji and Daguragu, Northern Australia*. Bilinarra, Gurindji and Malngin People; Department of Land Resource Management.
- Peter Sefton, Moises Sacal Bonequi, Alex Ip, Michael Lynch, Amanda Lawrence, Julia Colleen Miller, Sam Hames, Marissa Takahashi, River Tae Smith, Annie Cameron, Mark Raadgever, Nick Thieberger, Ben Foley, Adam Bell, Janet McDougall, and Michael Haugh. 2024. *Protocols for Implementing Long-term Archival Repositories Services (PILARS)*. <http://w3id.org/ldac/pilars>.
- Stian Soiland-Reyes, Peter Sefton, Mercè Crosas, Leyla Jael Castro, Frederik Coppens, José M. Fernández, Daniel Garijo, Björn Grüning, Marco La Rosa, Simone Leo, Eoghan Ó Carragáin, Marc Portier, Ana Trisovic, RO-Crate Community, Paul Groth, and Carole Goble. 2022. *Packaging research artefacts with RO-Crate*. <https://journals.sagepub.com/doi/full/10.3233/DS-210053>.
- Gari Tudor-Smith, Paul Williams, and Felicity Meakins. 2024. *Bina: First Nations Languages, Old and New*. Black Inc.

¹⁵DOI: 10.3565/kq2v-9g52

APPENDIX

Appendix A Website URLs

Table 1: URLs for Gurindji plant and animal websites

Birds	https://jurlaka.ngumpin.org.au
Bush tucker	https://bush-tucker.ngumpin.org.au
Bush medicine	https://bush-medicine.ngumpin.org.au
Fish	https://yawu.ngumpin.org.au

Appendix B Documentation

During this work we developed a "recipe" style guide to building plant and animal websites using RO-Crate tools. The guide is available online at the following URL.

<https://github.com/Language-Research-Technology/developer-documentation/blob/main/tutorials/ro-crate-html-lite/bird-site/ro-crate-preview-doc.md>