

Going to Market with Deweyfish: The Journey from Partnership to Commercialisation

Kylie Black

University Library, The University of Western Australia, kylie.black@uwa.edu.au

Cornelia Hooper

The University of Western Australia, Australia, cornelia.hooper@uwa.edu.au

Ian Castleden

The University of Western Australia, Australia, ian.castleden@uwa.edu.au

Nader Aryamanesh

The University of Western Australia, Australia, nader.aryamanesh@uwa.edu.au

Harvey Millar

The University of Western Australia, Australia, harvey.millar@uwa.edu.au

Kylie Black, Cornelia Hooper, Ian Castleden, Nader Aryamanesh, and Harvey Millar, "Going to Market with Deweyfish: The Journey from Partnership to Commercialisation." *Proceedings of the IATUL Conferences*. Paper 4.

<https://docs.lib.purdue.edu/iatul/2019/value/4>

GOING TO MARKET WITH DEWEYFISH: THE JOURNEY FROM PARTNERSHIP TO COMMERCIALISATION

Kylie Black

University Library, The University of Western Australia, Australia
kylie.black@uwa.edu.au

Cornelia Hooper

Australian Research Council Centre for Excellence in Plant Energy Biology,
The University of Western Australia, Australia
cornelia.hooper@uwa.edu.au

Ian Castleden

Australian Research Council Centre for Excellence in Plant Energy Biology,
The University of Western Australia, Australia
ian.castleden@uwa.edu.au

Nader Aryamanesh

Australian Research Council Centre for Excellence in Plant Energy Biology,
The University of Western Australia, Australia
nader.aryamanesh@uwa.edu.au

Harvey Millar

Australian Research Council Centre for Excellence in Plant Energy Biology,
The University of Western Australia, Australia
harvey.millar@uwa.edu.au

Abstract

In 2016-7, The University of Western Australia Library partnered with researchers in the Australian Research Council's Centre of Excellence in Plant Energy Biology to produce cropPAL2, a database providing the subcellular locations for proteins in crops significant for food production. The project was funded by the Australian National Data Service as part of its High Value Collections program, with the team consisting of computational biologists, software engineers and a librarian. The project involved many hours of manual article evaluation and data extraction by specialists in the plant species included in cropPAL, and the team decided that developing in-house software could make managing the process of article evaluation by multiple people much easier. Key software features were that it prevented assessing the same article twice, simplified finding and adding new articles to the database, provided real-time access by international group members, and the cut and drop function facilitated saving images and notes. Use of this software represented a 90% saving in time and therefore salaries. The team realised the in-house software could be applied across many areas of research. Known as Team DeweyFish, the group embarked on the CSIRO's ON Prime program in 2018 to learn how to commercialise the software. This process involved the team generating and testing 15 hypotheses about researcher behaviour through conducting 66 one on one interviews with potential users. This data led to some significant insights, clarifying the needs of various user groups and refining the software specifications. An initial target market has been selected, and the team is now working towards developing a commercialisable prototype. This paper will discuss the role of the Library as a key player in this collaboration, a first for the University of WA, both in the innovative process and as a key driver in directing the development towards the wider benefit of researchers at UWA and beyond.

Keywords: cropPAL; partnerships; commercialisation; market research; software development; DeweyFish

The University of Western Australia Library partnered with researchers in the Australian Research Council's Centre of Excellence in Plant Energy Biology in 2016-17 to produce cropPAL2, a database providing the subcellular locations for proteins in crops significant for food production (<http://crop-pal.org/>). The project was funded by the Australian National Data Service as part of its High Value Collections program, with the team consisting of computational biologists, software engineers and a librarian.

The original cropPAL database was developed to index and link the rapidly growing collection of data in plant genomics and breeding methodologies that were difficult to discover and therefore under-utilised. The database provides open access to protein subcellular locations and interaction data for wheat, barley, rice and maize. cropPAL indexes data generated over 10 years by over 300 institutions as reported in the published literature, and linked to global plant catalogues such as EnsemblPlants.

Improving the discoverability of this data can have a profound impact as it has the potential to be used in innovative crop development to increase global food production. In cropPAL1, the project team had problems sourcing the literature containing the protein interactions and subcellular locations to be indexed in the database. This is a notoriously difficult type of information to find as it is rarely mentioned in the title or abstract, and the relevant data is typically found in the materials and methods sections of reports and articles. The Library was not involved in this project and a great deal of staff time and energy was spent sifting through large numbers of search results – for wheat, only 10% of the articles retrieved were found to contain relevant data.

The cropPAL2 project ran from January 2016-June 2017 and it added protein data for the crops of banana, canola, grapevine, potato, sorghum (used to feed cattle), soybean and tomato to the existing cropPAL1 database. The Australian National Data Service funded the project and as with many of their projects, encouraged collaboration between the scientists and the Library. In this case the partnership was especially logical given the history of problems with searching, and that the relevant subject library was physically located next to the Centre for Plant Energy Biology.

This was the first time that the UWA Library has been a formal partner in a research project, and with our values [University Library, 2014, p.2] of collaboration, efficiency, responsiveness and innovation it was a good fit. Involvement in cropPAL2 provided an opportunity for librarians to contribute directly to a research project that has economic, social, health, environmental and academic impact. It also allowed the Library to assist the University in meeting a number of objectives in its Strategic Plan relating to internationally renowned research, including undertaking “research across all our disciplines, focused on issues of relevance to our communities and industries, while generating understanding and solutions of global value” and building “problem-oriented multidisciplinary teams” [The University of Western Australia, 2013, p. 7].

Traditionally, for a project gathering a large amount of literature such as this, Library involvement might have been limited to a couple of appointments with researchers to develop the search strategy, then possibly involvement with managing data and minting a DOI at the end of the project. With the Library as a partner in the project, there was involvement throughout the entire cycle to mutual benefit, as shown in Figure 1.

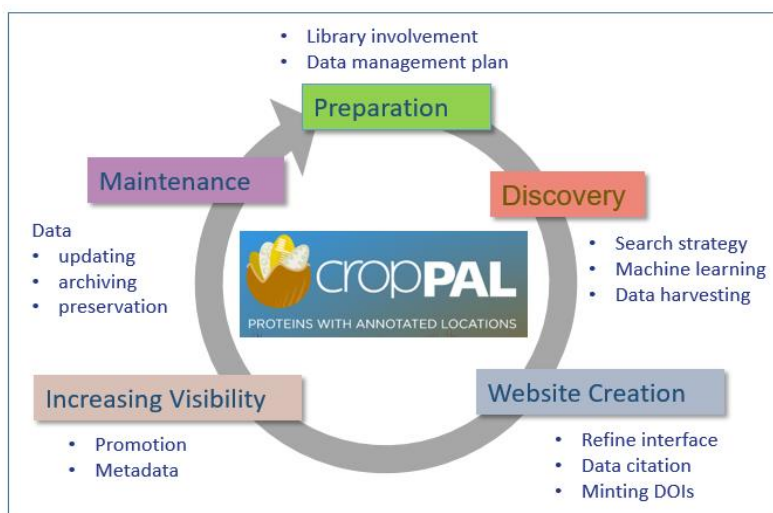


Figure 1 Library involvement in cropPAL2

The Library contributed to cropPAL2 through librarian expertise in searching and retrieving scholarly literature, formulating and executing complex search strategies, bibliographic database tools, metadata, and research data management. These skills were critical to the success of the project in terms of increasing the efficiency of the cropPAL2 data collation process and ensuring that the dataset was described, stored, and made available in an open access format. There is now an opportunity to further promote these skills and expertise to other parts of the UWA research community.

Additionally, the project provided a valuable and rare opportunity for a librarian to be fully embedded within a research team and within the research process. This resulted in an increased confidence in the librarian to be able to directly and positively contribute to the research process.

Two sections of the UWA Library were involved: Library Engagement, and Research Publication and Data Services. Both were involved throughout the 18 months of the cropPAL2 project and Engagement has ongoing involvement in updating the database and studying its impact as well as commercialisation.

Benefits for Library Engagement

This collaborative project demonstrated that for searching, capturing and dealing with permissions around literature, the collaboration between scientists and librarians was highly beneficial. Being embedded in a research team led to insights into how research projects are conducted and especially the need to balance time spent doing research with other tasks: the reporting required, attending project meetings, conferences and writing papers. At a more detailed level, it was very informative to see how computational biologists search for and manage biological information. Insights were also gained in relation to how co-authorship may be used as currency, to acknowledge the contributions of other researchers to the project. It is an especially valuable currency when the researcher has a high H index and track record of being highly cited.

Forming a close working relationship with the scientists in Plant Energy Biology has led to assistance in other areas, such as librarians providing data for grant applications and consultation with the Library for other projects involving large datasets. It is also very useful to have Library champions who are always willing to assist with any queries we might need to ask research staff.

Benefits for Research Publication and Data Services

In research data management there were benefits for the Library in working closely with real researchers with real data, such as versioning, allocating DOIs, and how to address the question of a new version versus an update. Scientists involved in cropPAL represented the senior user group for a Library data repository project reviewing dataset maintenance, and

advising in a repository platform migration project aimed at improving the hosting of research datasets.

The scientists in Plant Energy Biology were keen to maximise the discoverability of cropPAL, and with this in mind Library staff worked on the metadata and online discoverability, both through Google and subject-specific resources. This increased awareness of research data management within Plant Energy Biology has led to their use of better systems for storing research data and making it readily available through Research Data Australia.

Benefits for the Researchers

Librarian expertise in search strategy and data management was part of the original project proposal and resulted in a number of contributions. In the area of literature searching, the librarian was able to identify functionality within the EuropePMC (Europe PubMedCentral) database (<https://europepmc.org/>), which facilitated specific section searching techniques to improve the relevancy of overall search results. The cropPAL database relies on extracting relevant data from the published literature, so efficient searching is critical. While the researchers were already aware of the existence of EuropePMC, librarians were able to develop search strings that bypassed the functional limits of the interface. Library staff analysed how the section search could be developed further by referring to the literature [Kafkas,2015] and then applying these learnings to the problem of finding protein data for the selected crop species. Librarians were also able to leverage off the results of the literature searches that were carried out by researchers in the creation of the original cropPAL. The scientists had retained records of all the literature that had been manually assessed for cropPAL, which made it an excellent database to test the effectiveness of search strategies.

The collaboration between researchers and librarians was more complex in relation to research data management, as the cropPAL dataset was used as an example in relation to data promotion, storage, versioning and management. This was part of a Library project to decommission the internal system and migrate the data to the UWA Research Repository. In particular, discussion of the issue of data inheritance led to new approaches towards how data is submitted and linked to UWA staff to ensure ongoing data maintenance and accountability. Further discussions around the data archiving and security influenced the organisation of server and back-up server infrastructure within the wider university system. The benefit for the researchers was in understanding more about these issues and being able to apply them to the cropPAL dataset as it was being developed.

The project's final report best summarises the benefits to the researchers: "the positive change in attitude towards the value of a tight collaboration between the local institutional library services and lab-based scientists" [Hooper, 2017, p.3].

Both cropPAL projects involved many hours of manual article evaluation and data extraction, and the team decided that developing in-house software could make managing the process of article evaluation by multiple people much easier. Key characteristics of the software:

- Not assessing the same article twice
- Easy to find new articles – saved searches in PubMed and EuropePMC
- Simplifies adding new articles to cropPAL
- Provides real-time access by group members to the database
- Useful links to access the information regarding a paper such as full-text, the journal and number of citations
- Cut and drop function to save images and notes
- All information is tabulated so easy to import/export data or run queries.

Significantly, the cropPAL database was developed in a fraction of the cost (\$200k) for manual curation, representing a 90% saving in time and personnel salaries. The team realised the in-house software was highly innovative and could be applied across many areas of research - and may be worth commercialising.

ON Prime – first steps to commercialise the cropPAL in-house software

A group of scientists, a software engineer and the librarian involved in cropPAL formed a four person team to commercialise the software, known as DeweyFish. The group successfully applied for the CSIRO's (Commonwealth Scientific and Industrial Research Organisation) ON

Prime program for late 2018, to begin the process of market research, making industry contacts and learning more about potential sources of funding and issues around intellectual property. The role of the librarian was two-fold: to bring technical knowledge of literature searching, and to bring the perspective of researchers in fields of study outside of plant biology.

Once accepted, there was no financial cost to participants and the organisers facilitated team work by flying in one of the team members from Adelaide. However, ON Prime is a significant time commitment, with five face-to-face sessions run over six full days, spanning eight weeks. There is a large time commitment between sessions with further work as a team and a target of 100 interviews with potential customers or users to be conducted within this eight week period. This data is invaluable for researching customer pain points and market validation. For every team in the DeweyFish cohort the interviews resulted in a pivot point, where the direction of the project changed significantly.

Interviews

The backbone of the ON Prime program is each team developing several hypotheses to test by conducting a large number of interviews with a range of potential customers. Team DeweyFish had 15 hypotheses and these evolved as the interviews progressed. Early in the program participants learned about interview technique and formulated interview questions. Some of Team DeweyFish's hypotheses were:

- Advanced search techniques (going beyond keywords in title and abstract) are needed for precise searching
- Researchers waste time looking at studies more than once
- Scientists do not like to take notes and collect data out of individual studies into spreadsheets
- If it was easier to find and parse, scientists would reuse more of the existing data
- Data is not shared or retained after the research is completed

As a result of the 66 interviews conducted across a range of fields of research, Team DeweyFish had one pivot, realising the software would have limited applicability in the humanities due to the need to refer to historical documents and different formats of material. The interviews led to other significant insights, such as:

- Managing the literature is more difficult than initially finding it
- Lack of awareness of scholarly databases with researchers using Google or Google Scholar instead
- Research data is not widely shared or is perceived to be a practice for senior researchers.

From the library perspective, accessing the notes from interviews conducted by non-librarians is valuable in determining how researchers find, use and manage information when they have not sought assistance from librarians.

The Team has decided to focus its prototyping efforts in the area of medical research where the technology could greatly assist with systematic reviews, both in the original review and particularly in updating existing reviews. The team members with expertise in software development are working on the functional requirements as a result of the interviews conducted in ON Prime and by closely examining other products that do part or all of what DeweyFish does. Other team members are investigating sources of funding and will be conducting more specific interviews within the chosen market segment.

Throughout the cropPAL2 process and now commercialising the software as part of Team DeweyFish, the Library has played a key role in providing expert knowledge of literature searching and management, and research data management. Being an equal member of the research team also led to benefits in learning how scientists conduct their research and in building relationships with Faculty to mutual and ongoing benefit.

References

- Hooper, C. M. (2017). *Final Report: High Value Collections Program. The University of Western Australia - ARC Centre of Excellence in Plant Energy Biology - The compendium of cropPAL 2 [unpublished]*.
- Kafkas, S., Pi, X., Marinos, N., Talo, F., Morrison, A., & McEntyre, J. R. (2015). Section level search functionality in Europe PMC. *J Biomed Semantics*, 6, 7. doi:10.1186/s13326-015-0003-7
- University Library. (2014). University Library Strategic Directions 2015-2020. Retrieved from http://www.library.uwa.edu.au/__data/assets/pdf_file/0003/2778114/University-Library-Strategic-Directions-2015-2020.pdf
- The University of Western Australia. (2013). UWA 2020 Vision Strategic Plan: 2014-2020. Retrieved from http://www.web.uwa.edu.au/__data/assets/pdf_file/0010/2538343/114085-VICCHA-StrategicPlan-v3.pdf