KULA
Knowledge Creation, Dissemination, and Preservation Studies

Crowdsourcing Downunder

Rachel Hendery et Jason Gibson

Citer cet article

Hendery, R. & Gibson, J. (2019). Crowdsourcing Downunder. *KULA*, *3*(1), 1–12. https://doi.org/10.5334/kula.52

Résumé de l'article

In this paper we report on the experience of two research projects that intended to experiment with crowdsourcing models for opening up their scholarly materials to the wider public. Both the *Howitt & Fison* project, and *Mapping Print; Charting Enlightenment* were designed to take into consideration particularities of the Australian academic environment: in the former case, sensitivities around materials relating to First Peoples; in both cases, geographical distance from potentially interested communities, and the difficulties of formal recognition and categorisation of time spent on activities that lie at the intersection of research and outreach. They had similar challenges in terms of needing to process a large amount of data before analysis and progress towards the projects' main research goals could begin. They also had similar goals in terms of eventual use of the project data, for example, making historical texts available online, and producing maps, networks, timelines and digital exhibitions of images and texts. In the end, one project has found crowdsourcing invaluable for building connections with interested publics the other discovered that crowdsourcing was not necessary to produce the results the project needed, and has moved away from this to focus its efforts instead on the linking of existing data and automation of structuring and categorisation. This paper discusses how the projects came to take these different directions, and how the above-mentioned Australian contexts contributed to their evolution.

RESEARCH ARTICLE

Crowdsourcing Downunder

Rachel Hendery[1] and Jason Gibson[2]

[1] Western Sydney University, AU

[2] Deakin University, AU

Corresponding author: Rachel Hendery (r.hendery@westernsydney.edu.au)

In this paper we report on the experience of two research projects that intended to experiment with crowdsourcing models for opening up their scholarly materials to the wider public. Both the *Howitt & Fison* project, and *Mapping Print; Charting Enlightenment* were designed to take into consideration particularities of the Australian academic environment: in the former case, sensitivities around materials relating to First Peoples; in both cases, geographical distance from potentially interested communities, and the difficulties of formal recognition and categorisation of time spent on activities that lie at the intersection of research and outreach. They had similar challenges in terms of needing to process a large amount of data before analysis and progress towards the projects' main research goals could begin. They also had similar goals in terms of eventual use of the project data, for example, making historical texts available online, and producing maps, networks, timelines and digital exhibitions of images and texts. In the end, one project has found crowdsourcing invaluable for building connections with interested publics the other discovered that crowdsourcing was not necessary to produce the results the project needed, and has moved away from this to focus its efforts instead on the linking of existing data and automation of structuring and categorisation. This paper discusses how the projects came to take these different directions, and how the above-mentioned Australian contexts contributed to their evolution.

Keywords: Crowdsourcing; anthropology; history; digital humanities; Australia

Introduction

In this paper, we report on the experience of two Australian-Research-Council-funded projects with which the authors are involved, and which intended to experiment with crowdsourcing models for opening up their scholarly materials to the wider public.[1] Both cases are designed to take into consideration particularities of the Australian academic environment: in one project, sensitivities around materials relating to First Peoples; in both cases, geographical distance from potentially interested communities, and the difficulties of formal recognition and categorisation of time spent on activities that lie at the intersection of research and outreach.

Crowdsourcing is the practice of inviting people beyond the immediate project team to contribute to the work in some way. It is usually volunteer work, and frequently open to anyone who wishes to participate. Most commonly, the tasks in question can be completed online through the web browser, remotely, but it is not uncommon to also include activities or training at a central location such as a museum or university for those who desire face-to-face engagement. While on the face of it, this kind of distribution of labour across an online volunteer workforce is reminiscent of the use of 'playbour' (Kücklich 2005) or 'hope labour' (Kuehn & Corrigan 2013) by companies that use Amazon Mechanical Turk or competition-sourced design work, most academic crowdsourcing projects have been quick to point out the differences (e.g., Ridge 2012,

---

2013a; Hedges & Dunn 2012, 40; Terras 2016). Most cultural heritage projects that employ crowdsourcing define 'crowd' rather differently, for example. It is often the case that the volunteers are sourced from a small group of people who already have a relationship with the institution in question (Terras 2016), thereby building community and offering the opportunity to work towards a shared goal, both of which can be major motivations for participation (Rockwell 2012, 148). This 'nichesourcing' (de Boer et al. 2012) or 'community sourcing' (Phillips 2010) was the plan for both of the projects discussed in this paper.

Howitt and Fison's Archive

The project 'Howitt & Fison's Archive: Insights into Australian Aboriginal Language, Kinship and Culture' (hereafter referred to as 'the Howitt & Fison project') is an Australian-Research-Council-funded project led by Associate Professor Helen Gardner (Deakin University) that began in late 2016 and will run for three years. Howitt and Fison were anthropologists who were active in Australia in the late nineteenth century. Howitt was a magistrate in the Gippsland region of Victoria in south-eastern Australia and Fison was a Methodist missionary with experience in Fiji and Australia. Together, they carried out some of the earliest anthropological research in Australia and left extensive archival materials on language, kinship, and social organisation. Some of the most significant materials on Australian First Peoples from the colonial period are contained in the tens of thousands of handwritten, unpublished manuscript pages from their notebooks and correspondence. They published a landmark text *Kamilaroi and Kurnai* (1880) as well as a few smaller items (Howitt 1887; Torrance 1887), but the intended publication of the bulk of their collected materials never eventuated (see Gardner and McConvell 2015). The publications that did appear (Howitt 1904), as well as their manuscript notes, provided direct inspiration to early ethnographers in Australia and served as a foundation for later structuralist and functionalist approaches in anthropology (acknowledged by Radcliffe Brown 1931; Stocking 1995; Stanner 1972; and Needham 1967). Their archives include:

· over 7000 pages including correspondence, field-notes, circulars in Museum Victoria's Howitt Papers collection.
· 9 boxes of correspondence, articles, and primary data (13,200 manuscript pages including 400+ pages of images and postcards) in State Library Victoria's *Howitt Papers*.
· 6 Letterbooks, 800 pages each, in National Library of Australia *The Papers of Lorimer Fison*.
· approximately 800 pages of correspondence in Rochester University, USA.
· 2461 pages of Fison's correspondence together with 1000+ pages of drafts and primary data in St Marks Theological College Library in Canberra.

They contain important cultural and language histories of people of south-eastern Australia and Fiji and also material relating to other First Peoples of Australia and the Pacific. The records they kept are imbued with the cultural expertise of their informants, notwithstanding that Howitt and Fison's work was certainly shaped by the questions of colonial administration, social evolutionary theory, and missionising. The word lists, detailed information about kinship and social relations, traditional stories and songs, some recorded in original languages are of immense value and could only be gathered by close engagement with Australia's First Peoples. We treat these materials as artefacts of deep and prolonged colonial encounter.

The aim of this project is to return Howitt and Fison's papers to First Peoples communities and to the wider public in a way that brings to the fore their rich linguistic and ethnographic data as well as the relationships and histories of their original production. We aim to do this by describing and interpreting their contribution to contemporary linguistic and anthropological scholarship and to First Peoples' heritage, but also by building upon their records to provide new analyses of systems of First Peoples' cultures and languages. As with other recent projects aimed at bringing together and annotating early anthropological works (Glass 2014, 2017; Batty and Gibson 2014) the project mobilizes digital media to link together disparate collections, scholars, and First Peoples communities. This project is being carried out in partnership with Gunnaikurnai communities in Victoria through the involvement of the Victorian Aboriginal Corporation for Languages (VACL), including Partner Investigator Paul Paton, himself a Gunnai and Monaro man and the involvement of First Nations Legal and Research Services (formerly Native Title Services Victoria).[2] The Howitt and Fison collection contains important information for First Peoples from all over Australia but as the richest ethnographic materials derive from Howitt's time amongst the Gunnaikurnai population in

---

[2] Howitt spelled the name of this group as Kŭrnai, but other spellings, including Gunnai, are found in other nineteenth century sources, and the community now uses both spellings in the combined term Gunnaikurnai.

Gippsland, we have prioritised partnership with this group. The project also incorporates a (paid) program for community members wishing to learn more about archiving and curation and examine the cultural, historical and linguistic contents of the materials within the institutional setting of a library or museum. A PhD student from the local community is also expected to begin in 2019 and wider consultations with other Aboriginal communities across Australia are also underway.

The transcription of the archives is mainly being carried out through a crowdsourcing initiative, which is described further in the fourth and fifth sections, below. This process is being overseen by Dr. Jason Gibson, a long-time research associate at the Melbourne Museum and the *Howitt & Fison* project manager.

Mapping Print; Charting Enlightenment

The Mapping Print; Charting Enlightenment project (MPCE) is an Australian-Research-Council-funded project led by Professor Simon Burrows at Western Sydney University that began in mid-2015 and will run until mid-2019. The project aims to understand the European Enlightenment by examining the circulation of books in Western Europe, including piracy, illegal publishing, the French book trade, and reading habits. A large part of the project is concerned with compiling and curating large datasets about book publishing and sales networks, through the digitisation and structuring of the data present in logbooks and manuscripts.

An earlier project by Professor Angus Martin and collaborators compiled a large dataset of French prose fiction from 1700–1800 including a global survey of library holdings. The MPCE project is integrating these data into its comprehensive database and thus linking literary output and subject with data on the travels, writings and intellectual networks of authors, publishers and readers.

The aims of this project are to understand what the French were reading before the revolution of 1789–1799 and how their reading might have shaped ideologies and events; to understand what books were produced and whether there was such a thing as an integrated national or international publishing market; and to understand the extent and nature of religious publishing and the illegal book trade.

The process of digitising, annotating and linking data from private collections, booksellers and library catalogues was initially conceptualised as having a crowdsourced component, beginning with input from the Bibliothèque Nationale's collection of bookseller catalogues from 66 European towns (1770–1790). Eventually, the decision was made to prioritise machine-based automation instead of human labour, the reasons for which will be discussed below.

Comparison, Challenges, and the Australian Context

The two projects described above present an interesting comparison, because they both faced some similar challenges, and both intended to rely on crowdsourcing, but eventually took two very different directions. Moreover, there are similarities in terms of their eventual plans for the data they are compiling and producing, as both projects anticipate geospatial outputs such as maps, other data visualizations such as networks and timelines, and hyperlinked, annotated textual and image-based online exhibitions. We will begin here, however, with the differences between the projects' potential pools of crowdsourcing volunteers.

The *Howitt & Fison* project team anticipated that their pool of crowdsourcing volunteers would most likely be drawn from amateur historians, local history societies, genealogists, community members with First Peoples' backgrounds, and other members of the general public. As it eventuated, most participants are either museum volunteers, researchers working for First Peoples' organisations or with government departments, and academics (the 'community sourcing' approach described by Phillips 2010). Members of the research team also take part in transcription and checking of the transcriptions and a small number of Gunnaikurnai people have started making transcriptions. While the age range varies, the transcription volunteers at the museum are most typically older women (aged 50–70) with an interest in local history, First Peoples' culture, and heritage related issues. The *MPCE* team, on the other hand, expected that those who would be most interested in participating in their crowdsourcing project would be academic researchers planning to use the database for their own scholarly purposes, wishing to add annotations and expanded datasets of relevance to these. The data entry portal for MPCE was never opened to the wider general public.

One of the major challenges for both projects has been structuring the data. There is always a tension between ease of data entry, or coding, and the usability of the resulting dataset. To exemplify this from the *Howitt & Fison* project, the simplest approach for our volunteers would be to transcribe a notebook (e.g. see **Figure 1**) word for word into software they are familiar with such as Microsoft Word—with best guesses taken whenever a transcription is in doubt, and no attempt to mark up any non-word elements on the page besides punctuation. Every aspect of the task that deviates from this brings a small additional burden in terms of training, giving and recalling instructions, and familiarity with technology. At the other extreme, for usability,

Figure 1: Example of a 'kinship schedule' or word list of kinship terms (from Fison's Letterbook 3, 1872), showing the necessity for some structured mark-up of the notebook transcriptions.

we would ideally like to have every element of the data structured and encoded, using for example TEI markup of every element, tagging words structurally and semantically, including for a set of important categories such as place names, personal names, kinship terms, words from languages other than English, etc. We would like to have the crowdsourcing volunteers encode the text in a way that represents the formatting of the page—is it in a table, or in columns? What has been crossed out or written over? Are there scribbles in the margin? If there is a diagram or drawing on the page, we would like to transcribe a written description of it together with any text contained by it. If multiple diagrams or drawings appear in the manuscripts that are similar, it would be good to develop a controlled vocabulary for tagging them as particular types of image. Howitt used a convention of drawing a line down the centre of a page in his notebooks often to indicate that he had published the content of the text, so we need a way to tag pages based on whether or not this line appears.

Of course, doing all of this is unrealistic with a pool of volunteers from the general public who moreover each have slightly different idiosyncratic conventions in their transcriptions. Ultimately, we have settled for a small set of markup conventions that allow especially important categories of text to be tagged, and the use of software (*From the Page*) that automatically links such text and creates indexes and semantic networks (via linkage of concepts that are regularly found in proximity to each other.) The direct output of the crowdsourcing volunteers' work will need significant processing, cleaning, and analysis by the research team before it can be integrated into new or existing linguistic and historical databases, or used to produce data visualisations.

The MPCE project, on the other hand, planned an approach in which crowdsourcers (who, in this case, are academic researchers) input their digitisations directly into the project database. This made the design of such a database, the structuring of pre-existing data, and the design of the data entry tools particularly challenging. The tools that were created were designed as relatively traditional-style database entry forms. While the user interface is well designed, users still need some familiarity with data entry principles and a solid understanding of the structure of the MPCE database to make use of it, which is a further factor that might limit the general public's ability to engage with the data entry side of the project.

In general, data structuring and linking has been a much more serious challenge for the MPCE project than for *Howitt & Fison*. The MPCE project relies on a number of very large datasets that originate from different projects with different data schemata, goals, and user bases. The MMF database mentioned above was constructed to run on a DOS operating system and runs today in the DOSbox emulator. The integration of this dataset into the MPCE database has only become feasible thanks to scripts and workflows painstakingly constructed by Jason Ensor and, more recently, Katie McDonough. Despite this sterling effort there are remaining intractable problems, such as notes fields that cannot be exported and characters that do not output correctly. Other datasets have been compiled in spreadsheets with different fields, tables, and column names than those in the main database, and the main database itself contains tables that partially overlap or repeat data, but are not always linked in the ways a user might expect.

In the process of developing scripts and tools for incorporating these various datasets into the *MPCE* main database, it became clear to the project team that linking existing databases was likely to be a better long term approach than trying to restructure datasets and re-enter them into *MPCE's* database. The project team has therefore now turned its main attention to the challenges of restructuring the existing MPCE data along principles of Linked Open Data and finding ways to connect it to material in Linked Data formats elsewhere. Secondarily, the project is investigating whether we can automate the process of tagging and categorising the remaining MPCE data. Rather than asking crowdsourcers to manually complete the process, it should be possible to train a neural net on the entries that have already been tagged and then have it automatically classify the remaining entries. We may then choose to turn back to crowdsourcing at this later stage to check and refine the automatically generated tags.

The data structure challenges described above make *MPCE* a very different kind of project than *Howitt & Fison*, which has not yet tried to grapple with the complexities of integrating multiple datasets from different sources into a single resource. This may still become an issue, however, as *Howitt & Fison* intends to eventually attempt to connect the new crowdsourced data with several existing datasets, the main one being *AustKin,* the database of Australian kinship (Dousset et al. 2010).

Rather than data structuring and linking, the largest anticipated challenge for *Howitt & Fison* is the sensitivity around Australian First Peoples' materials. This is mainly being handled through the co-researching and consultation embedded in the project, as described above, but it also has to be taken into account in the workflow for releasing material for digitisation and crowdsourcing. Following an initial survey and cataloguing of the materials, Jason Gibson has been working closely with various Gunnaikurnai elders to identify manuscripts that contain any sensitive materials. Such materials might contain detailed information relating to ceremonies, delicate information about individuals, or describe gender-restricted ritual practices that cannot be made public (Gibson forthcoming). Even information, which may initially seem benign, such as the recording of placenames, need to be handled with care. The transcribers recently worked diligently to produce transcriptions of some of Howitt's notebooks, which because of their poor quality (scrawled handwriting and faded pencil) were left unexamined by past researchers. Once the transcripts were examined by the research team, we realised that Howitt had recorded a number of place names used by the original inhabitants of the Melbourne area, the Wurundjeri. As the names had never been made public before and because they could potentially be used in land claims, the research team is now in consultation with the Wurundjeri Council about how and when these details might be published.

Having transcripts of this material has therefore been a critical first step in unlocking the potential of the archive with an eye to these sensitivities. As deciphering the handwriting in the manuscripts is not always an easy task, sometimes the content of a page cannot be completely known until after crowdsourcers have worked with it. It is thus often necessary for the material to be transcribed first before its classification as 'sensitive' (or otherwise) can be accurately determined. The completed transcripts are, however, not released on the main public *Howitt & Fison* project website at this stage. Rather, they are checked by the project research team and discussed with representatives from the Gunnaikurnai community before being further circulated.

A related consideration for this project is a risk that it be seen as scholars asking for donations of free labour in order to unlock information that, by rights, belongs (at least in part) to local Gunnaikurnai people and has been kept by the same colonial institutions that now benefit from this labour. This is one reason why the collaboration, training and co-research with Gunnaikurnai people has been planned as a quite separate part of the project from the crowdsourcing component. The trainees, consultants, and PhD student are all paid for their time and expenses, and have a degree of control over the project development that the unpaid crowdsourcing volunteers do not have. While some Gunnaikurnai community members have chosen to participate in the crowdsourcing, they are not necessarily the target group for this volunteer work, and there are more central roles interested members of this community can play in the project.

In his discussion of crowdsourcing, Rockwell (2012) points out the role it can have in decentralising authority, noting that croudsourcing 'erases differences between professional scholars with degrees and those who are self-taught amateurs without formal training' (152). The project design for *Howitt & Fison*, similarly, seeks to erase distinctions between professional scholars, whose expertise is widely recognised and rewarded, and Gunnaikurnai elders whose (often far deeper) local expertise in many cases is not formally acknowledged by academic institutions or Australian society more broadly. A recent event organised by the project involved hosting four Gunnaikurnai men on a tour of the Melbourne Museum and State Library of Victoria's collections. Over two days the group spent considerable time looking at original manuscript items, but also read through the online transcripts that had been produced by the volunteers. This viewing elicited

lively discussion about historical events, the linking of contemporary Gunnaikurnai families with those named in the archive, and queries about how nineteenth century anthropology described Gunnaikurnai religious beliefs. Returning to the archival sources in this collaborative manner highlighted, as Glass (2014) argues, that 'all anthropological knowledge is co-constructed to a significant degree, in as much as it emerges from social encounter and interaction that is based on relations of consultation and complicity between scholars and research associates' (19–20). In acknowledging and respecting the dialogical origins of this material, it was thus imperative that we responded to, and integrated into our project, any present day Gunnaikurnai issues or concerns.

The MPCE project, of course, does not encounter the same potentially sensitive material. While some of the data in the project identifies individuals as being involved in illegal activities—piracy, for example—or in publishing or reading pornography, this is mitigated by considerations such as the greater age of the materials, the difficulty of identifying most descendants of the people involved, and most importantly, the different socio-cultural contexts of a majority demographic in Western Europe versus an Australian First Peoples community in postcolonial Australia.

Interfaces

The *Howitt & Fison* project experimented with a number of off-the-shelf crowdsourcing platforms with varying degrees of customisability. The initial intention was to use the Digivol platform that has been developed at the Australian Museum in Sydney and is already being used with great success for other Museums Victoria projects. Initially designed to crowdsource the transcription of entomology labels, the Digivol software has since expanded its functionality to incorporate materials from the humanities and social sciences. The advantages of Digivol were that it hosted a ready community of engaged online volunteers (over 20,000 people), and a stable, established interface that is relatively straightforward to use. It also features inbuilt motivation/gamification tools such as leaderboards and badges. The main disadvantages, however, were a lack of customisation options that made it harder for us to implement the kind of text markup we needed, and difficulty accessing a structured export of the transcribed data.

After experimenting with several other alternatives, we identified *From the Page* as a platform that offered great potential. It allowed for greater customisation of the transcription interface, easy markup solutions, a variety of data outputs (including TEI), and the facility to automatically link generated lists of marked up terms belonging to particular categories of interest (e.g., place names, personal names, kinship terms, etc.) and concept maps (see **Figure 2**). The concept map produced for the individual named 'Big Charley,' for example, immediately illustrates his links to a number of language groups in the Gippsland area (Brabralung, Bidwell etc), physical locations (Snowy River, Maneroo [Monaro]) and a keyword from the Gunnaikurnai language 'Thundung' (totem/Dreaming), as well as the names of other Gunnaikurnai individuals (King Charley and Billy Jumbuck).

In the end we uploaded pages to both *Digivol* (here) and *From the Page* (here), and compared the uptake statistics, and the usefulness of the outputs, and decided that the flexibility of *From the Page* was worth the lower user numbers. Initially 70 items were uploaded to *Digivol*. Most items were several pages long, so this represented several hundred pages. Within two days, without any attempt at recruitment of volunteers, we found that 25% of them had been transcribed. Although we have moved on to using *From the Page*, we left the initial items on *Digivol* and the transcriptions were quickly completed. On *From the Page,* in around two months just over 350 pages have now been transcribed by a team of 17 contributors. Many of these pages have now also been cross-checked by the project's research team, and indexed to link important keywords. One of the challenges we are currently addressing is to create a 'soundex,' or to regularise the spellings of the terms from Australian languages, so that they are searchable without necessarily knowing the idiosyncrasies of Howitt and Fison's own orthographic conventions in advance.

For MPCE, on the other hand, the complexity and specialised nature of the data meant that data entry and annotation was 'niche-sourced' by a relatively small group of people associated with the project, rather than open to interested collaborators more generally. The idea was originally to open access to the manuscript tool to more researchers eventually, once the integration of the MMF dataset was complete and the structure of the main database tables finalised. The interface for this entry of new data and annotation was custom designed and built for the project by Jason Ensor, and a screenshot of it is shown in **Figure 3**.

This is a rather different kind of crowdsourcing than that of most digital humanities projects (cf. Terras 2016), in that it is not mainly about transcription of text, but structuring and categorising the information in that data: building out a database. While the interface is therefore slightly more complex than a pure transcription interface, the results of the tasks completed are more immediately apparent, and the individual
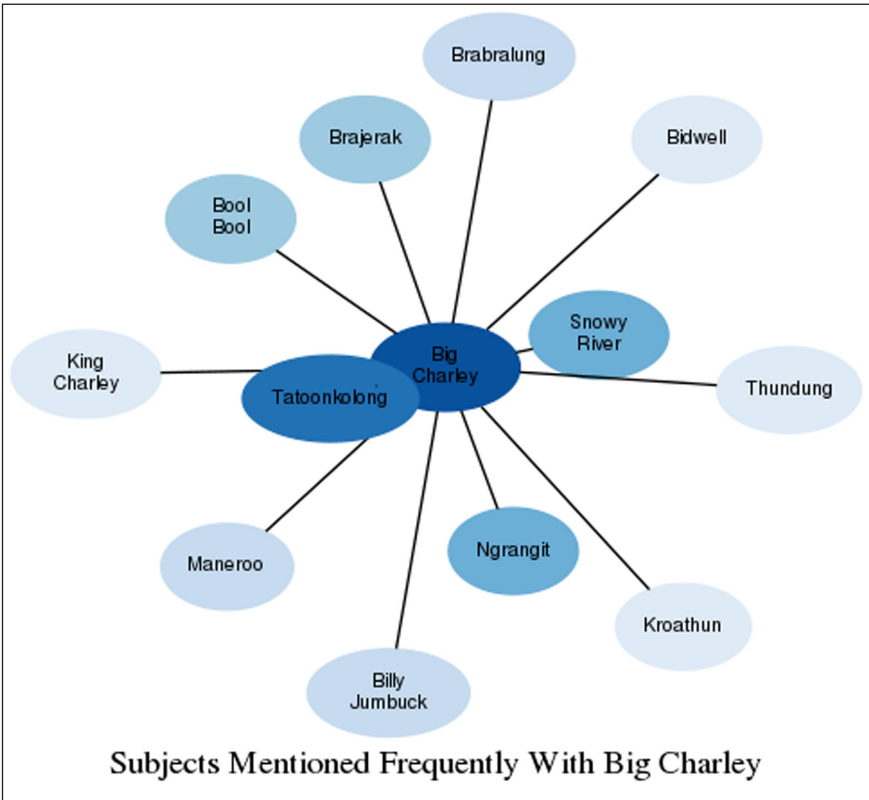
Figure 2: A conceptual network automatically generated by From the Page. Upon clicking any concept in this network, that concept becomes the centre of a new network, and all the subjects mentioned frequently in proximity to that new central concept are shown. Thus the whole network of concept relationships can be navigated.

Figure 3: Manuscript data toolkit, interface created by Jason Ensor (2015).

tasks (entering a single book or sales event) are quicker to complete. This may be more motivating for participants than the slow, incremental results of manuscript transcription (c.f. Rockwell's concept of 'small autonomous tasks' (2012); or Ridge's discussion of 'scaffolding' (2013b)). However, it also means that greater familiarity with the materials and training in how to use the interface was necessary. A small number of researchers closely associated with the project were trained in how to use this interface and have been entering data, but the portal has not, after all, been opened to the general public, and in the next section we will discuss the reasons for this.

Democratisation of Research, Community-building, and Motivations for Participation

One of the challenges with crowdsourcing projects is managing expectations – both those of the project team and those of the crowdsourcing volunteers themselves. It is very seductive to imagine that a mysterious public is waiting, enthusiastic about our research topics, and ready to leap into action like the shoemaker's elves to carry out our tedious and time consuming tasks while we focus on abstract, higher matters of scholarship and conceptual design. In reality, the process of engaging with crowdsourcing volunteers is complex and time-consuming, as is the set-up of platforms, interfaces and appropriate data structures. Checking transcribed material, processing it further, and analysing it all takes a great deal of time. If researchers were doing the initial transcription themselves, this processing and analysis would happen alongside the transcription rather than as a separate activity, and in many ways that would be more efficient. As a comparison, one of the largest, longest running academic crowdsourcing projects, *Transcribe Bentham,* estimates that:

> Had the Research Associates been employed purely to transcribe manuscripts on a full-time basis, they could have transcribed about 5,000 manuscripts between them over twelve months, or two and-a-half times as many as the volunteers would have produced had they continued transcribing at the same rate. Without having to invest in digitisation or programming, the AHRC grant could have employed both Research Associates for three years, allowing almost half of the remaining 40,000 UCL Bentham Papers to be transcribed. Instead, they spent the equivalent of a month's full-time labour moderating 1,009 submissions, with the rest of their time spent in development and testing of the interface, volunteer recruitment, publicity, maintenance, the conversion of legacy transcripts, and other editorial tasks. (Causer et al. 2012)

Similar concerns have been noted by Anderson (2011) and Zou (2011). It seems likely that in most projects, as with *Howitt & Fison* and *MPCE,* the benefit of crowdsourcing is not so much the prospect of faster transcription and free labour, but rather as a framework for establishing public engagement with the research and thereby identifying what in the material is of interest and significance to the world beyond the project team's own immediate circle. This ensures the project will develop in directions that will have maximum community impact. Already, transcribed pages are being read closely by Gunnaikurnai people for important cultural information. One of the Gunnaikurnai elders, Russell Mullet, has, for example, brought his breadth of knowledge to correcting transcriptions of Gunnaikurnai language material that were originally produced by the volunteers. In another case, a sketch map drawn into Howitt's papers was used to locate the site where the last *jeraeil* initiation ceremony was held on the McLellan Strait near the remote township of Seacombe in 1883.[3] As this ceremony was likely the last of its kind to be performed in Victoria, it holds considerable historical and cultural significance. The transcription of Howitt's notes and subsequent analysis by Gunnaikurnai elders will now be incorporated into an application to the Victorian Government to register the area a cultural heritage site.

The MPCE project community of 'volunteers' was always likely to be more specialised, and more remote. Unlike *Howitt & Fison*, we cannot rely on people local to the project to have a particular interest in the material. This means that even if a wider community of transcribers and annotators were to come on board, there would be additional complexities around managing participants at a distance. While some of the *Howitt & Fison* transcribers are remote, many are able to come to the museum, at least occasionally, and so it is easier to develop personal rapport with transcribers, to build a sense of community, discuss this often complex content with them, and provide training. Some transcribers do all of their volunteer work

---

[3] This is Howitt's spelling for this ceremony term. In a standardised modern orthography it becomes dyerrayil.

in situ at the museum, while others work at home. This personal interaction, although time-consuming, has been extremely important to maintaining enthusiasm for the project and the volunteers also welcome the opportunity to discuss the content of the manuscripts with experts in the field. Moreover, the face to face interaction provides an opportunity for the research team to provide feedback to the transcribers on how the transcripts are being used by First Peoples groups and the academic team. On the other hand, the development of, and engagement with, a remote community of volunteers for the MPCE project might have been beneficial to a project team that may otherwise feel geographically isolated from the subject of their research and from the most natural audience for it.

While academic researchers working on related projects are likely to have good reasons to support projects such as MPCE, harnessing this enthusiasm and converting it to volunteer labour is a different kind of challenge from attracting crowdsourcing volunteers from the broader community. Many crowdsourcers on other projects are retired, un(der)employed, or unable to work outside the home. They may have strong desires to take on intellectually stimulating tasks, to become part of a community, to learn new skills or information, or to engage with prestigious regional institutions such as museums or universities, and they often have plenty of free time in which to do so. Academics, on the other hand, already have associations with such institutions, have plenty of intellectually stimulating tasks on their to-do lists already, and have other channels for learning new skills or information and building community. Spare time in this group is a rare commodity. No matter how internally motivated a researcher might be to undertake data entry or annotation tasks to improve a resource that might ultimately benefit their research, the more tried and tested ways to motivate and engage them, such as meet-ups, email lists, training sessions at the museum or university, certificates or online acknowledgement, such as leaderboards and badges, may not be so successful. This is one of the reasons why MPCE decided that resources might be better invested in focusing on linked data and automating data structuring than on motivating human volunteers to input and structure new data.

One of the particular challenges for projects that rely on academic participants is the lack of recognition in academic institutions for 'non-traditional research outputs' such as datasets or digital curation and editing projects. As discussed in Hendery (2016), universities frequently do not know how to measure datasets and digital projects in the same way they count paper publications or citation metrics (although see Margetts et al. 2015 for some suggestions). Researchers counsel each other to prioritise publication of traditional papers rather than digital resources because of the expectations of universities, referees, and grant funding bodies. Moreover, researchers are more familiar with how to cite traditional publications and so the latter benefit from traditional measures of impact more easily. Finally, established methods of peer review lend papers and monographs a kind of authority for which there is as yet no clear substitute in non-traditional research outputs. All of these factors combine to ensure that academics' contribution to digital resource creation is not rewarded to the same extent that equal time spent on traditional publications would be.

While the participation of academics in crowdsourcing projects may not necessarily be incentivised or appropriately recognised and rewarded, academics who *create and manage* crowdsourcing projects may in fact find themselves gaining institutional recognition for this community engagement. Australian academic institutions have recently become very concerned with how to measure and incentivise 'impact' and 'engagement', which the Australian Research Council (2017) defines as follows:

> 'Impact':
>> *Research impact is the contribution that research makes to the economy, society, environment, and culture beyond the contribution to academic research.*
> 'Engagement':
>> *The interaction between researchers and research end-users outside of academia, for the mutually beneficial transfer of knowledge, technologies, methods, or resources.*
> 'End-user':
>> *An individual, community, or organisation external to academia that will directly use or directly benefit from the output, outcome or result of the research.*

The Australian Research Council conducted a pilot study in 2017 to measure research impact and engagement by asking universities to submit case studies. This pilot will continue in 2018. Research relating to Australian First Peoples is considered separately from other fields of research for the purpose of impact case studies. As yet, no explicit incentives are directly tied to these metrics, but now that impact and engagement are being measured, compared and recorded, and scholars are asked to construct narratives about their projects'

impact and engagement (for the pilot case studies, but also in national grant applications), this is likely to fuel universities to consider impact and engagement among other indicators of successful research when considering internal funding, promotion, and employment decisions.

Conclusions

The lessons learned so far from both the *Howitt & Fison* project and *Mapping Print; Charting Enlightenment* include corroboration of many findings from other crowdsourcing projects in the digital humanities, but also have some aspects specific to the Australian context. For example, the *Howitt & Fison* project has found, like many other crowdsourcing projects mentioned above, that most of the participation is from a small number of very active volunteers. Both projects have found that distributing the labour of transcription or data entry among a group of academics or the general public is not necessarily faster, cheaper, or more efficient than it would be for the project investigators to take on the tasks themselves. We have found, particularly with the *MPCE* project, that recent advances in open data and automation, including machine learning, provide new alternatives to human labour for some tasks that crowdsourcing has previously been used for. However, the creation of community, of a public interested in the work of the project, building of relationships, and distribution of knowledge through new networks are all reasons why such crowdsourcing experiments are nevertheless valuable.

The Australia-specific elements to these crowdsourcing projects discussed above included potential sensitivities in materials relating to Australian First Peoples, the complexity of the post-colonial context, and the geographical remoteness of the MPCE project from its main community of interest. Finally, while the creation of non-traditional research outputs such as datasets or digitisations is not yet fully considered equal to the more traditional modes of scholarly knowledge production, academics may find more incentives and rewards for engaging in crowdsourcing projects in Australia as institutional narratives increasingly turn towards measuring and rewarding impact and engagement.

Competing Interests

The authors have no competing interests to declare.

References

Anderson, K. 2011. "Even Crowdsourcing Can Get Too Expensive." *Scholarly Kitchen.* Accessed June 1, 2018. https://scholarlykitchen.sspnet.org/2011/03/14/even-crowdsourcing-can-get-too-expensive/. Archived at: https://perma.cc/HUB2-9ZCL.

Australian Research Council. 2017. *Engagement and Impact Assessment Pilot 2017 Report.* https://www.arc. gov.au/engagement-and-impact-assessment/ei-pilot-overview.

Batty, Philip, and Jason Gibson. 2014. "Reconstructing the Spencer and Gillen Collection Online: Museums, Indigenous Perspectives and the Production of Cultural Knowledge in the Digital Age." In: *Corpora Ethnographica Online Strategien Der Digitalisierung Kultureller Archive Und Ihrer Präsentation Im Internet*, Holger Meyer, Christoph Schmitt, Alf-Christian Shering, and Stephanie Janssen (eds.), 29–48. Munster: Waxman.

Causer, Tim, Justin Tonra, and Valerie Wallace. 2012. "Transcription Maximized; Expense Minimized? Crowdsourcing and Editing *The Collected Works of Jeremy Bentham.*" *Literary and Linguistic Computing*, 272: 119–137. DOI: https://doi.org/10.1093/llc/fqs004

De Boer, Victor, Michiel Hildebrand, Lora Aroyo, Pieter De Leenheer, Chris Dijkshoorn, Binyam Tesfa, and Guus Schreiber. 2012. "Nichesourcing: Harnessing the Power of Crowds of Experts." In: *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2012*, A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Aquin, A. Nikolov, N. Aussenac-Gilles, and N. Hernandez (eds.), 16–20. Berlin: Springer. DOI: https://doi.org/10.1007/978-3-642-33876-2_3

Dousset, Laurent, Rachel Hendery, Harold Koch, Patrick McConvell, and Claire Bowern. 2010. "Developing a Database for Australian Indigenous Kinship Terminology: The AustKin project." *Australian Aboriginal Studies*, 1: 42–56.

Fison L., and A. W. Howitt. 1880. *Kamilaroi and Kurnai.* Melbourne: George Robinson.

Gardner, Helen, and Patrick McConvell. 2015. *Southern Anthropology: a History of Fison and Howitt's Kamilaroi and Kurnai.* London: Palgrave-MacMillan.

Gibson, Jason. Forthcoming. "Dealing in the Sacred: Working with Restricted Collections." *Museum Anthropology.*

Glass, Aaron. 2014. "Indigenous Ontologies, Digital Futures: Plural Provenances and the Kwakwaka'wakw Collection in Berlin and Beyond." In: *Museum as Process: Translating Local and Global Knowledge*, Raymond Silverman (ed.), 19–44. London: Routledge.

Glass, Aaron. 2017. "Reassembling The Social Organization: Collaboration and Digital Media in Remaking Boas's 1897 Book." *Museum Worlds*, 5(1): 108–132. DOI: https://doi.org/10.3167/armw.2017.050111

Hedges, Mark, and Stuart Dunn. 2012. "Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research." *Arts and Humanities Research Council.* Accessed January 16, 2014. http://crowds.cerch.kcl.ac.uk/.

Hendery, Rachel. 2016. "'Writing about Music is like Dancing about Architecture': Integration of Multimedia into Linguistic and Anthropological Publications." In: *Language, Land and Song: Studies in honour of Luise Hercus*, Peter K. Austin, Harold Koch, and Jane Simpson (eds.). London: EL Publishing.

Howitt, A. W. 1887. "On Australian Medicine Men; or, Doctors and Wizards of Some Australian Tribes." *The Journal of the Anthropological Institute of Great Britain and Ireland*, 16: 23–59. DOI: https://doi.org/10.2307/2841737

Howitt, A. W. 1904. *The Native Tribes of South-East Australia.* London: Macmillan and Company, limited.

Kücklich, J. 2005. "Precarious Playbour: Modders in the Digital Games Industry." *Fibreculture*, 5. Accessed June 1, 2018. http://five.fibreculturejournal.org/fcj-025-precarious-playbour-modders-and-the-digital-games-industry/.

Kuehn, Kathleen, and Thomas F. Corrigan. 2013. "Hope Labor: The Role of Employment Prospects in Online Social Production." *The Political Economy of Communication*, [S.l.], 1(1): 1–17.

Margetts, Anna, Nick Thieberger, Stephen Morey, and Simon Musgrave. 2015. "Assessing Annotated Corpora As Research Output." *Australian Journal of linguistics*, 36(1): 1–21.

Needham, Rodney. 1967. "Introduction." In: *The Development of Marriage and Kinship*, Rodney Needham (ed.). Chicago, IL: University of Chicago Press.

Phillips, Carol. 2010. "'Crowdsourcing' vs. 'Community Sourcing.'" *Millennial Marketing*, September 10. Accessed June 1, 2018. http://millennialmarketing.com/2010/09/crowdsourcing-vs-community-sourcing/. Archived at: https://perma.cc/MMR8-7QAB.

Radcliffe Brown, A. 1931. *The Social Organisation of Australian Tribes.* Sydney: Oceania Monograph 1.

Ridge, Mia. 2012. "Frequently Asked Questions about Crowdsourcing in Cultural Heritage." *Open Objects (blog).* Accessed June 1, 2018. http://www.openobjects.org.uk/2012/06/frequently-asked-questions-about-crowdsourcing-in-cultural-heritage/. Archived at: https://perma.cc/9T6U-6XST.

Ridge, Mia. 2013a. "Digital Participation, Engagement, and Crowdsourcing in Museums." *London Museums Group (blog).* Accessed June 1, 2018. http://www.londonmuseumsgroup.org/2013/08/15/digital-participation-engagement-and-crowdsourcing-in-museums/. Archived at: https://perma.cc/8LJV-AP5A.

Ridge, Mia. 2013b. "From Tagging to Theorizing: Deepening Engagement with Cultural Heritage through Crowdsourcing." *Curator: The Museum Journal*, 564: 435–450. DOI: https://doi.org/10.1111/cura.12046

Stanner W. 1972. "Howitt, A.W." *Australian Dictionary of Biography*, 4. Melbourne: Melbourne University Press.

Stocking, George W. 1995. *After Tylor, British Social Anthropology 1888–1951.* Madison, WI: University of Wisconsin Press.

Terras, Melissa. 2016. "Crowdsourcing in the Digital Humanities." In: *A New Companion to Digital Humanities*, Susan Schreibman, Raymond Siemens, and John Unsworth (eds.), 420–439. Hoboken, NJ: Wiley-Blackwell.

Torrance, G. W. 1887. *Music of the Australian Aboriginals.* London: Harrison and Sons, Printers.

Zou, J. J. 2011. "Civil War Project Shows Pros and Cons of Crowdsourcing." *The Chronicle of Higher Education.* Accessed June 1, 2018. https://www.chronicle.com/blogs/wiredcampus/civil-war-project-shows-pros-and-cons-of-crowdsourcing/31749. Archived at: https://perma.cc/U6FH-P2C7?type=image.