

PRACTICE PAPER

Reviving an Old and Valuable Collection of Microscope Slides Through the Use of Citizen Science

John Pring¹, Lesley Wyborn² and Neal Evans¹

¹ Geoscience Australia, AU

² Australian National University, AU

Corresponding author: John Pring (john.pring@ga.gov.au)

Since the federation of Australia in 1901 Geoscience Australia, and its predecessor organisations, have amassed a significant collection of microscope slides of a variety of physical samples from across Australia, Antarctica, and adjacent regions. The extensive nature of the collection and the diverse and often remote nature of the source locations means that the cost of recreating the collection, if possible, would be \$AU100Ms. The original samples were collected as part of either extensive government geological mapping programs or more specific scientific expeditions conducted for major Government initiatives. They are technically open to anyone (industry, educational institutions, the public), but are essentially unknown and almost impossible to access.

Management of this collection was based on an aged card catalogue and ledger system developed in the pre-digital era. The aged management system, with increasing deterioration of the physical media, combined with loss of access to even some of the original contributors meant that rescue work was needed. Rescuing the collection made use of non-traditional means, including the extensive use of web-based citizen science and reference to a small number of onsite volunteers.

Through essentially a volunteer effort, from a group more used to biology related items, the project has seen the transcription of some 40,000 sample metadata records (more than 2.5 times our current electronic holdings). This paper examines the process undertaken and advocates the approach that has made it successful. It promotes the value and benefits to Geoscience Australia, participating volunteers and potential users of the collection.

Keywords: Citizen Science; Data Rescue; Geology; Thin Section; Microscope Slides; DigiVol; IGSN

1. Introduction

For more than a century Geoscience Australia, and its predecessors (the Australian Geological Survey Organisation (AGSO) and the Bureau of Mineral Resources, Geology and Geophysics (BMR)), have been collecting rock samples from around the world, with a particular focus on Australia and the surrounding region. Many samples in the collection are irreplaceable and come from locations that are now inaccessible and in some cases no longer exist. From these samples over 250,000 thin section microscope slides (also called petrographic sections or thin sections) have been produced.

Thin sections have been utilised in the geosciences since around 1850 (Folk 1965). Traditional thin sections are made using a diamond saw to cut a thin sliver of rock, which is mounted on a glass slide and ground down until the sample is only 30 μm thick, and can then be examined by a variety of techniques including plain polarized petrographic microscopes, reflection microscopes, electron microscopes and electron microprobes. The finished microscope slide is a useful aid in determining the mineralogy of the parent rock sample. Thin sections are one of the key ways of determining not just the mineralogical composition of a rock but the relationship between minerals within the rock. They are an aid in defining the history of a rock with respect to how the rock initially formed and subsequent events such as deformation and metamorphism. Minerals also control the chemical and physical properties of rocks (density, reflectance, magnetisation, etc.) and hence knowing the mineralogy of a rock can provide valuable insights into the

interpretation of other geophysical and geochemical datasets. The collection also includes microscope slides of microfossils, mud smears and nannofossils.

Today, by far the greatest volume of geoscientific data and information is more likely to be derived from high volume remotely sensed data sets (Wyborn & Lehnert 2016) (e.g., airborne data, satellite data, drones, etc.). Such collections tend to measure proxies of the real world (e.g., satellites can measure infrared radiation but the data needs to be mathematically manipulated to give temperature; airborne radiometric surveys measure gamma-rays produced by the radioactive decay of potassium, uranium and thorium are used to estimate concentrations of those elements). Hence, position-located, real world physical samples have a valuable role to play in the calibration of modern remotely sensed data sets, particularly those that can provide information on the minerals present at each site and their formation. It would be prohibitively expensive to collect the required samples from scratch. Increasingly, researchers are turning to historic collections of physical samples and thin sections in support of the calibration and enhancement of remotely sensed data sets. Indeed, one of the drivers for establishing this project to 'rescue' the historic Geoscience Australia collection was to demonstrate the potential usefulness of this collection to modern initiatives including the Australian Federal Government's current Exploring for the Future minerals, groundwater and energy programme to increase the attractiveness of Australia's North to investment (Geoscience Australia 2018).

However, the value and potential of this collection was not fully realised and the management system for this collection, started in 1901 in parallel with the collection of the samples, remained largely hardcopy based (Figure 1), with no online presence. Proposals to digitise the metadata records that show the distribution of the samples and (where available) sample descriptions have suffered from a lack of prioritisation, focus and available funding. While the collection is technically publicly available, the unstructured state of the hard copy management system has acted as an inhibitor to the wider discoverability, accessibility and reusability of this public asset. For example, it was impossible to use valuable mineral identifications to aid in the interpretation of very high cost geophysical and geochemical data sets. The opening up of this collection to electronic discovery is in line with Geoscience Australia's desire to maximize data potential and to improve access to collections. The opening up of the data through the project has broadened the Geoscience Australia stakeholder base and should enhance the relationship.

This paper describes the 'rescue' of a subset of 40,000 items from this collection, trialing

- 1) a method that made extensive use of a team of non-specialist citizen scientists to transcribe letter for letter, number for number, a series of handwritten cards into digital form;
- 2) incorporation of these data into a preexisting data system; and
- 3) the ways in which these data are now accessible on line and other potential uses.

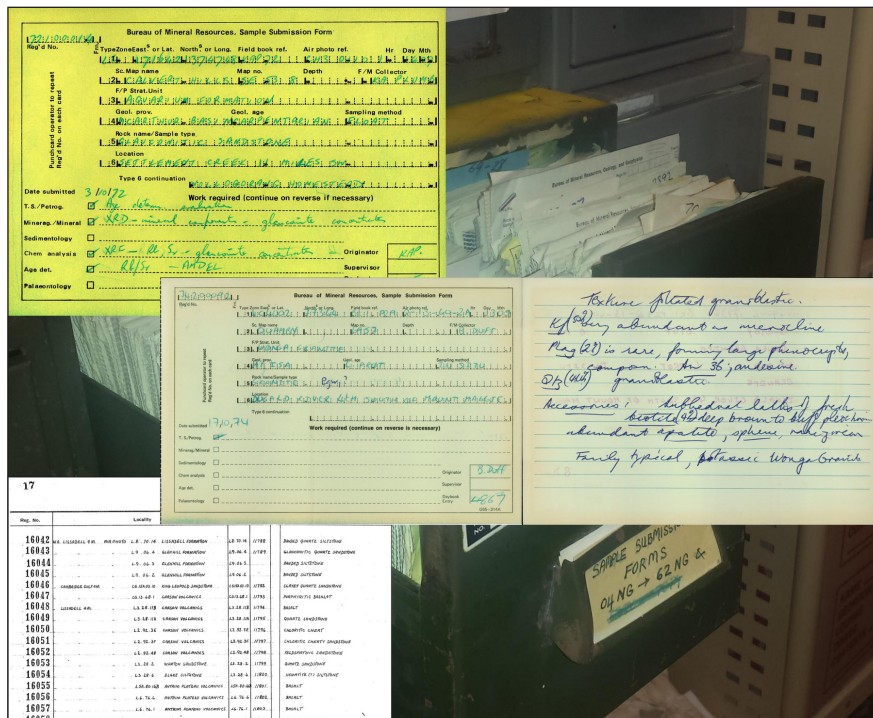


Figure 1: Illustration of current management 'system'.

The paper also describes future plans. Throughout this paper we will use the term “thin section slides” to describe microscope slides that are prepared from individual rock samples.

2. The Initial State

Most of the 250,000+ thin section slides are recorded only via hard copy records (**Figure 1**). These records stretch back into the first quarter of the 20th century and include important historic items such as microfossil slides made from samples taken by Sir Douglas Mawson during his 1911–14 expedition to the Antarctic (Australian Antarctic Division 2017). In undertaking the project it was decided to concentrate on the rock thin sections, i.e., thin sections made from rock samples.

Given the size and breadth of the collection, to reproduce the thin section slide collection today, would cost 100s of millions of dollars. Some localities no longer exist (e.g., sites that have been mined out), whilst for some others, land access issues inhibit the potential to recollect samples.

The primary challenge for the project was to convert thousands of handwritten paper entries, from multiple generations of authors, to meaningful machine-readable, structured metadata/data suitable for online consumption in the modern digital age. The labour-intensive nature of the traditional manual transcription process meant that staffing costs would be prohibitive for a standard approach, and hence if the cards were to be transcribed new alternative approaches to the project would need to be determined.

Due to the different media that the relevant data had been collected on over decades, the variability and quality of the handwritten source material (**Figure 1**), and the inconsistent and variable length of content originally provided (**Figures 3–6**), automated digitisation was not viable. It was clear that comprehensive human validation of the digitised material would be essential and that the project would support the use of a crowd source solution. Some initial hesitancy existed about the transcription being undertaken by non-experts. However, it was agreed that if the transcription was done in a letter for letter, numeral for numeral manner then non-expert resources could be crowdsourced for the initial transcription phase with a small number of subject matter experts available if/when required.

In looking for an approach DigiVol was identified as a possible crowdsourced alternative to traditional transcription. DigiVol (ALA 2019a) was developed by the Australian Museum in collaboration with the Atlas of Living Australia as part of the Australian Federal Government-funded National Collaborative Research Infrastructure (NCRIS) (Australian Government – Department of Education and Training 2018) initiative. While initially developed for transcribing records of non-geological content, it is fundamentally a crowdsourcing platform that facilitates the digitising of data including the transcription of historic scientific records. DigiVol has multiple capabilities, including visual analysis of data from camera traps and the transcription of hard copy material such as old journals and metadata cards. The resulting data are stored on the DigiVol platform with suitable provenance metadata and are readily retrievable.

3. The Rescue Process

This section describes the processes employed to identify the relevant records, establishing and using the DigiVol platform, and the methods of capturing and retaining the interest of volunteers. It also outlines a method of accessing the resulting metadata records and where this might lead in the future.

3.1. Identifying thin section slides within the area of interest

Given the scale of the collection it was decided to focus the initial rescue effort round a geographic region of current interest to Geoscience Australia to test the methodology of using citizen science in translating technical scientific data from handwritten cards into a digital database.

The area of interest for the project was loosely defined as Mt Isa (Queensland) to Tennant Creek (Northern Territory) and north (See **Figure 2**). In total it amounted to an area of approximately 400,000 km² with the bulk of the samples in these areas collected during mapping projects conducted during the 1960's and 70's that were based on national map sheets (1:250,000 equivalent).

3.2. Metadata

Prior to having a thin section made during these projects, a sample submission card had to be completed by hand (**Figure 1**) that provided minimal metadata including sample number, grid reference, map sheet, location description, and some basic information on the nature of the sample (e.g., rock type, texture, colour, mineralogy). Once the thin section was available to the scientist more detailed information was often filled in including mineralogy on the reverse side of the card (**Figure 1**). Unfortunately most of this additional information, which contained very valuable scientific detail, was handwritten, and at times, this handwriting was not very easy to decipher. It was definitely not of sufficient quality to enable scanning and automatically translate.

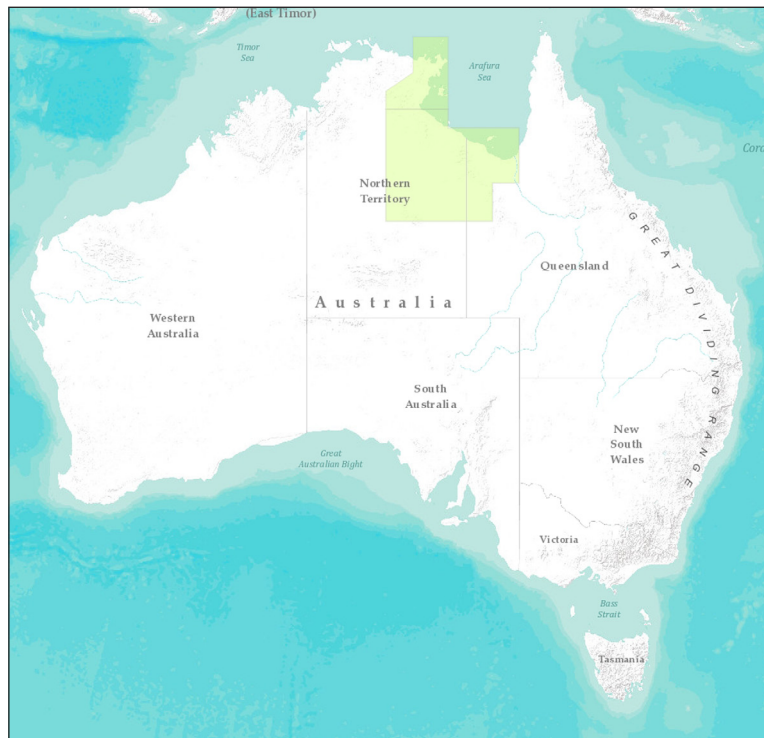


Figure 2: Map showing the 'Area of Interest' where the initial rescue operation was targeted.

3.2.1. Uniquely identifying each sample

Critical to the project was the assumption that each sample in the collection chosen was uniquely identified. Fortunately, each sample in the area was uniquely identified due to a field site numbering system started in Geoscience Australia during the 1960s (and still used today). An 8- to 10-digit code, e.g., 67846023, was assigned where:

- 67 = the year that the field site was mapped or the sample was collected.
- 84 = the two-digit identifier number of a mapping/sampling project, which could be related to the geographical region the sample was collected from. The region typically covered multiple map sheets.
- 6 = an identifier of the geologist that collected the sample/field site. Each field worker was usually assigned their own number by the project.
- 023B = a 3 digit, usually sequential sample number; these often had additional letters and/or number suffices to create a unique identifier for the sample or field site, e.g., 023A, 023B.

This 8-digit code was later increased to 10 by prefixing 2 digits to enable a 4 digit year to be included.

The systematic use of this numbering system throughout the area of interest proved to be very fortunate for the rescue effort, as for the most part, it did create a unique sample numbering system (although not always fail safe and some duplicate sample numbers did occur). Whilst the project identifier number only provided the general location of the sample site to be known. (e.g., Project 20 'Cloncurry, Qld', 'Qld. Mt Isa', or Project 10 'McArthur Basin, NT [Northern Territory]', or 'McArthur Basin, Mount Isa Inlier'), there was sufficient information encoded in this sample numbering system to then make discovery of additional metadata and/or identification of the original collector of the sample much easier.

3.3. Imaging sample submission cards and registers

The sample submission cards (**Figures 3–5**) were imaged by a commercial provider. Where data was also present on the reverse side (**Figure 4**), or within attached pages (**Figure 5**), these were concatenated to form a single image suitable for use within the DigiVol platform (**Table 1**). The generated images have also been used to fulfil archive requirements under the Australian Archives Act (Australian Government 2011) as described in the National Archives of Australia Scanning Specifications (as amended 22 Aug 2013) (National Archives of Australia 2013).

3.3.1. Examples

Figure 3: Sample Submission Card (Simple).

Figure 4: Sample Submission Card (Long #1).

Figure 5: Sample Submission Card (Long #2).

Table 1: DigiVol Image Requirements.

Criteria	Requirement (DigiVol) (ALA 2017a)
Format	JPEG
File Size	1–2 MB (2 MB max)
Number of images per task	One

3.4. Establishing a DigiVol 'Expedition'

Each activity or project within the DigiVol platform is referred to as an 'Expedition'. An Expedition comprises a quantity of 'like activities' all based around a common data capture template. The slide collection expeditions were based on transcriptions but other format expeditions exist within DigiVol. During the course of the transcription process six sample submission card-based expeditions and two expeditions based on bound registers (**Figure 6**) were conducted. The table below (**Table 2**) indicates the size of each expedition that was undertaken.

Currently (6 January 2019) DigiVol has 3,758 volunteers working on 13 'Expeditions' (Current DigiVol Landing page), but since launching DigiVol in 2011 1,246,277 tasks (DigiVol term for a record) have been completed.

3.5. Running an expedition

DigiVol allows the Administrator of an expedition to define the template by:

1. Selecting an underlying structure from a number of predefined formats (e.g., number of columns, spreadsheet format, camera trap/questionnaire, journal etc.);

Reg. No.	Locality	Age	Description	Sender	Date
2635 R.225	Agassiz's Mine, Tasmania	? Devonian	Granodiorite	W. G. Woodrough	25-11-'28
2632 R.226	Wentworth, Rosbery, Tasmania	Lower Devonian	Schist	"	16-11-'28
2633 R.227	Stoney Gully, Tasmania	banked Devonian	Lignite & Limestone	"	22-11-'28
2634 R.228	Little Cypriped Pass, D'Entrecasteaux Ch. Is.	Devonian	Including Diabase, Mafic Porphyry	"	9-11-'28
2635 R.229	South of Buckland, Tasmania	Mesozoic	Coarse Grained Gneiss	"	27-11-'28
2632 R.230	North crest hill off from Spindle Dr	Lower Devonian	Micro Schist (Mafic Porphyry)	"	21-11-'28
2632 R.231	Point Bluff, Wynyard, Tasmania	Mesozoic	Mafic Schist	"	"
2635 R.232	Rushlow, Tasmania	Devonian	Limestone	"	11-'28
2635 R.233	Agassiz's Mine, Tasmania	Devonian	Schist	"	25-11-'28
2635 R.234	South of St. Helens	? Devonian	Traps - Diabase	"	"
2632 R.235	Mt. Lyell Summit	banked Devonian	Schist	"	11-11-'28
2632 R.236	Patrick River at Mt. 1	banked Devonian	Quartzite	"	18-11-'28
2632 R.237	South of Huonua, Tasmania	Mesozoic	Gneiss	"	26-11-'28
2635 R.238	Rungroona	Devonian	Veinlet Basalt	"	28-11-'28
2632 R.239	Wentworth Mine, Rosbery, Tasmania	Lower Devonian	Schist	"	16-11-'28
2635 R.240			Fine Grained Iron, Lead Schist	"	"
2632 R.241		? Devonian	Porphyry	"	"
2635 R.242	Brisis, Tasmania	Devonian	Basalt	"	24-11-'28
2635 R.243	Bungarooma	Devonian	Basalt	"	28-11-'28
2637 R.244	Harford - Frankford, Tasmania	Mesozoic	Diabase (Dolomite)	"	22-11-'28
2632 R.245	3 miles from Brownfield, Tasmania	banked Devonian	Quartzite	"	"
2635 R.246	Huonua Mafic on Rungroona	? Devonian	Gneiss	"	28-11-'28
2632 R.247	near Circular Head, Tasmania	banked Devonian	Slaty Rock	"	19-11-'28
2632 R.248	South of Mount Cameron	? Devonian	Gneiss	"	24-11-'28
2635 R.249	Muscovite, Tasmania	Pre-Cambrian	Schist	"	21-11-'28
2632 R.250	La Porte District	Devonian	Traps	"	"
2632 R.251	Brisis, Tasmania	Devonian	Chert Basalt	"	24-11-'28
2635 R.252	Cristone Park, S. of Seamaner, Tas.	? Devonian	Handmade Granite	"	20-11-'28
2632 R.253	Head of N.W. Bay, Hobart, Tasmania	Devonian	Basalt	"	8-11-'28
2635 R.254	Croft, Tasmania	Mesozoic	Schist	"	27-11-'28
2638 R.255	Port Cresset, Tasmania	Devonian	Mafic Porphyry	"	9-11-'28

Figure 6: Rock (Sample) Register (example page).

Table 2: Expedition Sizes.

Item Transcribed (Register/Card)	Number of pages (Registers)	Number of Sample Submission Cards
Rock Register 2	471	
Rock Register 1	505	
Sample Cards #1		901
Sample Cards #2		959
Sample Cards #3 (Long format/multi page)		390
Sample Cards #4		990
Sample Cards #5		1409
Sample Cards #6		1208

2. Adding fields of default size (labelled to suit the specific expedition);
3. Amending the priority of each field to balance the layout and structure that the volunteer will see. Fields cannot be moved within the template other than by changing the priority;
4. Loading of pick lists to be associated with fields to assist in ensuring more consistent results.

Once the template was defined, images of the sample cards and register pages to be transcribed were uploaded as .JPG files in batches on to the DigiVol platform. Some larger files were reduced in resolution to meet size restrictions. Each file was automatically allocated a sequence number to allow later sorting.

3.6. Transcription process

The actual transcription approach within DigiVol is a two-pass process: firstly the initial transcription, followed by a separate validation process.

3.6.1. Initial transcription

The first pass allows all registered DigiVol users to select and transcribe a document. The user is presented with an image for transcription and fields in which to enter the data. (Figure 7). The user is provided with instructions on what is required via 'Tutorials'. If needed the user is able to seek specific or general advice on a topic via the expedition forum. Questions and comments entered by users can be addressed by either the expedition administrator or other users.

A transcriber is also able to make notes on a task to indicate areas of issue that could not be resolved (e.g., difficult handwriting). These notes are available to the subsequent validator and the expedition administrator and remain with the transcribed information.

3.6.2. Second pass

Once a task (DigiVol term for a record) has been transcribed it is then available for the second pass, 'Validation'. This allows a selected group of users for that expedition to review the transcribed documents and to:

- validate the data;
- modify; or,
- reject the transcription.

In a similar manner to the initial transcriber, the validator is also able to provide notes on a task to provide feedback to the expedition owner.

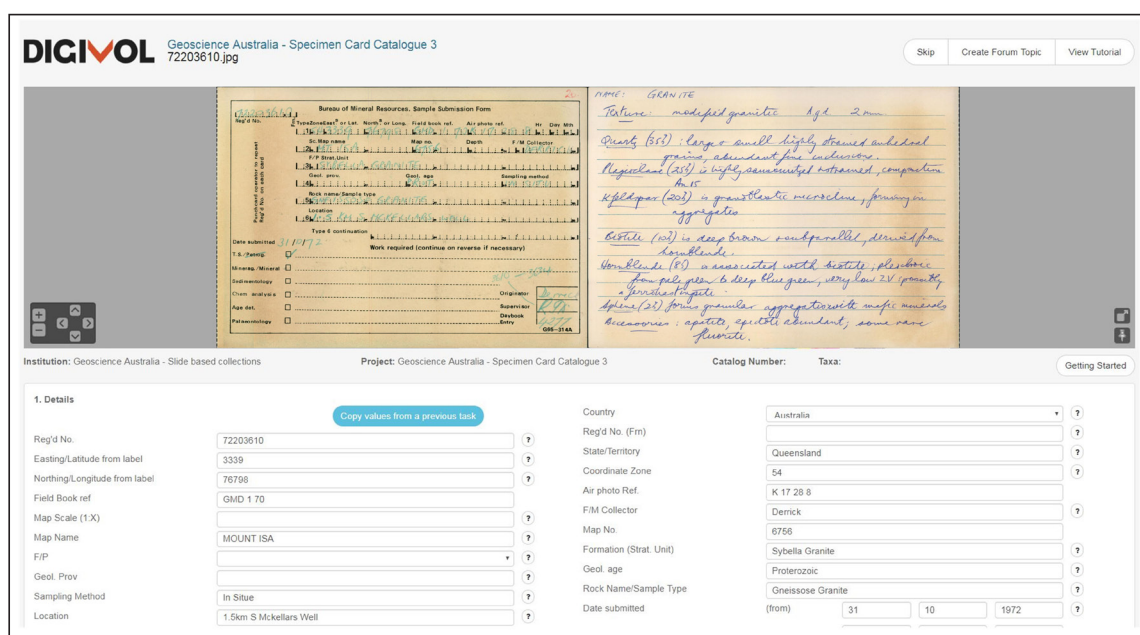


Figure 7: Example DigiVol transcription User Interface.

3.7. Social aspects – capturing and keeping interest

Working with volunteers requires some adjustment of approach from paid employees in a conventional workplace: capturing and keeping their interest is critical to the success of an 'expedition'. By and large the volunteers are participating in a DigiVol expedition out of interest and while they remain interested they are likely to keep contributing. The DigiVol platform comes with an existing user base. The current user base is 3,758 volunteers (ALA 2019b). Although this created a ready 'market' for selling the expeditions to, attention had to be paid to capturing and retaining their interest. For a community used to capturing information on biological specimens, rocks were quite different.

3.7.1. Capturing interested volunteers

For this project the initial capture of the volunteer's interest was achieved by presenting the Geoscience expeditions as interesting, both from a visual sense (using related images that are visually appealing (**Figure 8**) and from a story sense (giving a sense of the history and background as to why the collection is important).

3.7.2. Maintaining the interest of the volunteers

Given the uniqueness of the task, once volunteers were 'captured', maintaining the interest, and through that, the continued active support of volunteers, was achieved by:

1. **Platform-based support.** The DigiVol platform provides a forum environment that allows transcribers to ask questions of the administrator, expedition author and each other. This virtual community in addition to addressing questions also provides positive feedback when transcribers reach milestones (e.g., 100 expedition transcriptions). The mutual support offered by the forum partially ameliorates the impact of time zones on the ability of the expedition author to respond to issues raised;
2. **Timely response to questions.** We endeavoured to address questions raised via the expedition forum within 24 hours. Anecdotal feedback from volunteers indicates that they found that this allowed them to comfortably proceed more rapidly and they also indicated that the responsiveness demonstrated the expedition administrators' respect for contributors. (ALA 2019c);



Figure 8: Interesting Cover Photograph to capture Volunteer interest.

3. **Additional related information.** Where a volunteer would ask for further information on a tangential topic we would develop a tutorial both to address the immediate question, and also try to provide a more complete background on the topic. These tutorials would then be stored and referenced within the DigiVol expedition and contributed to building up a Frequently Asked Questions resource. An example of this is enabling users to understand how 'Air Photos' were used in determining spatial information in the pre-GPS era. (ALA 2016b);
4. **Additional items of interest.** Periodically, to help encourage continued support of the expeditions we would release a 'tutorial' that raised the profile of the expeditions. These tutorials did not directly relate to the transcription but drew upon the work of the volunteers to provide examples of transcribed minerals (ALA 2016c), background on the making of thin sections (ALA 2016a), and even the meaning of words through the use of well-known board games (ALA 2017b); and
5. **Acknowledgement of contribution milestones.** On the basis that we all seek acknowledgement for our contributions we would note, on the expedition forums, when volunteers had achieved significant milestones (e.g., multiple of 100 transcribed tasks for the expedition) (ALA 2017c). This is in addition to the virtual rewards system built into the DigiVol platform where volunteers are rewarded with virtual tokens for achieving milestones. Other volunteers would often respond to these posts with congratulations of their own.

3.8. Download, review and prepare

Once all the cards within an expedition had been firstly transcribed and subsequently validated the information was downloaded from the DigiVol platform as a CSV file. This file was then reviewed for consistency (e.g., 'Queensland', 'QUEENSLAND' or 'QLD') by Geoscience Australia employees. Any inconsistencies are made more evident given the column and row nature of the data presentation.

In order for the resource information to be used further, the actual location of the sample (and its uncertainty) needs to be known. Due to the variations in the manner in which spatial data was recorded for samples over the decades, volunteers were asked to simply transcribe exactly what was written letter by letter, number by number. The spatial information provided for each sample varied in:

1. Level of detail. Some geologists recorded positions with all the accuracy possible, whilst others simply noted the map sheet with a general reference to being near or in direction X from town/homestead Y.
2. Manner of recording. (Figure 10) When a position was given it might have been given as grid coordinates, Latitude/Longitude, points on a particular air photo or a description of the location. Due to the considerable effort involved in determining locations based on air photos it was decided that these samples would use the centroid of the map sheet provided as the sampling location. To preserve the integrity of the digital database, the accuracy of the location of the observation was also recorded.

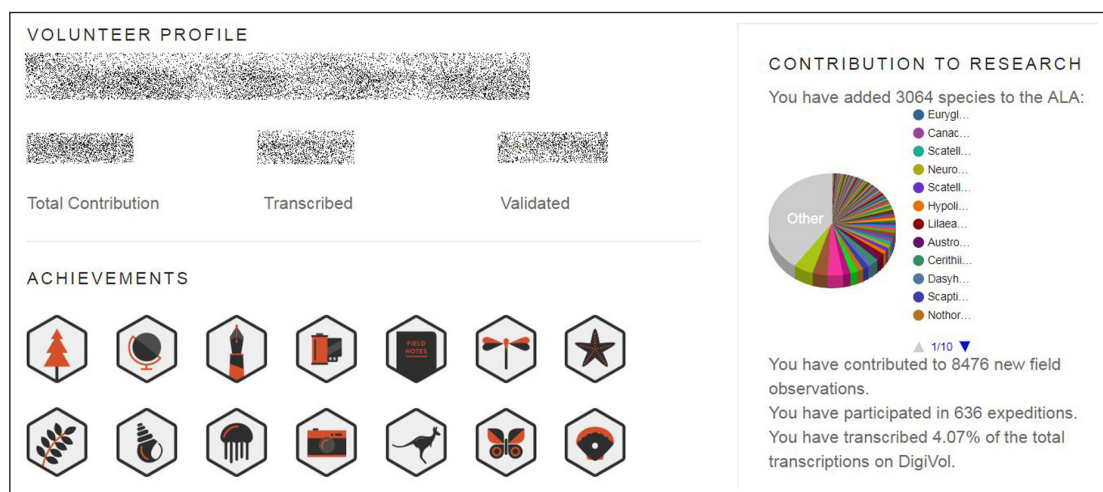


Figure 9: Example DigiVol User Profile.

1 : 250,000 } sheet <i>Cloncurry</i> 1 : 50,000 }	Fieldbook ref		
	Zone	Easting or Latitude	Northing or Longitude
Sample Co-ordinates	6	3 2 9 5	2 4 0 2 0
Grid Reference (Map Sheet Cloncurry)			
<i>3 1/2 miles W of Chumvale</i>			
Location Description			
Rock name/ Sample type	Airphoto ref <i>523/3-13019</i>		
1 : 250,000 } sheet <i>GASON SG 54-13</i> 1 : 50,000 }	Fieldbook ref <i>AMG 30</i>		
Air Photo Reference (Map sheet Gason)			

Figure 10: Example Spatial References (Sample Submission Card).

3. Projection used. Even when a point location is given via coordinates (e.g., Lat/Long or grid reference), due to the time over which the samples have been collected the various mapping projections and datums have changed. Determining the correct projection involved additional investigation during preparation of the data for upload. Some of the earliest submissions predate any form of modern spatial projections (e.g., AGD66, AGD84, etc.) but were defined using the Clarke 1858 spheroid latitudes and longitude.

3.9. Access and discovering the newly digitised data

Once the data from the cards had been transcribed, validated and downloaded from the DigiVol platform, it was then uploaded into the relevant Geoscience Australia databases. In addition, each parent sample, as well as the derivative thin section slide was assigned an International Geo Sample Number (IGSN), a global identifier system designed to provide an unambiguous globally unique persistent identifier for physical samples and facilitate the location, identification, the citation of physical samples and the ability to link any sample to other data or any publications derived from that sample. By assigning a unique IGSN identifier to both the 'parent' rock, as well to the 'child' thin section slide, means that other derivative sample preparations (e.g., mineral separates, rock powders, etc.) and derived data sets (e.g., geochemistry, physical rock properties, etc.) that these also have unique persistent identifiers assigned to them (Devaraju et al 2017).

The use of the existing internal Geoscience databases and the minting of IGSNs on each thin section slide and its parent sample, means that a wealth of tools already developed for these can now be applied to the rescued samples. For example, from the Geoscience Australia Databases, data and metadata for the 40,000 thin sections rescued in this project, were integrated seamlessly with the existing records of thin sections. Progressively this combined resource is now being made available via an Open Geospatial Consortium (OGC) web services that enables users to search spatially for thin section slides within their area of interest (Geoscience Australia 2019). Once the desired samples have been located by the user, via the use of the webservice, the samples can be requested for viewing or borrowing.

Since commencing the work Geoscience Australia has received external requests for thin section slides to support PhD research and internal investigations. Provided the thin section slides are contained within the 67,000+ catalogued in Geoscience Australia, initial retrieval of the thin section slide only takes 10–15 minutes. This is a significant efficiency gain for internal collection managers and results in faster delivery to clients (Collection managers have suggested that retrieval could take some days).

The use of OGC web services also allows access to the thin section metadata via any tool capable of consuming OGC data services. Hence the metadata can be utilised by other external tools and data systems and the user is not dependent on using only those tools provided by Geoscience Australia on its website.

Industry interest in the slide collection to date has been limited. This is believed to be in large part due to the lack of knowledge and inaccessibility of the slide collection. With the move of the handwritten metadata to an electronic format followed by the minting of an IGSN for each thin section slide, as well as its parent sample which in turn, will enable the linking to other derivative data such as geochemistry and petrophysics, the value of the new collection to industry will only increase.

Figure 11 indicates the Australian-based Thin Sections that can now be accessed.

3.10. Future developments

With the provision of web services (Thin Sections Collection 2019), users will be able to access the sample metadata and, where available, descriptions of that sample. To access the actual thin section slides, they need to be physically shipped to the requester or viewed at Geoscience Australia. This creates a number of risks and issues, relating to delivery timeframes and the risk to the thin section slides themselves, including breakage and loss. Many of the older thin section slides are extremely fragile, with some of the material used in their manufacture deteriorating over time (e.g., resin used for attaching labels obtained from the Balsam Fir tree ('Canada balsam')). For this reason, some slides are unavailable for shipping. An alternative is to 'deliver' quality digital images of the thin section slides for screening and potential analysis purposes to help reduce the numbers that need to be actually shipped. One example of this approach is the British Geological Survey's Britrocks web application (**Figure 12**). Britrocks allows users to find thin sections and examine them visually both in plane and cross polarised light (BGS 2019).

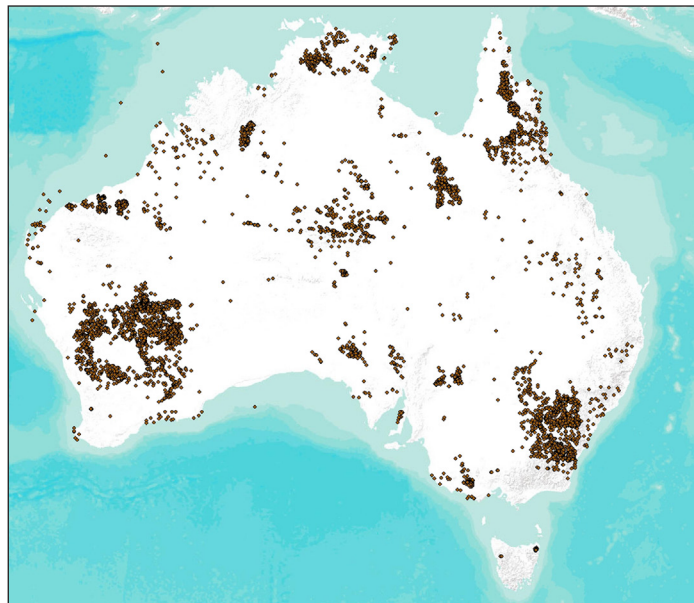


Figure 11: Map of Thin Sections within Australia.

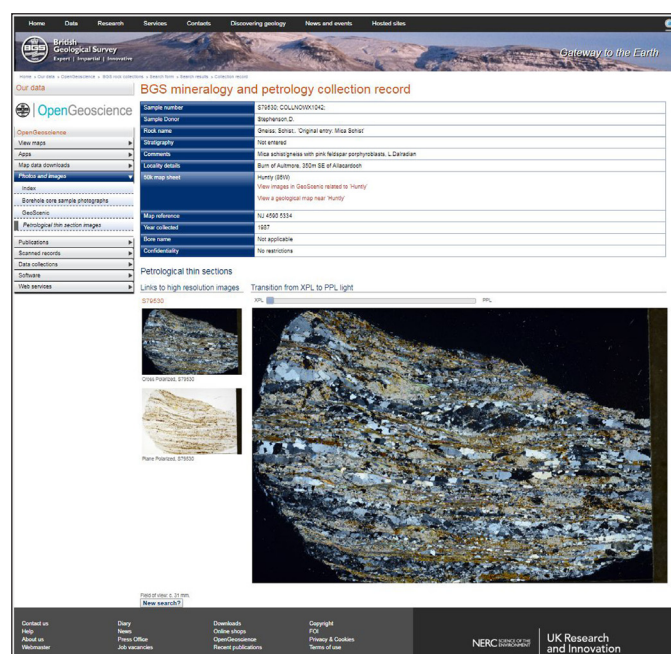


Figure 12: British Geological Survey Online discovery tool with thin section images. (sample S79350).

It is worth noting that in the Scottish component of the British Geological Survey project, some 100,000 thin sections were photographed, in both plane and cross polarised light, by volunteers over a period of approximately 15 months (Simmons 2012). Geoscience Australia could follow a similar path and progressively provide images of their thin section slides.

4. Discussion

This thin section data rescue project was an undertaking that would not have been possible from a financially or resource availability perspective without the active participation of volunteers who were supported by access to current subject matter experts from within Geoscience Australia.

The volunteers consisted of two main groups:

1. Online: These volunteers (49 (ALA 2018)) were mostly Australia based and undertook both the initial transcription of records and the subsequent validation through the web based DigiVol platform.
2. Onsite: This small number of volunteers (3) provided detailed subject matter expertise in the areas of geology, contemporary processes and cartography. These were all past employees of Geoscience Australia and its predecessor and had participated in the collection of some of the samples being transcribed.

4.1. Working with the DigiVol volunteers

4.1.1. What sort of person is volunteering for DigiVol and the Rock Expedition?

Figure 13, shows examples of the volunteers that worked on the Geoscience Australia expeditions.

Lang, in a report for the Australian Museum 'DigiVol online volunteer evaluation report' (Lang 2015), noted as an initial observation that DigiVol volunteers tend to be mature (54%- retirees represent the single largest group), female (63%), working from their home computer (64%) and most likely with a higher education qualification (Figures 14–16).



Figure 13: Example DigiVol volunteers working on our rock expeditions.

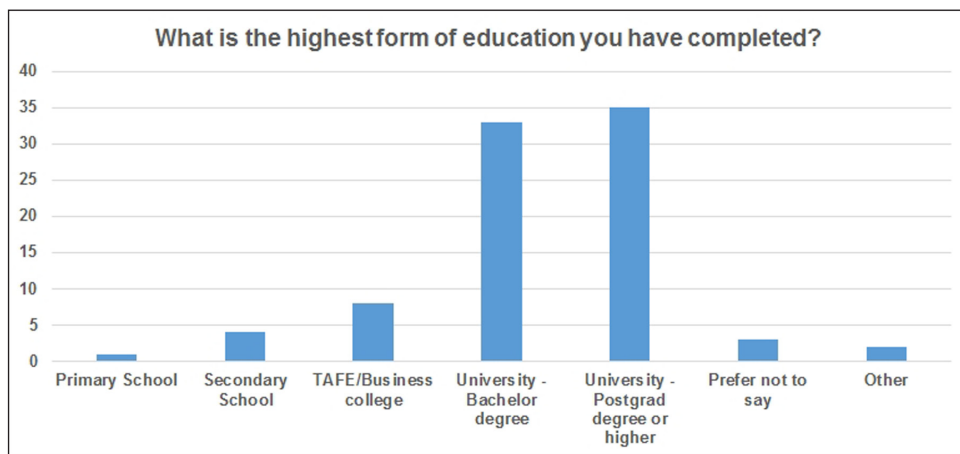


Figure 14: Highest level of Education for DigiVol Volunteers (Lang, 2015).

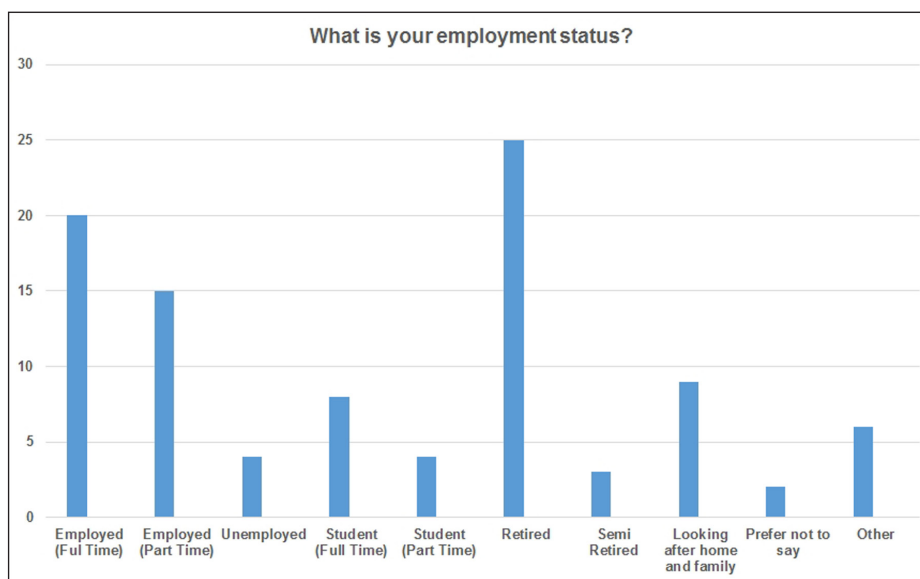


Figure 15: Employment status – DigiVol Volunteers (Lang, 2015).

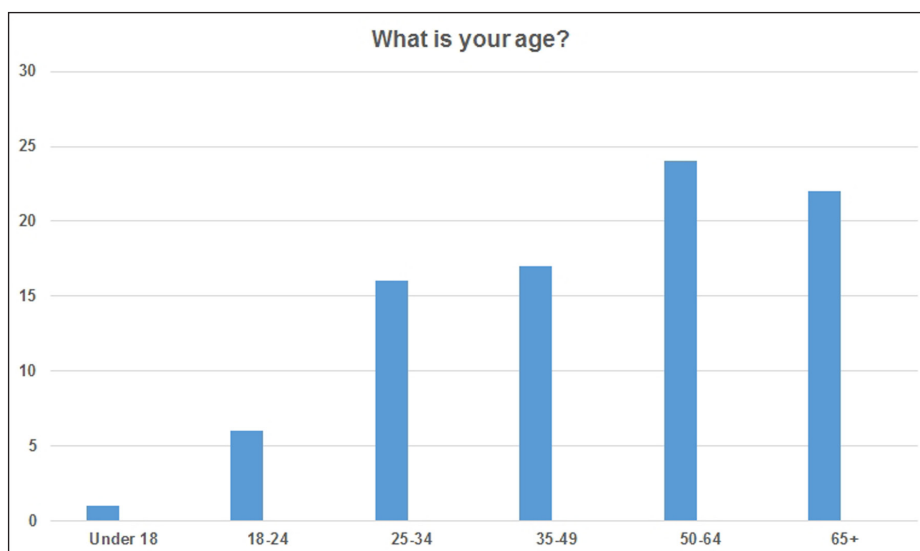


Figure 16: Age of DigiVol volunteers (Lang, 2015).

The evidence from the Thin Section Slides Project supports Lang’s findings with the top 2 transcribers (both female retirees) transcribing almost 44% of the total tasks (**Figure 19**).

4.1.2. Dedication and drivers

Table 3 indicates the estimated time taken to transcribe each style of sample document.

While acknowledging that these numbers are only indicative it does suggest that the volunteers expended significant effort in transcribing the data for no tangible return to themselves.

4.1.3. What drives people to expend considerable amounts of their own time?

Discussions with some of the volunteers involved with the Slide Based collections expeditions indicated a variety of reasons for participating these include:

1. They like doing the work and seeing the outputs of their efforts while gaining more knowledge.
2. Keeping mentally active is a key concern.
3. Sense of Satisfaction. Volunteers indicated that participating in a DigiVol expedition provided them with a sense of ‘job’ satisfaction that was also contributing to society.

4. 'Discovering' new places. Volunteers indicated that their transcriptions had informed their subsequent travels within Australia. This, they indicated, added a level of 'reality' to the sample information that they were transcribing on their return.
5. Friendly rivalry. With the DigiVol platform providing statistics and a ranking system volunteers often indicate a competitive attitude. One volunteer indicated that, as the register transcriptions were taking longer than normal tasks, they were doing additional transcriptions from other (shorter) expeditions to 'keep up' their statistics.
6. Personal statistics and virtual prizes. The platform provides individual users with information on the number of transcriptions and validations that they have done. The individual is also able to earn virtual achievement awards (**Figure 9**).
7. Formation of an online 'community'. Access to the platform's forum facilities encouraged the forming of an online community and a sense of belonging.

The below figures (**Figures 17–18**) indicated that most transcriptions occurred during what would be standard business hours for Australia (particularly in eastern Australia.)

Figure 19 indicates that approximately 64% of all the transcriptions done as part of Geoscience Australia Slide Based collections project were undertaken by 3 transcribers.

As mentioned, the project also made use of a much smaller group of onsite volunteers (3). These onsite volunteers were all past employees of Geoscience Australia, or the organisation's predecessors (AGSO and

Table 3: Level of Volunteer effort in transcribing tasks (Time to transcribe each type of record based on average of independent anecdotal advice from Users).

Type of card	Length of time to Transcribe	Number of items	Estimate of effort by volunteers
Rock Register page	45 min	976	658.8 hours
Sample Card (simple)	5 min	2850	614.9 hours
Sample Card (long)	20 min	390	130.0 hours
		Total	1143.7 hours

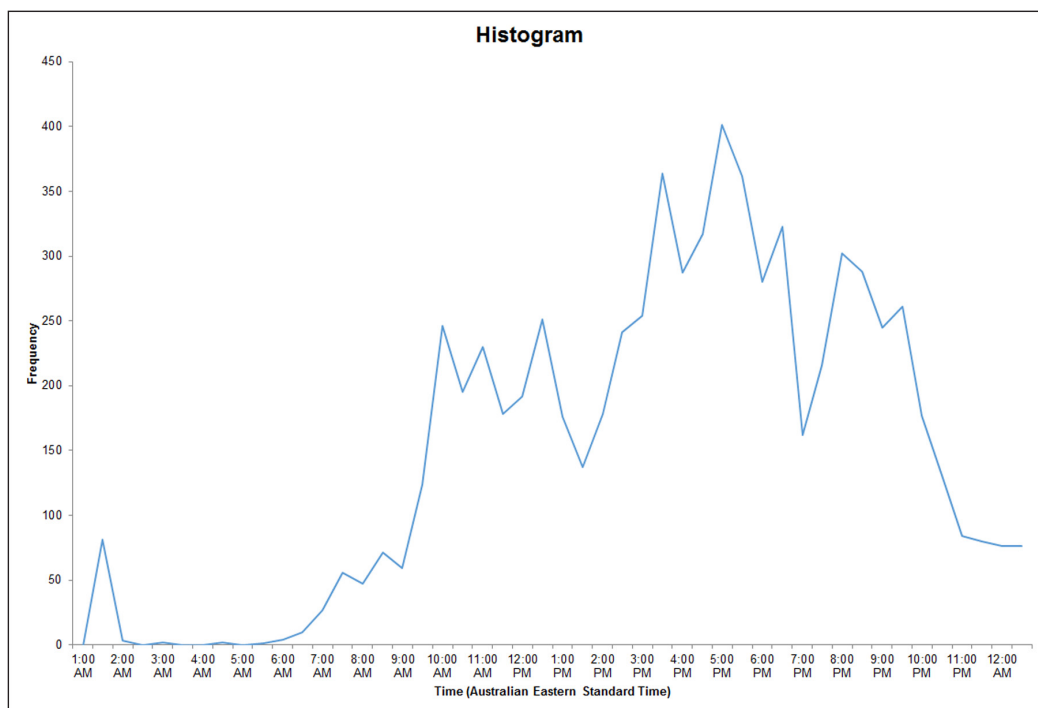


Figure 17: Graph – Total Number of Items Transcribed vs Time of Day (Based on transcription metadata generated from Geoscience Australia Slide Based collection expeditions (DigiVol).

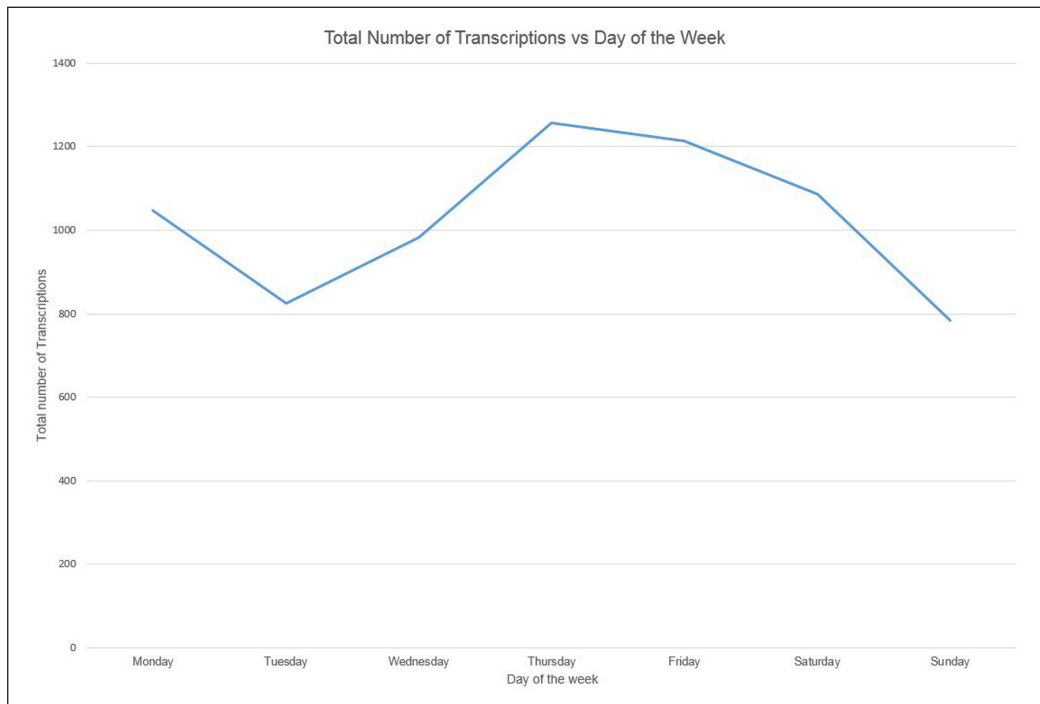


Figure 18: Graph – Total number of Transcriptions vs Day of the week.

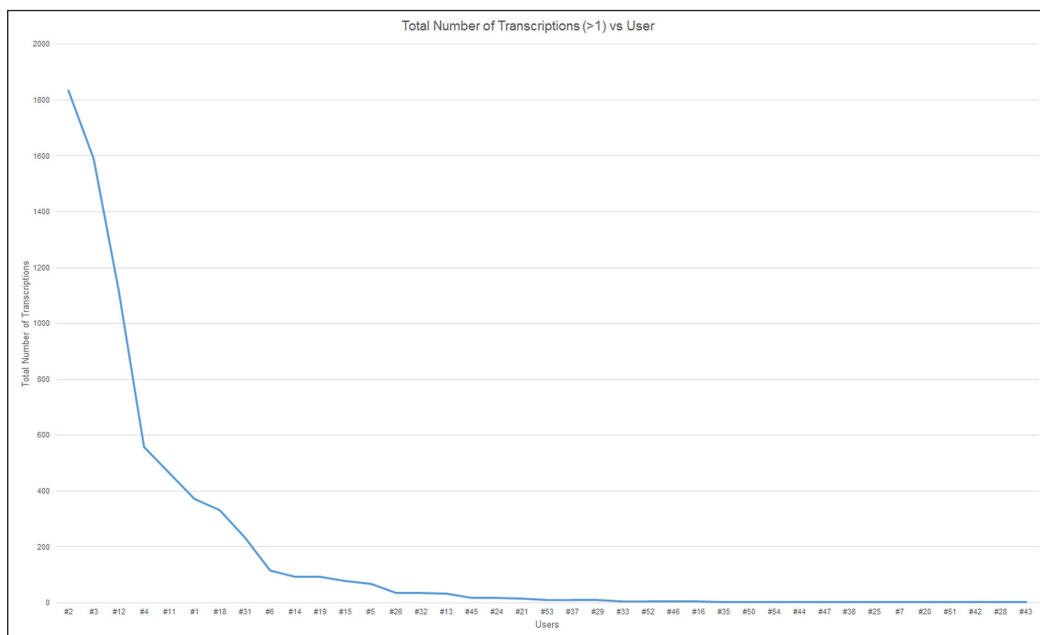


Figure 19: Graph – Total Number of Transcriptions (>1) vs User.

BMR), in the early 1960s and had participated in many of the field programs that collected the samples, and/or were involved in the curation of the samples and the various generations of that curation process. Access to this subject matter expertise and experience aided greatly in determining the practices and processes of the period from which the samples came. These individual volunteers did not participate full time but acted in a consultancy style capacity.

4.1.4. Relative 'costs'

DigiVol showed that valuable information could be captured by letter for letter, number by number transcription of aging pre-digital data formats stored on varying hard copy formats. The DigiVol volunteers came with an ability to decipher often appalling handwriting, and it could be suggested that with respect

to interpreting handwriting they were 'subject matter experts'. This transcription process showed that the transcribers did not need firsthand geological knowledge, although the provision of a lexicon was useful.

Subject matter experts were a scarce resource and the ability to focus their attention to specific issues made more effective use of this limited resource. They were also invaluable in the development and review of the DigiVol tutorials that in turn educated and helped many of the DigiVol volunteers.

The project would not have been possible without either group of volunteers.

5. Conclusions

The use of volunteers in the rescue of this valuable data resource has proved beneficial to Geoscience Australia in terms of the availability of the data and the ability to access physical samples that would otherwise have continued to languish. The volunteers have indicated that they also found the work beneficial in the form of the mental stimulation, the sense of achievement and the social interaction.

Access to the collection will potentially help industry and academic researchers to conduct virtual preliminary geological surveys of areas and refine their planned field surveys without the expense of field time.

The project opened up an old collection to modern access methods and had the added benefit of raising Geoscience Australia's profile, and geology more generally with a new segment of the community.

Acknowledgements

This paper is published with the permission of the CEO, Geoscience Australia. We are sincerely grateful to the numerous volunteers of DigiVol: without their efforts this collection would not be available online. We also thank Irina Bastrakova and David Champion whose reviews have considerably improved the manuscript.

Competing Interests

The authors declared no potential conflicts of interest exist with respect to the research, authorship, and/or publication of this article.

References

- Atlas of Living Australia, Australian Museum.** 2016a. *Slide Based Collections (Thin Sections)*, 28 November 2016. Available at https://volunteer.ala.org.au/data/volunteer/tutorials/Geoscience%20Australia_Slide%20Based%20Collections%20-%20Thin%20section%20background.pdf, [Last accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2016b. *Explanation of 'Air Photos'*, 25 November 2016. Available at https://volunteer.ala.org.au/data/volunteer/tutorials/Geoscience%20Australia_Air%20Photos%20-%20explanation.pdf [Last accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2016c. *Examples of what it is about*, 14 December 2016. Available at https://volunteer.ala.org.au/data/volunteer/tutorials/Geoscience%20Australia_Examples%20of%20what%20it%20is%20about.pdf [Last accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2017a. *DigiVol Administration Manual*, January 2017. Available at https://volunteer.ala.org.au//data/volunteer//tutorials/DigiVol%20Administration_Admin%20Manual%20Nov%202017.pdf [Last accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2017b. *New Scrabble Words*, 31 January 2017. Available at https://volunteer.ala.org.au/data/volunteer/tutorials/Geoscience%20Australia_New%20Scrabble%20Words.pdf [Last accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2017c. *Well done Ross on Reaching 400*, 20 June 2017. Available at <https://volunteer.ala.org.au/forum/viewForumTopic/18740687> [Last accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2018. *DigiVol – Geoscience Australia – Slide based collections*, 4 October 2018. Available at <https://volunteer.ala.org.au/institution/index/15762158> [Last accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2019a. *DigiVol*. Available at <https://volunteer.ala.org.au/> [accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2019b. *DigiVol – Statistics*, 26 February 2019. Available at <https://volunteer.ala.org.au/stats/index> [Last accessed 26 February 2019].
- Atlas of Living Australia, Australian Museum.** 2019c. *DigiVol forum*, 23 February 2019. Available at <https://volunteer.ala.org.au/forum/index> [Last accessed 26 February 2019].

- Australian Antarctic Division, Department of the Environment and Energy.** 2017. *Pioneers in Antarctica – Sir Douglas Mawson*, 6 March 2017. Available at <http://www.antarctica.gov.au/about-antarctica/history/people/douglas-mawson>, [Last accessed 26 February 2019].
- Australian Government.** 2011. *Federal Register of Legislation – Archives Act 1983*. Available at <https://www.legislation.gov.au/Details/C2012C00025> [Last accessed 26 February 2019].
- Australian Government – Department of Education and Training.** 2018. *National Collaborative Research Infrastructure (NCRIS)*, 25 June 2018. Available at <https://www.education.gov.au/national-collaborative-research-infrastructure-strategy-ncris>, [Last accessed 26 February 2019].
- British Geological Survey.** 2019. *BGS Rock collections*. Available at <http://www.bgs.ac.uk/data/britrocks/britrocks.cfc?method=searchBritrocks> [Last accessed 26 February 2019].
- Devaraju, A, Klump, J, Tey, V, Fraser, R, Cox, SJD and Wyborn, LAI.** 2017. A Digital Repository for Physical Samples: Concepts, Solutions and Management. In: Kamps, J, Tsakonas, G, Manolopoulos, Y, Iliadis, L and Karydis, I (eds.), *Research and Advanced Technology for Digital Libraries (TPDL 2017)* (1st ed., Vol. 10450, pp. 74–85). Cham, Switzerland: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-67008-9_7 [Last accessed 6 March 2019].
- Folk, RL.** 1965. Henry Clifton Sorby (1826–1908), the Founder Of Petrography. *Journal of Geological Education*, 13: 2, 43–47. DOI: <https://doi.org/10.5408/0022-1368-XIII.2.43> [Last accessed 6 March 2019].
- Geoscience Australia.** 2018. *Exploring for the Future minerals, energy, groundwater – Invest in resource exploration*. Available at www.ga.gov.au/eftr [Last accessed 26 February 2019].
- Geoscience Australia.** 2019. *Thin Section Collection*. Available at <https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search#/metadata/74505> [Last Accessed 29 September 2019].
- Lang, C.** 2015. Australian Museum, May 2015. *DigiVol online volunteer evaluation report*.
- National Archives of Australia.** 2013. *Scanning specifications*, 22 August 2013. Available at http://www.naa.gov.au/Images/scanning-specifications_tcm16-93663.pdf [Last accessed 26 February 2019].
- Simmons, I.** 2012. British Geological Survey, 2012. *BGS thin sections: 150,00th image taken!*, 13 December 2012. Available at <https://britgeopeople.blogspot.com/2012/12/bgs-thin-sections-150000th-image-taken.html> [Last accessed 26 February 2019].
- Thin Sections Collection.** 2019. Available at <https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search;jsessionid=AE038BE79BC617C2C1BA839CD346BFBF#/metadata/74505>. DOI: <https://doi.org/10.26186/5c9aae9d36172> [Last accessed 28 March 2019].
- Wyborn, L and Lehnert, K.** 2016. Exploiting the Long Tail of Scientific Data: Making Small Data BIG. *Extended Abstracts, eResearch Australasia Conference, 2016*, Melbourne, Australia. https://eresearchau.files.wordpress.com/2016/03/eresau2016_paper_88.pdf [Last Accessed 29 September 2019].

How to cite this article: Pring, J, Wyborn, L and Evans, N. 2019. Reviving an Old and Valuable Collection of Microscope Slides Through the Use of Citizen Science. *Data Science Journal*, 18: 57, pp.1–17. DOI: <https://doi.org/10.5334/dsj-2019-057>

Submitted: 29 March 2019 **Accepted:** 14 October 2019 **Published:** 14 November 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.