RECORD 2023/17

THE PRESERVATION OF LEGACY COLLECTIONS PROJECT: A TEMPLATE FOR PRESERVING HIGH-VALUE COLLECTIONS FOR FUTURE RESEARCH

ER Blereau and A Bellenger

Government of Western Australia
Department of Mines, Industry Regulation and Safety

Geological Survey of Western Australia

Government of **Western Australia**
Department of **Mines, Industry Regulation and Safety**

RECORD 2023/17

THE PRESERVATION OF LEGACY COLLECTIONS PROJECT: A TEMPLATE FOR PRESERVING HIGH-VALUE COLLECTIONS FOR FUTURE RESEARCH

ER Blereau[1] and A Bellenger[2]

1 John de Laeter Centre, Curtin University, Kent St, Bentley WA 6102
2 Curtin University Library, Curtin University, Kent St, Bentley WA 6102

PERTH 2023

Geological Survey of Western Australia

MINISTER FOR MINES AND PETROLEUM
Hon Bill Johnston MLA

DIRECTOR GENERAL, DEPARTMENT OF MINES, INDUSTRY REGULATION AND SAFETY
Richard Sellers

EXECUTIVE DIRECTOR, GEOLOGICAL SURVEY AND RESOURCE STRATEGY
Michele Spencer

REFERENCE
The recommended reference for this publication is:
Blereau, ER and Bellenger, A 2023, The preservation of legacy collections project: a template for preserving high-value collections for future research: Geological Survey of Western Australia, Record 2023/17, 22p.

ISBN 978-1-74168-998-3
ISSN 2204-4345

John de Laeter Centre acknowledgement
The Preservation of Legacy Collections project was jointly funded by AuScope, the Geological Survey of Western Australia, Australian Research Data Commons and Curtin University. Isotope analyses were conducted using the SHRIMP ion microprobe facilities at the John de Laeter Centre (JdLC), Curtin University, with the financial support of the Australian Research Council and AuScope National Collaborative Research Infrastructure Strategy (NCRIS).

Acknowledgement of Country
We respectfully acknowledge Aboriginal peoples as the Traditional Custodians of this land on which we deliver our services to the communities throughout Western Australia. We acknowledge their enduring connection to the lands, waterways and communities and pay our respects to Elders past and present.

Published 2023 by the Geological Survey of Western Australia
This Record is published in digital format (PDF) and is available online at <www.dmirs.wa.gov.au/GSWApublications>.

Further details of geoscience products are available from:
First Floor Counter
Department of Mines, Industry Regulation and Safety
100 Plain Street
EAST PERTH WESTERN AUSTRALIA 6004
Telephone: +61 8 9222 3459 Email: publications@dmirs.wa.gov.au
www.dmirs.wa.gov.au/GSWApublications

Cover image: Episodic tides discharge sediments along a rocky shoreline at Cable Beach, Broome. Photo by Robin Bower

Contents

Figures

Table

Appendix

The Preservation of Legacy Collections Project: a Template for Preserving High-value Collections for Future Research

ER Blereau[1] and A Bellenger[2]

Abstract

The aim of the Curtin University Preservation of Legacy Collections project was to create a template for preserving globally significant high-value geoscience sample collections as research infrastructure. The template was developed and tested during the curation of the McNaughton Collection, which is a collection of 1095 sample mounts representing 25 years of Sensitive High-Resolution Ion Microprobe (SHRIMP) U–Pb geochronology research by Professor McNaughton and collaborators. Before this project, there was no unified approach for preserving such a research collection. This Record describes the project results and workflow as a template for preserving similar collections, and provides suggestions for similar projects. Sample metadata and associated digitized images and documents are publicly available through GeoVIEW.WA, AusGeochem and Research Data Australia. Samples were minted (the process of capturing sample metadata and linking it to data and research outputs) using the International Generic Sample Number system for persistent identification of materials. The SHRIMP mounts were donated to the State government and are available from the GSWA Perth Core Library and are being re-used for new research, highlighting the importance of curating physical sample collections and making them discoverable in the digital age.

KEYWORDS: catalogues, documentation, geochronology, information storage and retrieval, mass spectrometry, research management, sample logs, zircon

Introduction

The Curtin University Preservation of Legacy Collections project was established to secure, digitize, curate and make publicly available the physical specimens and associated records collected by Professor McNaughton and collaborators. Professor McNaughton was one of the founding investigators of the Sensitive High-Resolution Ion MicroProbe (SHRIMP) Facility established at Curtin University's John de Laeter Centre (JdLC). On his retirement in 2019, Professor McNaughton donated his collection (the McNaughton Collection) of rare and unique geochronology samples to the State government for preservation and future re-use in research.

The McNaughton Collection comprises 1095 mounts prepared during collaborative projects with government, industry and academic researchers over 25 years of operations at the Curtin University SHRIMP Facility between 1994 and 2019. More than 800 of the mounts have associated supplementary research materials, such as log files, notes, reports and images. Many of the samples are from globally significant mineral deposits that are inaccessible, or would be very costly to resample. Due to the high number of Western Australia samples, the Geological Survey of Western Australia (GSWA) agreed to host the curated collection, and make it available for future research. A primary objective of this project was to make the collection digitally discoverable and available for further research.

This project developed a template for use by other institutions, to preserve and facilitate re-use of high-value physical samples and collections. Such a template can be used by the broader Earth Sciences community who may experience similar legacy collection issues. Although government agencies have mandates and budgets to manage their geoscience collections, the university sector has restricted resourcing to preserve the collections of its prominent researchers (Simpson, 2003). The majority of institutions are unlikely to have succession plans for their key researchers who hold high-value collections and are planning to retire. This project captured metadata associated with the mounts so they could be discoverable to a wider audience.

Project establishment

In 2018, Professor McNaughton donated his collection of physical materials to GSWA. He agreed to provide support and expertise to the project group who would work to describe, process and make the collection discoverable to the general public. An agreement was reached between Curtin University and GSWA to relocate sample mounts for secure archival storage at the GSWA Perth Core Library, and for sample metadata to comply with GSWA metadata standards. Supplementary written materials and photos were scanned into digital format. Metadata and digitized materials were made publicly available through GeoVIEW.WA (<www.dmirs.wa.gov.au/geoview>), Research Data Australia (RDA; <https://researchdata.edu.au/neal-mcnaughton-legacy-collection/1457834?source=suggested_datasets>) and AusGeochem (<https://ausgeochem.auscope.org.au>).

1 John de Laeter Centre, Curtin University, Kent St, Bentley WA 6102
2 Curtin University Library, Curtin University, Kent St, Bentley WA 6102

Financial support was provided to Curtin University's JdLC by GSWA and AuScope. Curtin University provided in-kind financial support for Professor McNaughton's time in organizing the collection, and Professor McInnes (JdLC Director) in his role as Project Director. The two aims of the project were to report the number of physical samples curated and archived in the GSWA Perth Core Library, and the number of sample records discoverable online via data portals and digital repositories. The project sought to adhere to the FAIR (Findable, Accessible, Interoperable and Reusable) Data Principles, with the published dataset made available under a Creative Commons Attribution (CC-BY) license with use of persistent identifiers (digital object identifiers, DOIs, and handle services) where possible.

The project took about 18 months from securing funding to making the records discoverable to the public. The project was overseen by a steering committee comprising JdLC and GSWA representatives. Operational matters were managed by a project committee comprising technical staff and librarians. Work to describe the collection, collect and collate metadata, digitize the supplementary materials, and process the physical mounts for relocation involved a team of two researchers and five research assistants recruited through *Curtin University's Earn While You Learn* program, a campus employment initiative to help students build skills while undertaking university studies. Library staff provided advice on metadata standards for discovery platforms, the minting process using the International Generic Sample Number (IGSN; Klump et al., 2021, <www.igsn.org>), data publishing, and project reporting and statistics. An IGSN is a globally persistent and unique identifier for physical materials that is connected to a digital metadata description and a record of any parent or daughter material associations (e.g. a mineral separate, daughter, made from X sample, parent; Klump et al., 2021). IGSNs were used in this collection to ensure the samples were readily identifiable worldwide and to ensure the preservation of the collection into the future.

Project workflow

Sample metadata cataloguing

All identifying information relating to the mounts and their associated samples were compiled into a Microsoft Excel spreadsheet. These records contained:

- the distribution of samples within the mount
- names of the researchers involved
- sample numbers
- mineral(s) within the mount (e.g. zircon, monazite, rutile).

Additional information (e.g. lithology, sample location) was added with each column representing a required metadata field for GSWA, other discovery platforms (e.g. GeoVIEW, AusGeochem, RDA), and working fields to process the records for IGSN minting (discussed below). This spreadsheet would become the Legacy Catalogue for the project, an inventory of all mounts, samples and associated publication records. Figure 1 summarizes the workflow used for compiling the Legacy Collection and Table 1 details the metadata fields included in the Legacy Catalogue. The Legacy Catalogue is discoverable from GeoVIEW.WA, RDA and AusGeochem.

Each mount and sample had a unique number used by Professor McNaughton and recorded in his notebooks. These identifiers were used to search external data sources such as research theses, published journal articles and published/unpublished reports for information including lithology and sample location. The most useful identifier in published works was the mount number, although this was referenced in less than 50% of publications and not always in the main text. Sample numbers often contained subtle variations between the notebooks and publications (e.g. MG1 vs. MG-1). This will likely be an inherent problem in any subsequent data source containing work by multiple researchers. Some flexibility is required when searching the collection for a specific item, although searching for the mount number or involved researcher may be the most direct approach. It is possible that mount numbers could have the same issue seen with sample numbers if alphanumeric (e.g. BR18-01, BR18-1, BR-18-1) although this was not an issue for this collection.

Once a data source was identified, the sample location was added to the spreadsheet. Some publications quoted coordinates in the text or data tables. However, many locations had to be estimated from journal or thesis figures. Other samples, particularly relating to drillholes, were given generic locations, such as the mine site, as the drillhole name was not always provided, was abbreviated, or could not be located from publicly available information. This issue remains prevalent in research publications and journal authors, reviewers and editors have a responsibility to provide accurate sample location data. All sample locations were converted to decimal degrees for consistency and for ease of plotting.

GSWA has an interconnected, dual level of metadata for a sample's rock type within their digital sample catalogue of field observations, WAROX for 'Western Australian Rocks' (Riganti et al., 2015). Two specific lists of possible naming choices are used in-house by GSWA, with the selection from the initial broader rock type list (e.g. igneous granitic) limiting selections from the latter more specific lithology list (e.g. monzogranite). To conform to GSWA metadata requirements, this lithological hierarchy was implemented when registering McNaughton Collection samples. Where only one field for the rock type was presented in the collection, the more specific lithology such as 'monzogranite' was used rather than the more general descriptor.

A metadata field was included for the nature of the sample mount, that is, whether the mount contains separated mineral grains or pieces of petrological thin sections. Although mounts containing collections of grains are most common, with the increased emphasis on in situ microanalysis, it was important to distinguish mounts that contained drilled or cut out portions of thin sections, allowing researchers to determine whether a mount may suit a variety of needs.

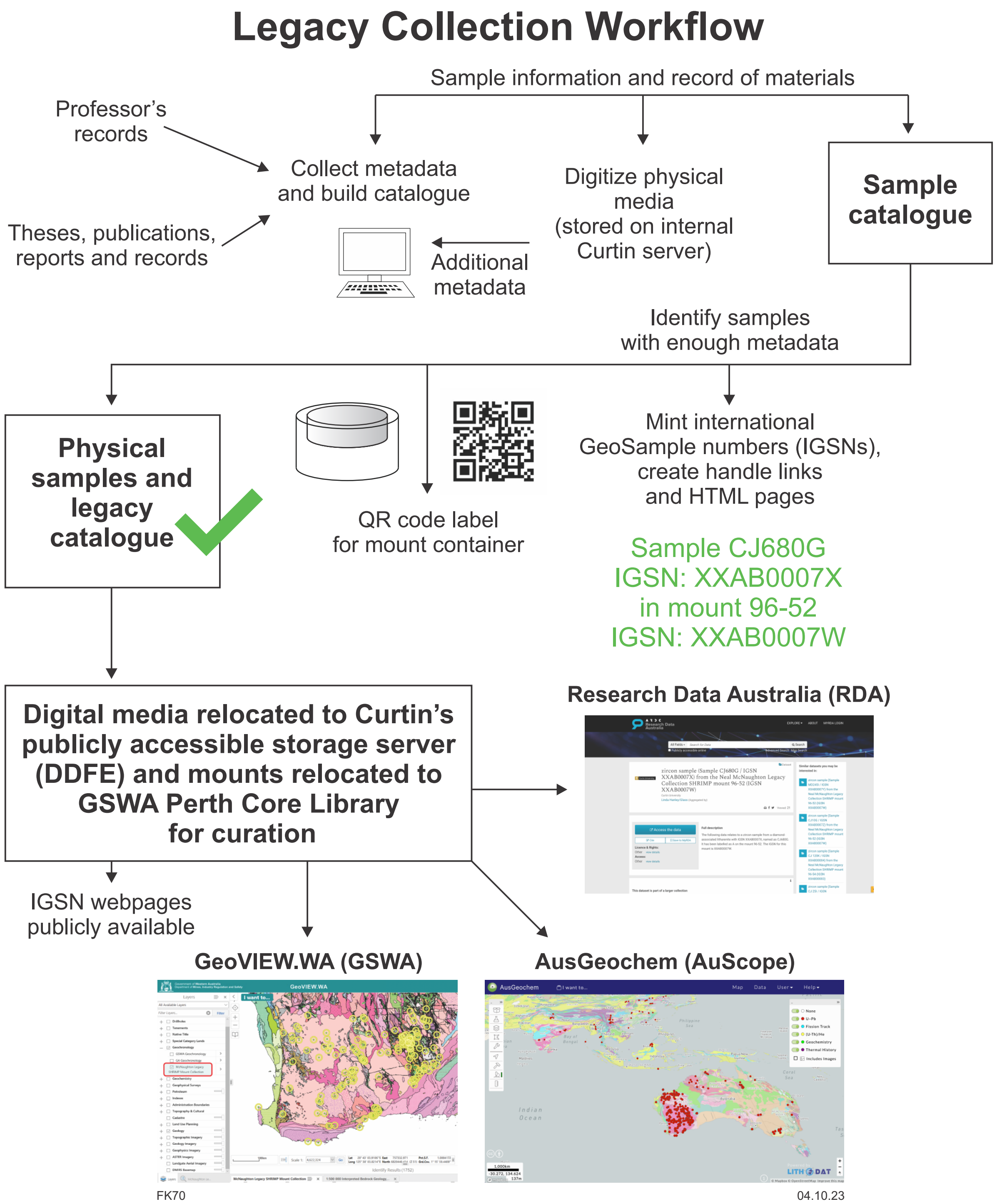


Figure 1. Workflow used in the Preservation of Legacy Collections project to compile and collect metadata, create the catalogue, digitize media and publish online

Table 1. Legacy catalogue metadata fields and descriptions

Metadata field	*Description*
Location	Abbreviation for filtering data by broad location (WA - Western Australia, AUS - Rest of Australia, INT - International)
SHRIMP Mount ID	ID number of the SHRIMP mount
Mount label	Subsection within the sample mount (usually letters or numbers)
Sample ID	ID number of the sample
Grains or thin section	Whether the mount contains multiple grains or pieces of thin section slides
Mineral one	Analysed mineral
Analysis age 1	Geochronological result
Uncertainty	Uncertainty on geochronological result
Confidence level	Confidence level of result
Event dated	Interpretation of result
Mineral two	Secondary mineral or same mineral but second age etc
Analysis age 2	" "
Uncertainty	" "
Confidence level	" "
Event dated	" "
Mineral three	" "
Analysis age 3	" "
Uncertainty	" "
Confidence level	" "
Event dated	" "
Rock type	Broad geological rock type (e.g. igneous granitic, sedimentary siliciclastic etc)
Lithology	Specific rock name
Tectonic unit	Tectonic unit that the sample lies within
Stratigraphic unit	Stratigraphic unit that the sample lies within
General location	Further location details, such as a mine site
Lat	Latitude in decimal degrees
Long	Longitude in decimal degrees
Location method	What type of information was used to locate the sample: Mine site, GPS, from a topographic map etc.
University of Primary/Secondary Researcher	University researchers who were involved at the time the mount was made
Primary Researcher First Name	Name of primary researcher involved
Primary Researcher Surname	
Secondary Researcher First Name	Name of secondary researcher involved
Secondary Researcher Surname	
Published	Paper (P), Thesis (T), Unpublished (U)
Publication information	Citation information for publication
Mount avaliable	Yes (in the possession of Neal McNaughton and now GSWA), No (Mount has been previously borrowed and not returned or destroyed – noted in comments)
Digitized information of the whole mount	Yes or No; Digitized imaging at the most zoomed out level, i.e. the entire mount
Image type	Transmitted light (TL), Reflected light (RL), Backscattered Electron (BSE), Cathodoluminesence (CL), Secondary Electron (SE)
Digitized information of a whole sample	Yes or No; Digitized imaging of a subsection of the mount
Image type	Transmitted light (TL), Reflected light (RL), Backscattered Electron (BSE), Cathodoluminesence (CL), Secondary Electron (SE)
Digitized information of a few grains or grain	Yes or No; Digitized imaging of very zoomed in sample information
Image type	Transmitted light (TL), Reflected light (RL), Backscattered Electron (BSE), Cathodoluminesence (CL), Secondary Electron (SE)
Digitized documents	Yes or No; Digitized documents (e.g. SHRIMP analytical log sheets)
Comments	Any additional information or notes
Mint	1 (mint IGSN), 0 (Do not mint IGSN)
Parent IGSN	IGSN pertaining to the SHRIMP mount as a whole
IGSN	IGSN pertaining to an individual sample within the SHRIMP mount
IGSN Handle	Handle URL link for the sample IGSN

A 'general comments' field was included to record any additional text-based metadata that did not fit within other metadata fields. This included information on physical materials not digitized, the presence of physical storage devices, drillhole numbers if different from sample numbers, additional sample numbers if more than one was assigned to a single sample, and any other information that was deemed useful.

Digitization of ancillary physical records

Approximately 75% of mounts in the McNaughton Collection had an accompanying box of physical media (Fig. 2a) containing records such as sample images and mount maps (e.g. backscattered electron, BSE; transmitted and reflected light), log sheets of analytical sessions and storage media including floppy disks and CD-ROMs. The physical materials were examined and the types of materials for each sample in the mount identified and recorded in a physical media cataloguing spreadsheet. Ancillary records such as draft reports, emails, notes and annotated figures were used to obtain sample location information. After cataloguing, physical materials deemed most useful to future research were then digitized.

All images and analytical logs were scanned as a PDF at 600 dpi. The files were scanned to Curtin University's internal research storage drive and later transferred to Curtin University's DDFE server for public access (see below). Due to time constraints on the project, analytical data printouts were not digitized although are stored in the physical media storage facility as the associated data file is likely retained by the original researcher or has been published.

Once physical records were digitized, they were returned to a storage facility at The University of Western Australia (UWA) in Shenton Park, where a sign providing information about the project, including the objectives, sponsors and collection-level DOI link was posted in front of the shelved material to provide context for the physical items (Fig. 2d).

The digitized files are stored on Curtin University's DDFE server and are accessible via <http://ddfe.curtin.edu.au/gswa-library/mcnaughton/>. The DDFE is a data delivery front-end server that uses a web interface to allow public access to Curtin University datasets. The system is managed by the Curtin University Digital & Technology Solutions group and runs on the RedHat Enterprise Linux operating system. The McNaughton Collection URLs for the DDFE file locations use a standard format including the IGSN. Landing pages were created at the collection-level, mount-level and sample-level on the DDFE, using a template and program that populated the template with mount-specific information such as lithology, IGSN and sample number. The index page also described the project, citation information, licensing information and was linked to the complete spreadsheet as stored on the DDFE server. Digitized files are available for download on the mount-level pages. As files were loaded onto the DDFE server pages, a custom program developed in-house updated the Legacy Catalogue spreadsheet with the URL relating to the DDFE file location.

Processing the metadata

Early in the project, it was recognized that the metadata records in the McNaughton Collection would not be fully compliant with the contemporary recording standards in GSWA metadata templates. When a section of the spreadsheet (mounts for one year) was deemed as complete as possible, the entries were checked for consistency of terminology and formatting, and tectonic and stratigraphic units were updated for Australian samples where regional geology in publications or theses had since been updated. Samples were categorized in the Legacy Catalogue spreadsheet using three systems; 1) colour-coding for the complete, partially complete or incomplete level of metadata capture relating to the records; 2) an additional column with either a '1' or a '0' indicating either to 'mint' or 'not mint' an IGSN; and 3) to indicate whether the sample originated from Western Australia, the rest of Australia, or an international location, which provided the processing staff an indication of priority for processing the records. As this project was funded by state and federal government agencies, all Australian samples were prioritised. Mounts with complete or partially complete metadata (i.e. rock type and sample location at a minimum) were flagged for the minting of IGSNs and long-term storage labels.

Due to the complexity of the collection from the materials produced by multiple researchers, and mounts containing multiple samples often from different locations, each sample was minted two IGSNs: one at the mount level (parent) and one at the sample level. IGSNs facilitate the discoverability of records. The minting of IGSNs was undertaken using the IGSN Application Programming Interface (API) service provided by the Australian Research Data Commons (ARDC, www.ands.org.au/online-services/igsn-service). An API was chosen rather than the ARDC IGSN Minting Service web interface due to the large number of records and to reduce the amount of manual workflows. ARDC's API is a customized version of software developed by CSIRO. Use of the API required liaison with ARDC to inform them of our intention to use the service and to register the institutional account, including the IGSN prefix to be allocated to the collection.

As part of the minting process, specific metadata fields were provided to the IGSN registry describing the object type, resource title and type, material type and curator. These fields were extracted from columns in the Legacy Catalogue and converted into the CSIRO IGSN Schema (version 3.0) in XML format, from the website <https://igsn.csiro.au/schemas/3.0/doc/>, to parse the information through the API, from the website <https://support.datacite.org/docs/registering-igsn-ids-1>. To mint IGSNs, IGSN identifiers must resolve (i.e. redirect to the latest information about the object it identifies, even if the object changes or moves) to a metadata record that describes the sample. The IGSN API process used the current DDFE file location in the XML output used for the ARDC's minting service. For the final version of the Legacy Catalogue spreadsheet and metadata published in GeoVIEW, RDA and AusGeochem platforms, the DDFE file location was not provided; however, it was replaced with a 'handle' link that was generated at a mount and sample level. The use of a handle link means a more persistent URL, as the direct URL for the file location may change over time whereas the handle link can be modified to redirect to the current location. Handle links were created using the free software Handle.Net Software (version 9.2) supplied by the Handle.Net Registry (HNR) from the website <www.handle.net/download_hnr.html>.

Figure 2. Selected photographs of physical items in the McNaughton Collection and processed as part of the Preservation of Legacy Collections project: a) physical media boxes at the UWA storage facility after digitizing, stacked two deep; b) an example of a digitized SHRIMP mount map; c) SHRIMP mounts with QR code labels and those flagged to receive labels; d) a sign posted in front of the boxes in storage after cataloguing. This sign points any visitors towards the collection and the digitized information via QR code

In 2021, IGSN entered into a partnership with DataCite requiring IGSN identifiers to be transitioned to the DataCite infrastructure. During the migration process in 2022–23, the management of the IGSN identifiers of the McNaughton Collection was also transitioned. The 10273/XXAB IGSN namespace prefix was previously managed by ARDC with the Curtin University Library having ownership of the handle IGSN identifiers. Curtin University and ARDC handed over the 10273/XXAB legacy handle IGSN identifiers to Lithodat for re-registration and ongoing management in April 2023.

Publishing the dataset

Curtin University Library provides a research data service to enable Curtin University researchers to openly share their data, and meet data-sharing requirements of funders and publishers. The library has an established process for publishing datasets to RDA on behalf of Curtin University researchers using its ReDBox/Mint installation. Datasets are published to RDA only, as there is no data discovery platform with institutional branding that is specific to Curtin University datasets. The data publication service includes DOI minting and guidance on metadata and licensing. Data files are stored on the DDFE server, which is under the joint custodianship of Curtin University's Digital & Technology Solutions group and the library. ReDBox is open-source software to manage metadata and publish it to online repositories. Mint is an authority control service that ensures that each authorized heading for a proper name or a subject (e.g. 'people' and 'groups') is consistent. It was decided that ReDBox would not be used due to the large number of records within the collection.

The McNaughton Collection comprises 1095 physical mounts, with 741 mounts having sufficient metadata to describe them on the GeoVIEW.WA and RDA discovery platforms. The Legacy Catalogue spreadsheet comprises approximately 4500 listings at sample and mount level, with a subset of 2433 metadata records contributed to RDA with accompanying IGSNs and DOIs minted.

Instead of using ReDBox, the library elected to mint DOIs using the API service provided through the ARDC as a member of the DataCite initiative, and then to contribute McNaughton Collection records through the RDA user interface directly by uploading XML files with the metadata in ARDC's required RIF-CS format. DOI minting required the library to register test and production accounts with the ARDC.

Prior to generating the records in RIF-CS format, the library team had to associate 'groups' and 'people' with the record set using Mint. As the collection comprises records relating to researchers from various institutions including Curtin University and UWA, and researchers with unknown affiliations, it was decided to associate all researchers with the JdLC as a research group. Due to the lack of detailed metadata regarding individual researchers, the researcher records included their first and last name, and affiliation with JdLC, with no unique linking information, such as ORCIDs (open researcher and contributor identifier) or contact details. This was due to limited resources available to process the dataset. Researchers are added as 'people' to Mint, which then generates this information as a People.xml file in RIF-CS format and presents this information for harvesting by RDA. Mint takes a sequence number relating to 'people' records and generates a standardized URL for each researcher (in the format: curtin.edu.au/parties/people/<seq-num). A 'group' key is used to represent the group record for JdLC. The library staff completed a curation process in Mint, which linked new 'people' records to the 'group' record, so the records would show up in the subset of McNaughton researchers and be sent through using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) feed to RDA.

Following the completion of the association of records with 'people' records and the JdLC 'group' record, RIF-CS formatted records were generated using code developed at the Curtin University Library. As part of this process, sample metadata export files were generated to upload to RDA to test the upload process. Records were contributed with two hierarchical levels, with the primary level being the collection, and the secondary level being the mounts and samples. The final result, from initial stages of metadata capture to publication of records to RDA, took 18 months to complete. The final stages of contributing records were particularly challenging due to COVID-19 restrictions.

The final stage of publishing records to discovery platforms was the contribution of records to AusGeochem, an AuScope geochemistry data repository funded via the National Collaborative Research Infrastructure Strategy (NCRIS). The AusGeochem platform displays geospatial data aggregated from 12 Australian research institutions on a geographic map (Fig. 3). Records that are captured by GSWA's interactive online geological platform GeoVIEW.WA include only Western Australia samples.

Curation and archiving of the McNaughton Collection

Within the scope of this project was the physical preservation of the sample mounts and relocation from Curtin University to GSWA facilities. Curtin University archival experts recommended storing the mounts in a 3.7 cm diameter acrylic storage case with a foil-backed archival-quality label, showing mount number and a QR code, attached to the outside (Fig. 2c). See the Appendix for a detailed procedure of how the QR codes and labels were generated. The QR code directs the user to the mount IGSN handle link, which in turn directs them to the online landing page for the mount and any digitized materials. A paper copy of the QR code label, without any adhesive, was printed and placed inside the mount container as a backup label. QR code labels were created for all Australian mounts with at least a minimum level of metadata and for international sample mounts with at least a minimum level of metadata that were also physically available as part of the collection (as some sample mounts were not held with the collection and presumably with the researcher or unable to be located). Mounts that did not have the minimum level of metadata and were held with the collection were kept in containers with the mount number on the outside, and on a slip of paper on the inside. Once mount samples were in their acrylic container with labels, they were relocated from Curtin University to the GSWA Perth Core Library for storage and potential re-use via the GSWA sample access procedure.

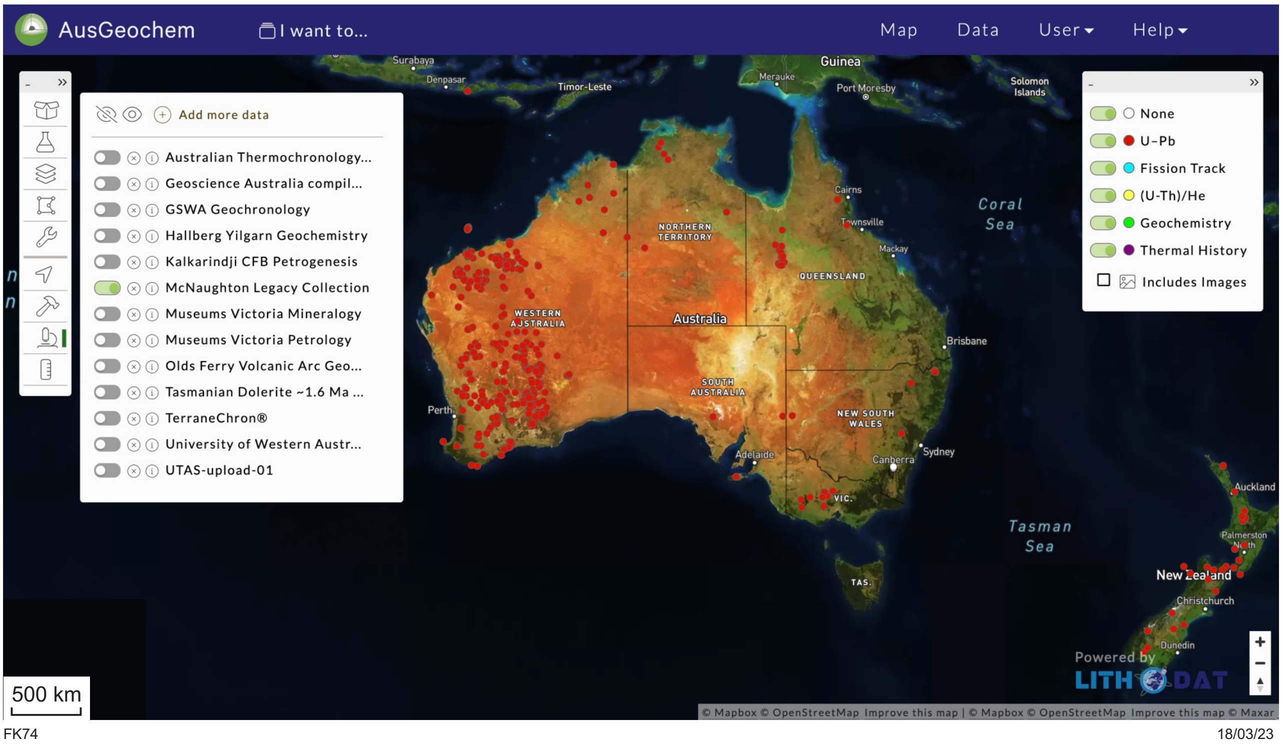


Figure 3. An AusGeochem web portal screen view of sample points from the McNaughton Legacy Collection in Oceania. Each dot represents an individual geological sample that is correlated with the related SHRIMP mount and sample metadata

Researchers using mounts on loan from the McNaughton Collection are requested to cite the collection (i.e. McNaughton, 1994–2019), cite this record on publication, use the IGSNs and mention the following paragraph (or a shortened version without ORCIDs if required) within the acknowledgements of any resulting publications:

'Mount XX used in this project is part of the McNaughton Collection, which was created by the John de Laeter Centre (Prof. Brent McInnes, 0000-0002-2776-0574; Dr Eleanore Blereau, 0000-0001-8850-397X and Prof. Neal McNaughton, 0000-0002-4384-0128), Curtin University Library (Amanda Bellenger, David Lewis, John Brown, and Colin Meikle) and the Geological Survey of Western Australia (Dr Michael Wingate, 0000-0003-2528-417X). The project was jointly enabled by NCRIS via AuScope, GSWA and Curtin University.'

Issues encountered

The Legacy Catalogue was built from notes containing basic information such as: mount number, named researchers, mineral(s) within the mount, sample numbers and a sketch of the mount. Without this basic level of information, the creation of the Legacy Catalogue would not have been possible. Where samples lacked location information or descriptions, this information was obtained by reading and extracting information directly from theses (PhD, Masters and Honours) and journal articles containing the geochronology data. Not all theses were available in digital format or physically available at university repositories.

Issues encountered with the physical materials included the degradation of elastic bands resulting in materials becoming disconnected and stained. Mount image maps created by joining multiple 6-inch by 8-inch photographs together (Fig. 2b) were often damaged from folding and storage, making it difficult to produce reliable digital products.

A key cataloguing issue involved named researchers changing jobs or institutions and the time between when data was collected and published (in some cases, more than 20 years). The majority of the catalogue information missing or assumed lost from the collection is due to researchers leaving academia before publishing their data. The project did not have the necessary resources to contact all named researchers.

Another complication preventing the preservation of information was the variation of sample numbers within publications relative to the sample number specified on the mount and in the Legacy Catalogue. Many researchers use multiple sample numbers, including different or general names for their samples (e.g. sample 1, sample 2) that do not correspond to those connected to the mount. There were also instances where samples did not have unique identifiers, or were marked as 'miscellaneous', which had implications for programming created by the project to work with APIs resulting in the inability to match samples with mounts or related publications (for instance). Published work often did not list the mount numbers within their data tables or within the manuscript text. Unless the researcher mentions the mount number, it was not possible to definitively connect the data in these papers to the Legacy Catalogue. A problem also occurred during IGSN minting where instances of samples having the same sample number or name (e.g. same rock sample; however, different mineral population with no change in sample number) resulted in the created IGSNs and associated metadata being misaligned.

Many publications and theses connected to samples in the catalogue did not list sample locations accurately, if at all. Some publications only showed sample locations on maps without any latitude and longitude information or that were not readily relatable to a geospatial location. In cases where generalized locations were given, a location area was used instead of a singular point.

When dealing with large collections of materials and spreadsheets, it is possible to overlook some metadata or create incorrect entries through transcription error. To mitigate against this, several quality control checks were undertaken during the digitization and cataloguing process, and during the creation of mount labels. When cataloguing or digitization of a particular sample was completed, a checklist spreadsheet was annotated and the presence of digital files was confirmed to make sure that no materials were missed or misidentified. When mount labelling began, this was an opportunity to check if IGSNs were required and if the mounts were available or being held with the related researcher.

When computational procedures were used to process and publish the catalogue, metadata quality control checks were used to identify where sample locations were outside the latitude and longitude range of Australia, and inconsistencies in how named researchers were recorded in the spreadsheet (e.g. CJ Adams, C.J. Adams). In terms of the DDFE process of moving digitized files to the repository portal, a bottleneck occurred due to the large size of the PDF files created during the scanning process. For future projects that include images, it may be more adventageous to use file formats such as TIF or JPEG to speed up this process.

Where sample metadata and analytical data were available and physical mounts were missing from the collection and could not be assigned an IGSN, the orphaned materials were digitized and saved in Curtin University's research storage server until such time that an IGSN may be minted for the associated mount and the corresponding materials be placed online. This approach meant that the research preservation of the collection was as complete as possible and future-proofed the collection if a missing mount was returned.

Recommendations

1. Record-keeping: there is a need for improved record-keeping of sample collections by academic researchers, including the following sample metadata as recommended by Boone et al. (2022):
    a. sample name (sample number assigned at the time of collection)
    b. sample type (description of the type of sample taken, for example, outcrop, bore hole)
    c. lithology or mineral (identified rock type or mineral)
    d. location (latitude and longitude in decimal degrees using WGS84 reference frame)
2. Use of persistent identifiers, such as IGSNs: this will ensure that sample numbers are unique.
3. Project infrastructure: the operation and organization of this project aimed to minimize the required computational and digitization infrastructure due to the short duration of the project. The downside of infrastructure limitations is the need to create greater efficiencies in the use of staff time. Building automated quality control processes into the earliest stage of the project mitigates risks associated with transcription and labelling errors.
4. Project team: it is highly desirable to carry out the archiving of research collections where the principal researcher is still actively practicing research. In terms of selecting the archival team, it is helpful to employ people with direct knowledge of the collection materials, knowledge in geology and geochronology, and scanning electron microscopy being beneficial although not essential. This will allow work to be done with autonomy as individuals or as part of a team with handovers, with no training required to identify materials other than to outline the classification plan for the project and desired products. The amount of metadata capture within this collection was the most significant portion of work, both in terms of digital and physical materials. The varied nature and the exact amount of work was not entirely apparent from the start of the project and would be a likely problem for other collections, particularly if information is not clearly recorded. More staff may help speed up the process although will require careful liaising and tracking of what metadata has been collected to avoid duplication. Research collection efforts, in particular digitization efforts, suffer from the dilemma of infrastructure vs time, in particular those involving more than one researcher.
5. Project management and resourcing: there can be a strong desire to continuously improve collection catalogues although this process eventually leads to diminishing returns. It is important therefore for project managers to set a clear end point for project completion.
6. Project dissemination: future projects should endeavor to advertise the generated collections and perpetuate any new advances and suggestions. Successful formats used in this study included conferences and webinars.

Communication

An important task after the completion of the project has been communicating the results to the wider research community. Key information to communicate included the method, the overall success of this project and the availability of the materials for future research. This project has been presented at research conferences such as eResearch Australasia in 2020 <https://conference.eresearch.edu.au/preservation-of-legacy-geoscience-collections-the-mcnaughton-collection-as-a-template-for-preserving-high-value-collections-for-future-research/> and at EGU 2021 (Blereau et al., 2021). An online webinar on the project was also given by Dr Eleanore Blereau as part of the AuScope Geochemistry Network webinar series and is publicly available on YouTube <www.youtube.com/watch?v=IwBM2rgZKGQ&t>. Based on general feedback, communicating our results via the conferences and webinar has resulted in an overwhelmingly positive response and will likely lead to similar projects.

Acknowledgements

We would like to thank Professor Neal McNaughton (ORCID: 0000-0002-4384-0128) who served in the role of SHRIMP geochronologist for the UWA Centre of Global Metallogeny (1994–2005), the UWA Centre for Exploration Targeting (2005–07), and the John de Laeter Centre at Curtin University (2007–19). Dr Michael Wingate, GSWA, (ORCID: 0000-0003-2528-417X) provided advice throughout the project and assistance in arranging housing of the physical mount collection at the Perth Core Library. The project sponsor was Professor Brent McInnes (ORCID: 0000-0002-2776-0574; John de Laeter Centre) who received funding from the NCRIS AuScope Earth Composition and Evolution program, GSWA, the Australian Research Data Commons and Curtin University.

Ms Amanda Bellenger oversaw the library team as Library Project Manager. Peter Green, David Lewis, John Brown and Colin Meikle contributed their librarian and data management expertise throughout the project. Dr Eleanore Blereau oversaw the metadata collection and managed the digitization aspects of the project as JdLC Project Manager. Shereen Roy, Ravi Patel, Hannah Whittaker, Brenton Lynn, Dr Nicole Nevill and Payal Panchal were responsible for the review, digitization, labelling and curation of physical materials. Dr Alexander Prent facilitated the upload of the McNaughton Collection to the AuScope AusGeochem data repository.

Professor Brent McInnes, Dr Alexander Prent and Dr Bryant Ware are thanked for editing the original version of this manuscript.

References

Blereau, ER, Bellenger, A and McInnes, B 2021, Preserving high value legacy collections for future research: the McNaughton Collection, EGU General Assembly 2021, online, 19–30 April 2021, EGU21-14458, doi: 10.5194/egusphere-egu21-14458.

Boone, SC, Dalton, H, Prent, A, Kohlmann, F, Theile, M, Gréau, Y, Florin, G, Noble, W, Hodgekiss, S-A, Ware, B, Phillips, D, Kohn, B, O'Reilly, S, Gleadow, A, McInnes, B and Rawling, T 2022, AusGeochem: an open platform for geochemical data preservation, dissemination and synthesis: Geostandards and Geoanalytical Research, article no. ggr.12419, 15p., doi:10.1111/ggr.12419.

Klump, J, Lehnert, K, Ulbricht, D, Devaraju, A, Elger, K, Fleischer, D, Ramdeen, S and Wyborn, L 2021, Towards globally unique identification of physical samples: governance and technical implementation of the IGSN Global Sample Number: Data Science Journal, v. 20, no. 1, p. 33, doi: 10.5334/dsj-2021-033.

McNaughton, N 1994–2019, Curtin University Preservation of Legacy Collections project, The University of Western Australia, Curtin University and Geological Survey of Western Australia, doi: 10.25917/5e7aee0406487.

Riganti, A, Farrell, TR, Ellis, MK, Irimies, F, Strickland, CD, Martin, SK and Wallace, DJ 2015, 125 years of legacy data at the Geological Survey of Western Australia: capture and delivery: GeoResJ, v. 6, p. 175–197, doi: 10.1016/j.grj.2015.02.015.

Simpson, A 2003, The plight of geological collections in the Australian tertiary education system: Museologia, v. 3, p. 37–44.

Appendix 1: Guide to creating archival labels with QR codes

1. Creating/Finding what mounts need labels

Step 1.

Open or copy the Legacy Catalogue as required.

Step 2. Formatting spreadsheet

The columns that are needed are **Shrimp Mount ID, Mount Label, Mount Available and Mint.**

Hide all other columns. **Option:** make a new column labelled QC comments if you want to make notes of anything along the way.

SHRIMP MOUNT ID	MOUNT LABLE	MOUNT AVALIABLE (Y/N)	MINT	QC Comments
05–01	A		0	
05–01	B		0	
05–01	C		0	
05–01	D		0	
05–02	A	Y	1	
05–02	B	Y	1	
05–03	A		0	
05–03	B		0	
05–03	C		0	
05–04	A-M	Y	0	
05–05	A	Y	0	
05–05	4–11	Y	0	
05–06	A-D		0	

Step 3. Physical Check

Retrieve the relevant physical boxes of mounts.

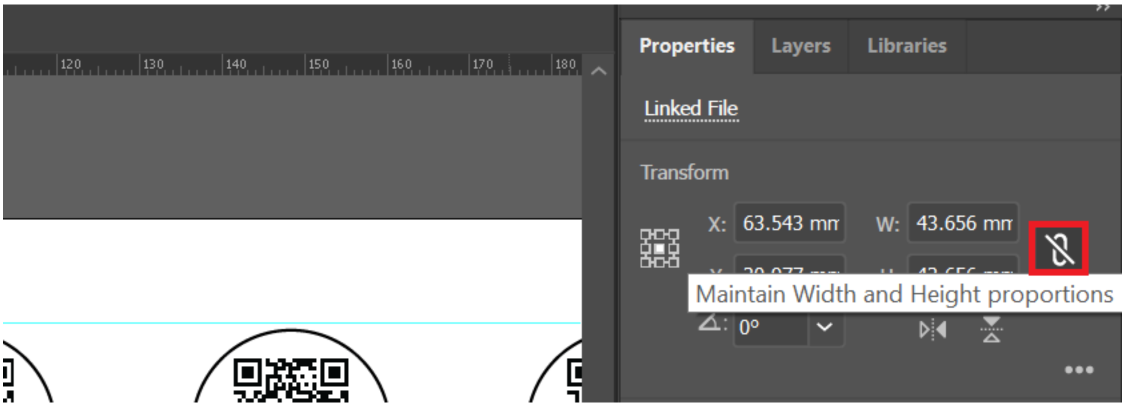
This step is a quality control check between the availability of the mounts and the catalogue.

As you go through the box look for anything that doesn’t match the spreadsheet.

Examples are mount availability is incorrect, mint status is not 1 or 0. Mount container is empty but marked as present, or anything else that is inconsistent.

For all mounts being minted, mark them with a Post-it note sticker or anything else (This helps with another QC step when applying labels further along).

In the Properties Tab first select the link W and H option, this will mean you only need to change one and not the other.



Now change the width to **16MM** now the QR code is the correct size drag it onto the correct mount in the correct place. Once you have the first one placed you will see the next QR code snapping to it along common axis with a pink line. This means once you have the top row you should have a perfect reference for all the others.

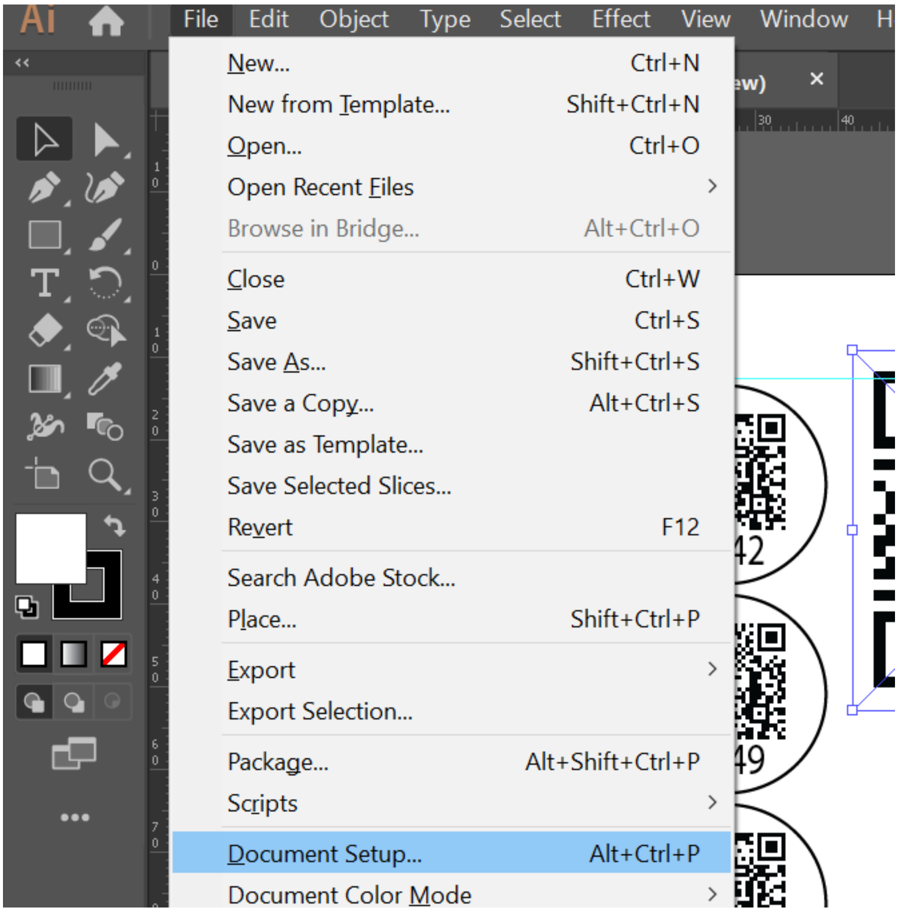
Note:

Save the sheet and move it to the completed folder for American letter size (the size of our archival sticker sheets).

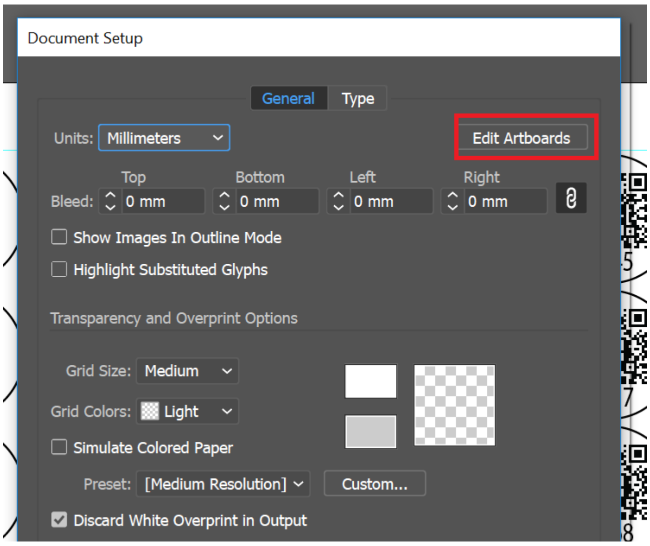
Now you need to copy the sheet and make an A4 size sheet as well as we are making a sticker label as well as a paper label in the container. There is another folder in the R drive for A4 paper sheets.

Open the copy of your American letter sheet

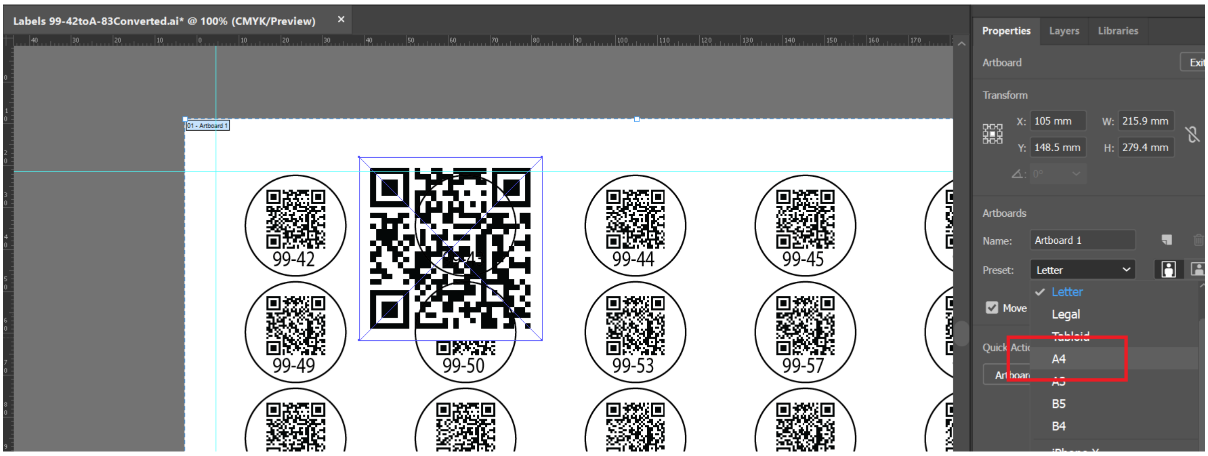
And select file and document setup



Select Edit artboards

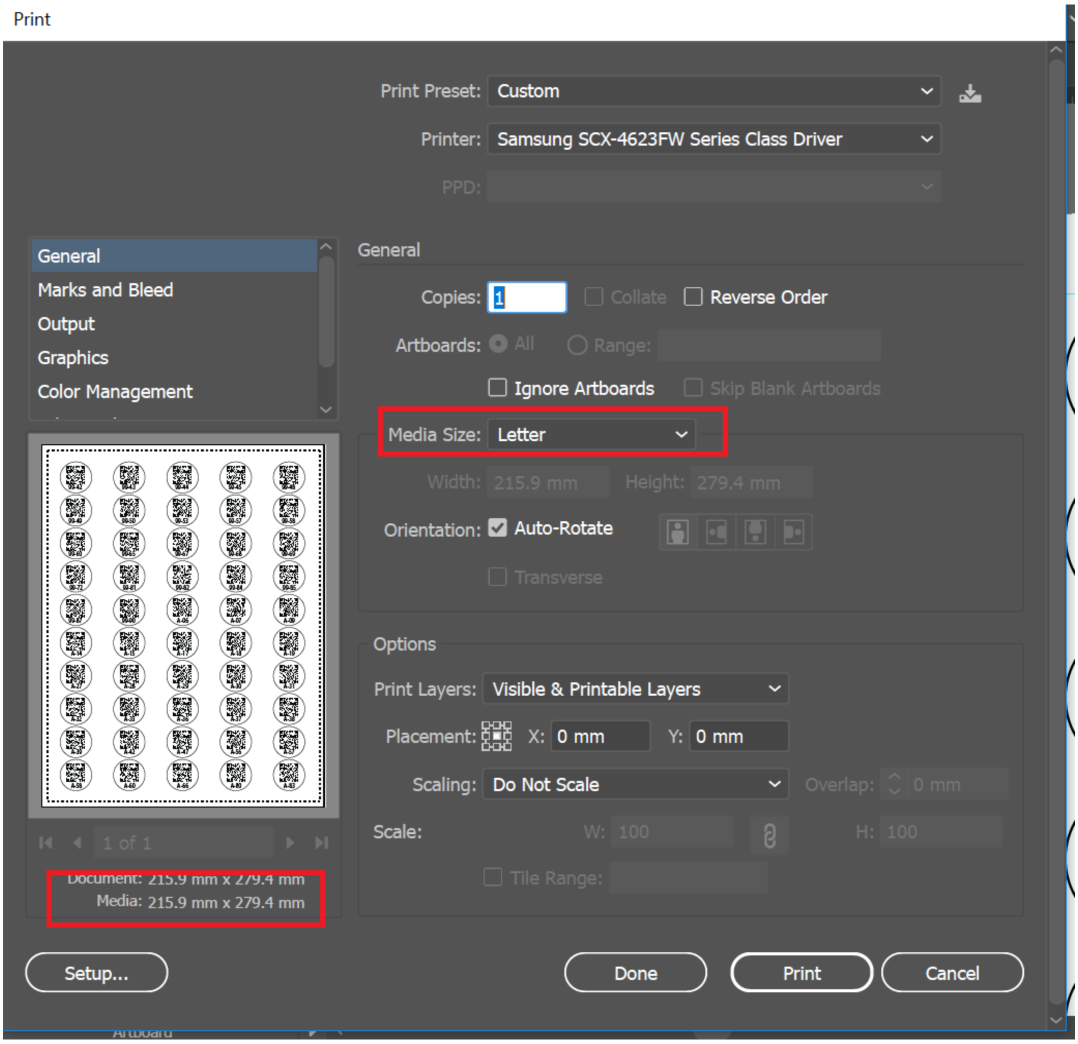


Now under properties and change the preset to A4



Save, you can now print both.

When printing make sure under the media type and dimensions are correct, you don't want it applying any scales.



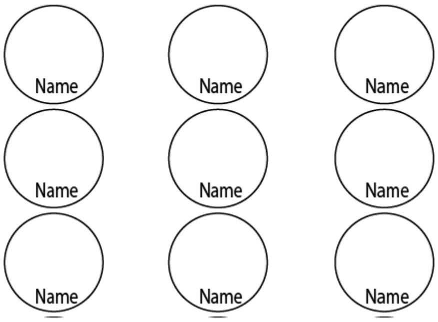
Step 4. Start the creation of labels part 1

We used Adobe Illustrator for the creation of the labels following an initial template to match the storage containers.

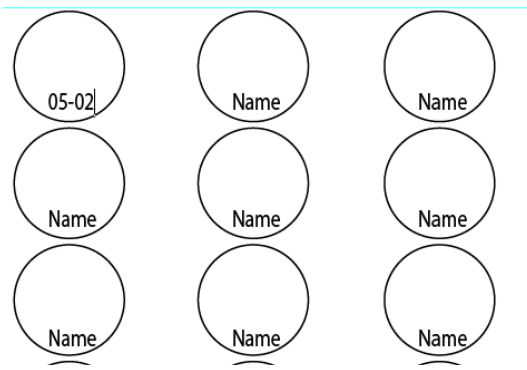
Copy the label template illustrator file and rename it to match the samples in question.

Rename each part of the template to a shrimp mount ID that will be minted and requires a label.

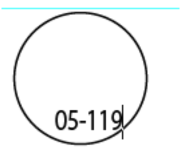
Before



After



Note : Because Mount Id can contain different amounts of characters and different characters are different sizes the labels may have to be manually adjusted. Just make sure it's all inline once completed.

Example 

In this case a 6 letter label causes the label to be too close to edge, realign as required.

Make sure to save often.

2. Making batches of QR codes

Starting Note

This part requires the highest precision, basically if you generate a wrong QR code and don't realise it's very hard for someone else to find it until they go to scan it. You can always download a QR scanner on your phone to make sure everything is going to the right place. You can only make up to 100 at a time, so to keep it simple and generate one year at a time.

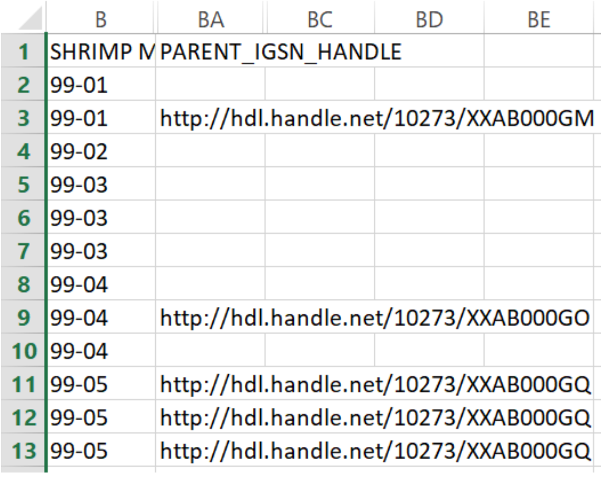
Step 1. Formatting the excel spreadsheet for generation.

To quickly create a lot of QR codes in one shot, the data must be formatted in a particular way.

Open the spreadsheet for the year you wish to generate with the IGSN handle links. You may want to make a copy to not alter the original.

You only need the columns **Shrimp Mount ID, Parent_IGSN_Handle.**

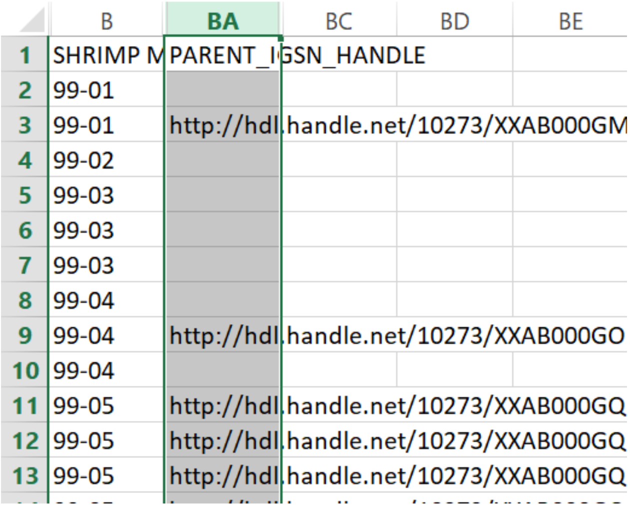
Hide every other column so it looks like the below.



	B	BA	BC	BD	BE
1	SHRIMP M	PARENT_IGSN_HANDLE			
2	99-01				
3	99-01	http://hdl.handle.net/10273/XXAB000GM			
4	99-02				
5	99-03				
6	99-03				
7	99-03				
8	99-04				
9	99-04	http://hdl.handle.net/10273/XXAB000GO			
10	99-04				
11	99-05	http://hdl.handle.net/10273/XXAB000GQ			
12	99-05	http://hdl.handle.net/10273/XXAB000GQ			
13	99-05	http://hdl.handle.net/10273/XXAB000GQ			

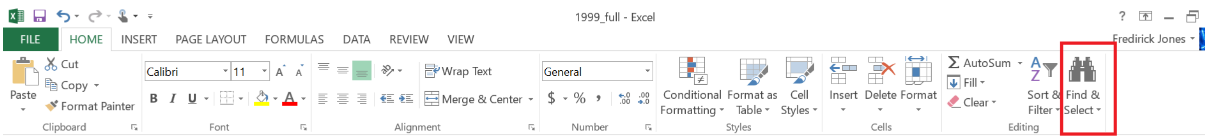
Step 2. Hiding the Blanks

Not all shrimp mounts are being given a label.
Select the column for the IGSN handle (in this case BA but it does change)

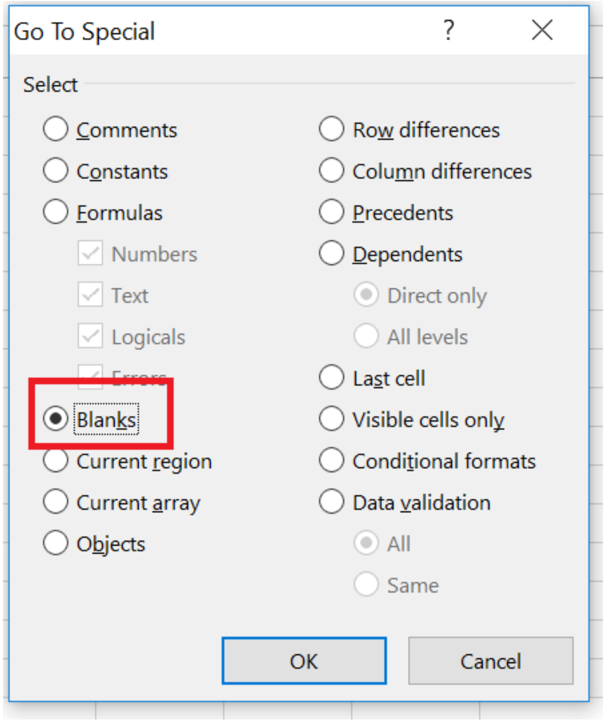


	B	BA	BC	BD	BE
1	SHRIMP M	PARENT_IGSN_HANDLE			
2	99-01				
3	99-01	http://hdl.handle.net/10273/XXAB000GM			
4	99-02				
5	99-03				
6	99-03				
7	99-03				
8	99-04				
9	99-04	http://hdl.handle.net/10273/XXAB000GO			
10	99-04				
11	99-05	http://hdl.handle.net/10273/XXAB000GQ			
12	99-05	http://hdl.handle.net/10273/XXAB000GQ			
13	99-05	http://hdl.handle.net/10273/XXAB000GQ			

With that column selected under click **Find & Select** under the **Home** Tab



This brings up quick filters select **Blanks**



IMPORTANT STEP

If you just hide the columns now it will shuffle the data and the URL will be aligned with the incorrect mount ID.

With the blank cells select press SHIFT + SPACE to select the entire rows of selected cells.

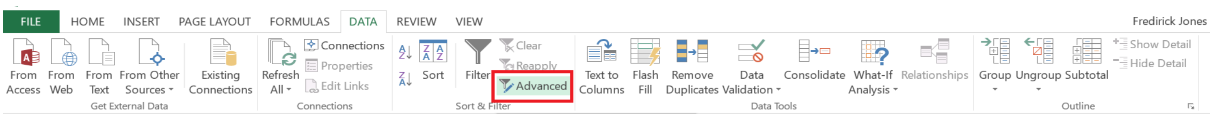
Press Crtl + 9 (shortcut for hide). You should now have something like below.

B	BA	BC	BD	BE
SHRIMP M	PARENT_IGSN_HANDLE			
99-01	http://hdl.handle.net/10273/XXAB000GM			
99-04	http://hdl.handle.net/10273/XXAB000GO			
99-05	http://hdl.handle.net/10273/XXAB000GQ			
99-05	http://hdl.handle.net/10273/XXAB000GQ			
99-05	http://hdl.handle.net/10273/XXAB000GQ			
99-05	http://hdl.handle.net/10273/XXAB000GQ			
99-06	http://hdl.handle.net/10273/XXAB000GV			
99-06	http://hdl.handle.net/10273/XXAB000GV			
99-06	http://hdl.handle.net/10273/XXAB000GV			

Step 3. Removing Duplicates

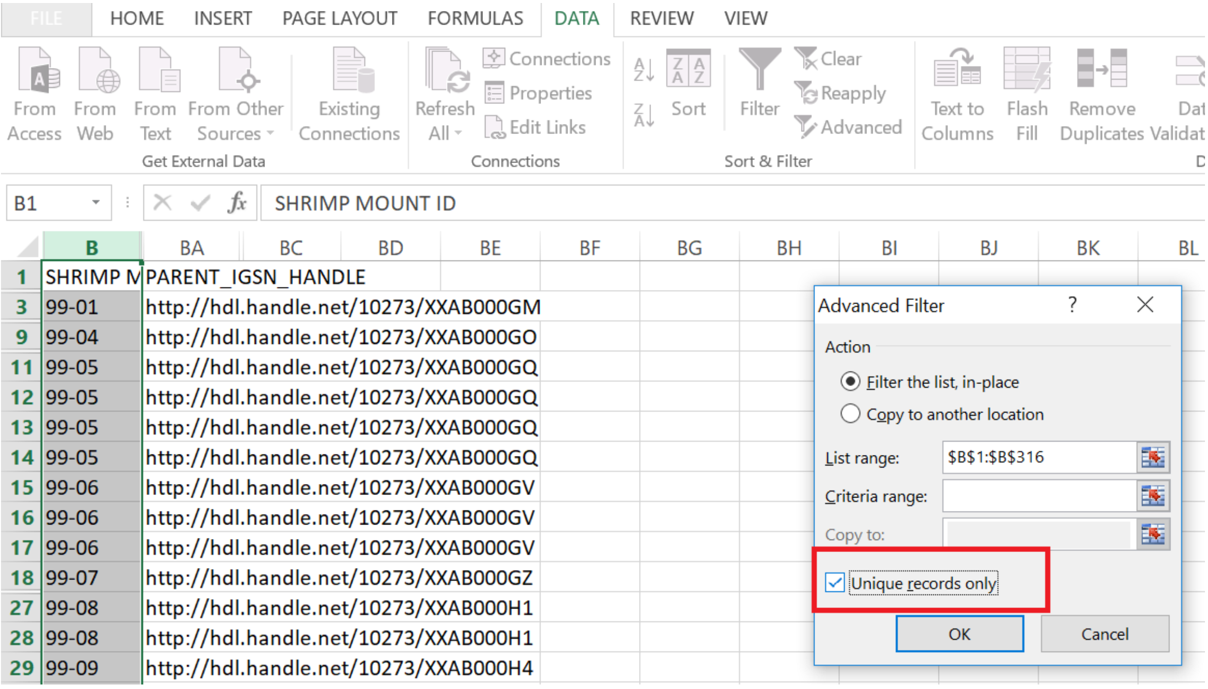
Select the **Parent_IGSN_Handle** column

Then under the **DATA** tab click the **Filter** button and then **advanced filter.**



Select Unique Records Only and the OKAY

In the below screenshot I have selected shrimp mount column, this is incorrect have PARENT _IGSN_HANDLE Column selected to filter as mentioned.



This will filter out all duplicate values for the IGSN Handles.

There are two outcomes, And it seems random

Either it will be filtered correctly with just the data we need. Like below

1	SHRIMP M	PARENT_IGSN_HANDLE
2		
3	99-01	http://hdl.handle.net/10273/XXAB000GM
9	99-04	http://hdl.handle.net/10273/XXAB000GO
11	99-05	http://hdl.handle.net/10273/XXAB000GQ
15	99-06	http://hdl.handle.net/10273/XXAB000GV
18	99-07	http://hdl.handle.net/10273/XXAB000GZ
27	99-08	http://hdl.handle.net/10273/XXAB000H1
29	99-09	http://hdl.handle.net/10273/XXAB000H4
34	99-11	http://hdl.handle.net/10273/XXAB000H7
42	99-12	http://hdl.handle.net/10273/XXAB000HF

If so you can skip the next filter step, if all your blank columns come back **follow step 3B,** I cannot figure out what causes the different output.

Step 3B.

Select Parent _IGSN_Handle Column

Go to the find & Select under home tab again.

Select Blanks.

SHIFT + SPACE to highlight entire rows

CRTL + 9 to hide these rows.

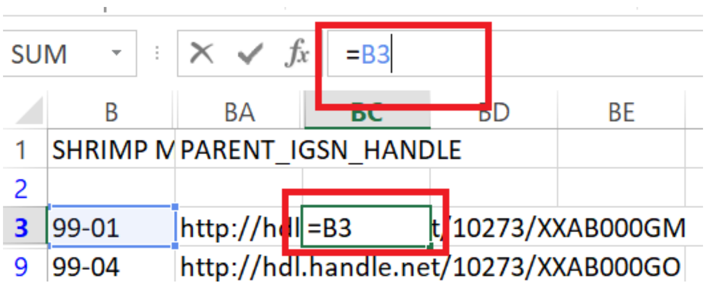
Step 4. Arranging Data for QR creation.

The website used to create these QR codes needs the data in the order of URL,File name.

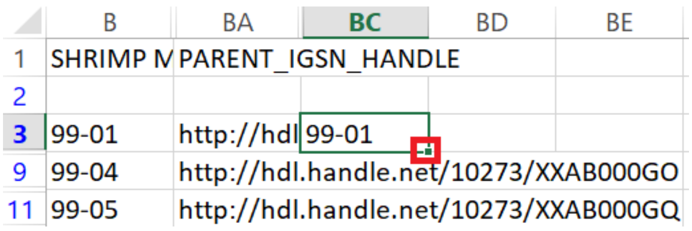
Note for later the comma is 100 percent necessary missing this comma means creating incorrect QR code. (this will make sense later)

The Shrimp Mount ID column to the right of the Parent_IGSN_handle column.

There are multiple ways to do this. How I do it is entering the equation to make the column next to Parent_IGSN_handle to equal the Shrimp mount Id column. Like below



Then apply this rule to the entire column. By dragging the small green square at the bottom right of the selected cell and highlight all other cells in the column.



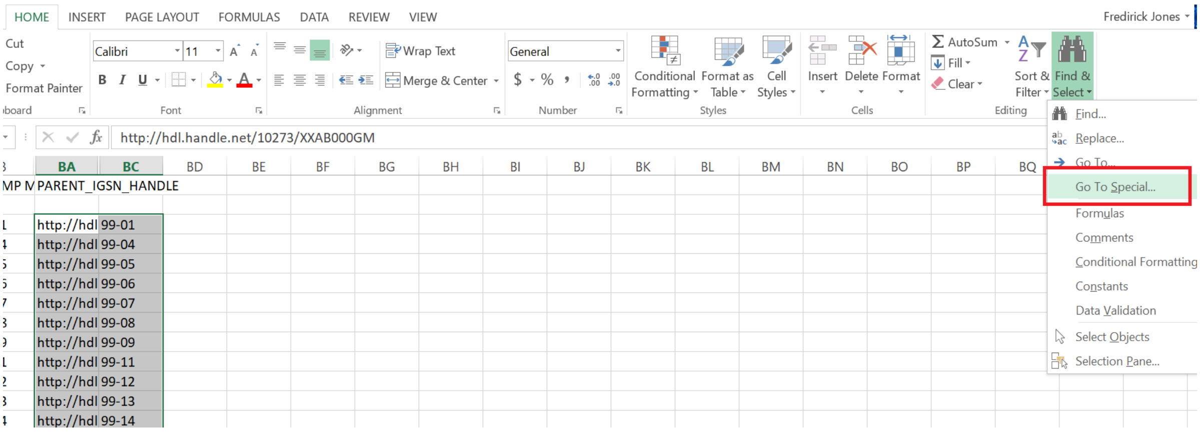
3	99-01	http://hdl	99-01
9	99-04	http://hdl.handle.net/10273/XXAB000GO	
11	99-05	http://hdl.handle.net/10273/XXAB000GQ	
15	99-06	http://hdl.handle.net/10273/XXAB000GV	
18	99-07	http://hdl.handle.net/10273/XXAB000GZ	
27	99-08	http://hdl.handle.net/10273/XXAB000H1	
29	99-09	http://hdl.handle.net/10273/XXAB000H4	
34	99-11	http://hdl.handle.net/10273/XXAB000H7	
42	99-12	http://hdl.handle.net/10273/XXAB000HF	
52	99-13	http://hdl.handle.net/10273/XXAB000HI	
55	99-14	http://hdl.handle.net/10273/XXAB000HK	
57	99-15	http://hdl.handle.net/10273/XXAB000HM	
79	99-20	http://hdl.handle.net/10273/XXAB000HR	
83	99-21	http://hdl.handle.net/10273/XXAB000HW	
88	99-22	http://hdl.handle.net/10273/XXAB000I1	
99	99-24	http://hdl.handle.net/10273/XXAB000I4	
108	99-26	http://hdl.handle.net/10273/XXAB000IA	
117	99-28	http://hdl.handle.net/10273/XXAB000IF	
120	99-29	http://hdl.handle.net/10273/XXAB000IJ	
131	99-32	http://hdl.handle.net/10273/XXAB000IL	
133	99-33	http://hdl.handle.net/10273/XXAB000IO	

	B	BA	BC	BD
1	SHRIMP N	PARENT_IGSN_HANDLE		
2				
3	99-01	http://hdl	99-01	
9	99-04	http://hdl	99-04	
11	99-05	http://hdl	99-05	
15	99-06	http://hdl	99-06	
18	99-07	http://hdl	99-07	
27	99-08	http://hdl	99-08	
29	99-09	http://hdl	99-09	
34	99-11	http://hdl	99-11	
42	99-12	http://hdl	99-12	
52	99-13	http://hdl	99-13	
55	99-14	http://hdl	99-14	
57	99-15	http://hdl	99-15	
79	99-20	http://hdl	99-20	
83	99-21	http://hdl	99-21	
88	99-22	http://hdl	99-22	
99	99-24	http://hdl	99-24	
108	99-26	http://hdl	99-26	

Step 5. Copying the correct data

You need to copy the Parent_IGSN_handle and the new column we made for shrimp ID (in this case Column BA and BC without the headings.) But if we simply copy we will select all out hidden cells as well.

Highlight the data you need and go back to Find & Select and **Go to special**



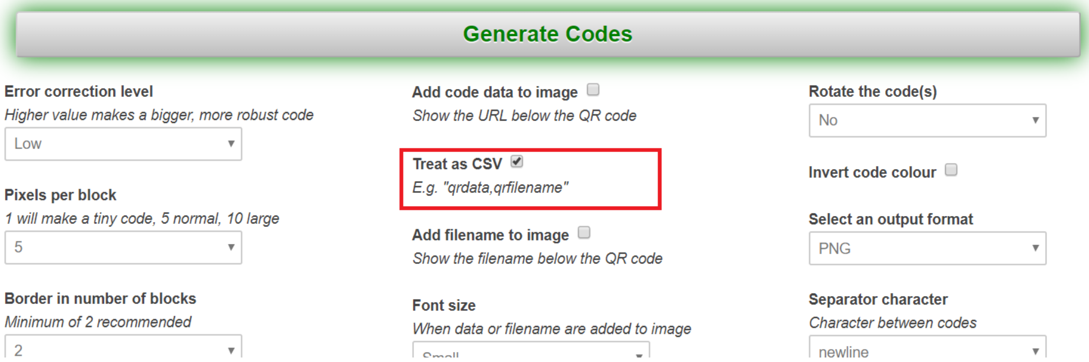
Now we can select the option **Visible Cells Only**

And finally we have selected the correct data which can be copied with CRTL + C

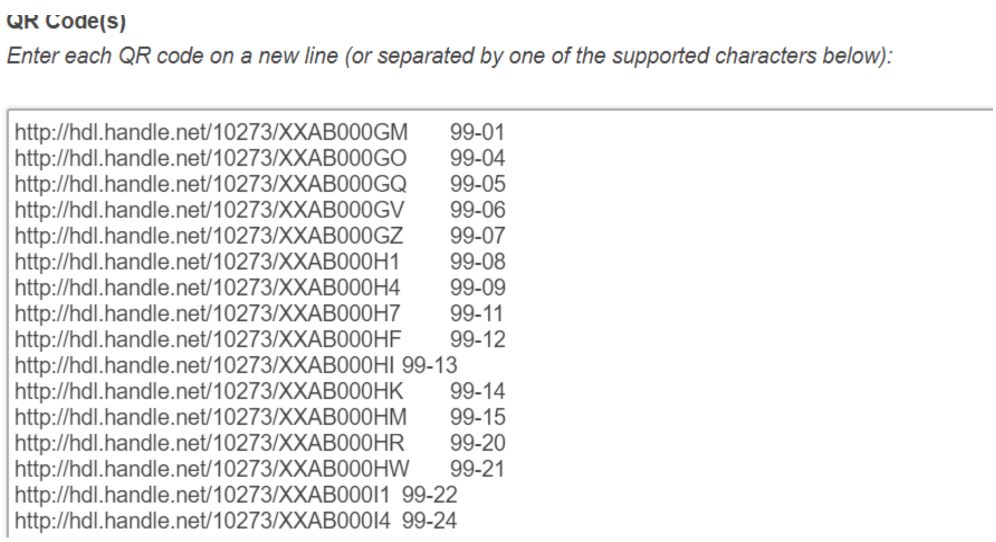
Step 6. Creating QR Codes

The website to use is https://qrexplore.com/generate/

The only option to change is to **treat as a CSV** make sure it's selected



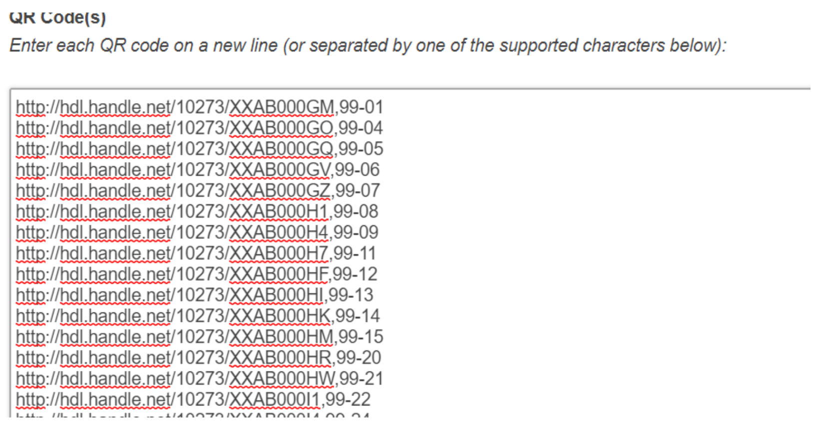
Paste the data into the window



Some will have a large gap and some will have a small gap this is just random

All the lines have to be the URL,Name (don't forget that comma).

Be sure not to delete or change any other data



Press generate codes and then download as Zip.

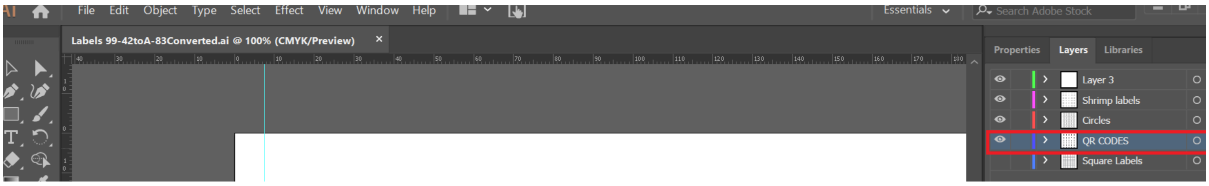
Extract the Zip file and name it the year. In the R drive there is a file named bulk QR codes.

In the file there is an excel sheet named generated codes this lists the data encoded in the QR code and the QR code name, I would strongly recommend opening this spreadsheet and making sure everything is correct. If you have a misnamed code it probably means comma was missed or two links where on one line or some other incorrect data, you can simply regenerate effected data.

3. Creating Complete Labels/Printing labels.

Open the relevant sheet in Adobe Illustrator, each label should already have a shrimp mount label from previous steps.

When placing the QR codes make sure you are working in the QR CODES layer, this will make it much easier to fix an issue or do a more detailed edit if needed. Click on it to select the layer you want to place in.



Then use the PLACE command (CRTL + SHIFT + P) to bring up the file explorer to select the QR code to place. Don't use CRTL + C and CRTL + V in Illustrator as a linked image which is not good.

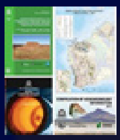
Once you have selected the correct QR code click to place it in the template, it will be too big which is fine for now.

RECORD 2023/17

THE PRESERVATION OF LEGACY COLLECTIONS PROJECT: A TEMPLATE FOR PRESERVING HIGH-VALUE COLLECTIONS FOR FUTURE RESEARCH

ER Blereau and A Bellenger

Access GSWA products



All products
All GSWA products are free to download as PDFs from the DMIRS eBookshop <www.dmirs.wa.gov.au/ebookshop>. View other geoscience information on our website <www.dmirs.wa.gov.au/gswa>.

Hard copies
Limited products are available to purchase as hard copies from the First Floor Counter at Mineral House or via the DMIRS eBookshop <www.dmirs.wa.gov.au/ebookshop>.

Fieldnotes
Fieldnotes is a free digital-only quarterly newsletter which provides regular updates to the State's exploration industry and geoscientists about GSWA's latest programs, products and services.
Access by subscribing to the GSWA eNewsletter <www.dmirs.wa.gov.au/gswaenewsletter> or downloading the free PDF from the DMIRS eBookshop <www.dmirs.wa.gov.au/ebookshop>.

GSWA eNewsletter
The GSWA eNewsletter is an online newsletter that contains information on workshops, field trips, training and other events.
To keep informed, please subscribe <www.dmirs.wa.gov.au/gswaenewsletter>.

Further details of geoscience products are available from:

First Floor Counter
Department of Mines, Industry Regulation and Safety
100 Plain Street
EAST PERTH WESTERN AUSTRALIA 6004
Phone: +61 8 9222 3459 Email: publications@dmirs.wa.gov.au
www.dmirs.wa.gov.au/GSWApublications