AUSTRALIAN ARCHAEOLOGY
2023, VOL. 89, NO. 1, 1–13
https://doi.org/10.1080/03122417.2022.2159751

ARTICLE

OPEN ACCESS

Check for updates

SahulArch: A geochronological database for the archaeology of Sahul

Wanchese M. Saktura[a], Emma Rehn[b], Lauren Linnenlucke[b], Henry Munack[a], Rachel Wood[c,d], Fiona Petchey[b,e], Alexandru T. Codilean[a], Zenobia Jacobs[a], Tim J. Cohen[a], Alan N. Williams[f,g] and Sean Ulm[b]

[a]ARC Centre of Excellence for Australian Biodiversity and Heritage and School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, NSW, Australia; [b]ARC Centre of Excellence for Australian Biodiversity and Heritage and College of Arts, Society and Education, James Cook University, Cairns, QLD, Australia; [c]Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, School of Archaeology, University of Oxford, Oxford, United Kingdom; [d]Research School of Earth Science, Australian National University, Canberra, ACT, Australia; [e]Radiocarbon Dating Laboratory, Te Aka Mātuatua – School of Science, University of Waikato, Hamilton, New Zealand; [f]ARC Centre of Excellence for Australian Biodiversity and Heritage and Climate Change Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, Australia; [g]EMM Consulting Pty Ltd, St Leonards, NSW, Australia

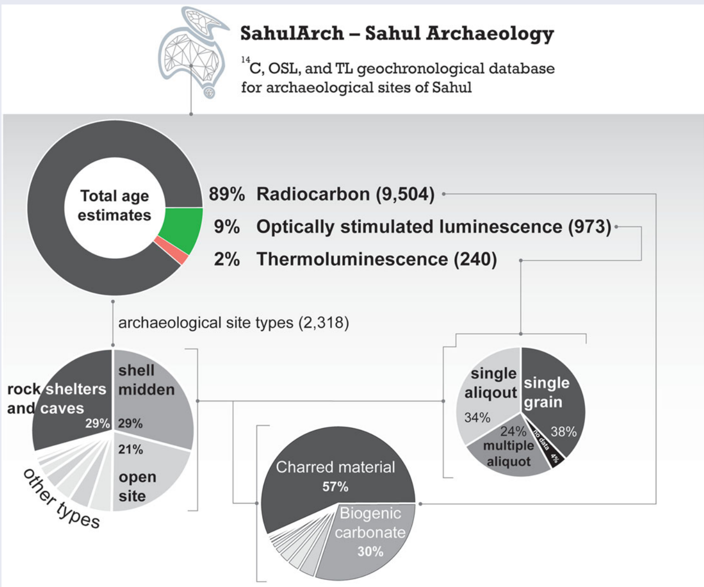
ABSTRACT

Reliable chronological frameworks for archaeological sites are essential for accurate interpretations of the past. Geochronology represents the core of interdisciplinary research because it allows integration of diverse data on a common timeline. Since the radiocarbon revolution in Australian archaeology in the 1950s, thousands of ages have been produced across Sahul (combined landmass of Australia and New Guinea). Methods such as thermoluminescence (TL) and optically stimulated luminescence (OSL) have also been used on Australian archaeological deposits and enabled the study of the deep past beyond the limits of radiocarbon dating. After seven decades, these geochronological methods no longer provide just a 'date', but instead, the geochronological community is focussed on providing the most reliable, precise, and reproducible ages. These aspects of age estimation are central to the framework of the SahulArch geochronological database. SahulArch is a new publicly available continental-scale dataset in which context and quality assurance criteria of each dated sample are considered as important as the age itself. SahulArch contains a total of 10,717 ages (9,504 radiocarbon, 973 OSL, and 240 TL) from 2,318 sites across the Sahul landmass. We describe the structure of SahulArch, types of auxiliary data collected, and provide a summary of the data in SahulArch.

ARTICLE HISTORY
Received 29 September 2022
Accepted 14 December 2022

KEYWORDS
Sahul; online database; radiocarbon dating; optically stimulated luminescence dating; thermoluminescence dating; geochronology; Australian archaeology

GRAPHICAL ABSTRACT



CONTACT Wanchese M. Saktura wanchese@uow.edu.au, saktura.wm@gmail.com ARC Centre of Excellence for Australian Biodiversity and Heritage and School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, NSW, Australia

Introduction

The ability to estimate the elapsed time since an event occurred is essential to any inquiry into the past. Geochronological methods such as radiocarbon (^{14}C), optically stimulated luminescence (OSL) and thermoluminescence (TL) dating can quantify the passage of time during the Holocene and Pleistocene and are important tools in the study of archaeological deposits. Since the first application of each of these methods to Australian archaeology, starting during the 1950s (radiocarbon: Gill 2010; TL: Huxtable and Aitken 1977; OSL: Roberts et al. 1993), thousands of ages have been obtained from archaeological sites across Sahul (Figure 1; e.g. Thomas et al. 2018; Williams et al. 2014). Age can be seminal for our understanding of the past or can be completely unusable, depending on its context, sample quality, and analytical procedures. Therefore, documenting and sharing the contextual information of a sample and analytical procedures used to produce an age estimate is as important as publishing the age itself.

Sahul is the exposed continental landmass that encompassed mainland Australia, New Guinea, Tasmania, and interspersed islands during Pleistocene low sea level stands (Figure 1). The initial peopling, subsequent dispersal and occupation of Sahul, as well as understanding changes in societies and cultures over time, are key research themes in Australian archaeology (e.g. Bradshaw et al. 2021; Clarkson et al. 2017; Crabtree et al. 2021; Williams et al. 2015). All these themes are time-based and require chronological datasets with comprehensive contextual information to provide reliable insights. The SahulArch database was designed and compiled to provide a repository of all known archaeological uncalibrated radiocarbon ages, and OSL and TL ages, for use by the wider community, within and beyond academia.

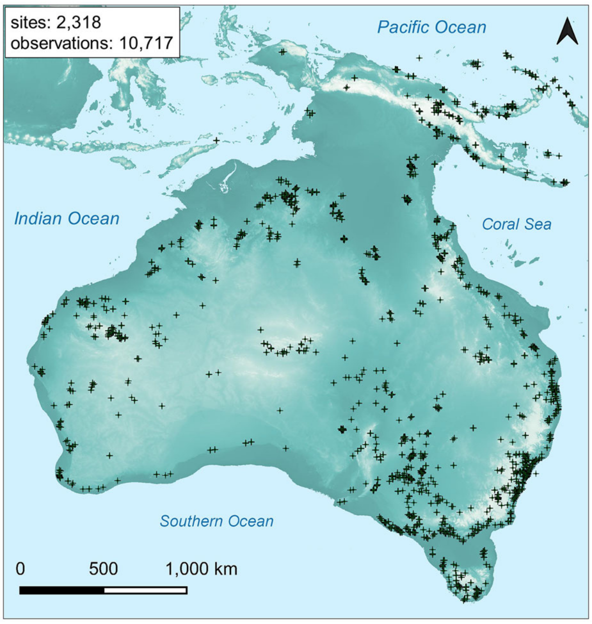


Figure 1. The Sahul landmass showing an approximation of the coastline during the Pleistocene assuming sea level 125 m lower than at present. The crosses mark SahulArch sites ($n = 2{,}318$) for which locations are obfuscated by a randomising algorithm. Digital elevation model: GEBCO Compilation Group (2020).

SahulArch builds on several earlier compilations from archaeological sites on national (Williams et al. 2014) and regional scales; for Victoria by Godfrey et al. (1996) and Thomas et al. (2018); the Top End by Brockwell et al. (2009) and Williams and Smith (2012); Queensland by Ulm and Reid (2000) and Linnenlucke (2022); arid zones by Williams et al. (2008); and southern Australia and Tasmania by Williams and Smith (2013).

The scope of SahulArch expands the geographical bounds of previous data compilations and is focused on the capture of auxiliary and quality assurance data that can be used for the evaluation of each age. The structure of SahulArch was designed with input from expert practitioners of each dating method to ensure that all essential quality assurance data are included in conformity with best-practice data reporting standards (e.g. Millard 2014). We acknowledge that these are subject to change as dating methods evolve, and so the database was designed with an update option. Attention was also given to the compilation of contextual information for each sample and exhaustive information regarding its site of origin. Therefore, SahulArch enables the capture of large amounts of additional data (where available) for each age determination, produced as part of dating procedures, as well as extensive qualitative information about the site, sample context, and material used for dating. Such information is rarely captured in most compilations, which are typically focused on age estimates alone. This has meant that every time a new regional study has been undertaken, investigators have been required to revisit all previous works to carry out even the most basic of chronometric hygiene protocols. This process is simplified by SahulArch through extensive data inclusion in the database, which preserves essential information from each study and provides users with the ability to quickly access and evaluate the accuracy, reliability, and reproducibility of age estimates. The purpose of SahulArch is to empower users to scrutinise auxiliary data for each sample and assess it before its age is used. It also provides a guide to what auxiliary information should be published alongside new age estimates according to current data reporting standards.

The aim of this paper is to introduce SahulArch to those interested in Australasian archaeology and to provide a summary of the available data within its collections, an overview of its features, how it can be accessed and used, and to outline future opportunities and benefits to Indigenous and archaeological communities. An overview of OCTOPUS v.2, the database that hosts SahulArch, is also provided but more extensive information is available in Codilean et al. (2022b).

The OCTOPUS v.2 database

SahulArch is hosted by the OCTOPUS v.2 database (Codilean et al. 2022b) and can be accessed via a bespoke web-browser interface at: https://octopus-data.org. User guides are also provided (Rehn 2022; Rehn et al. 2022) and showcase how to use and access data through various means. OCTOPUS v.2 contains three core data collections including global and Australian cosmogenic radionuclide ages and denudation rates (CRN), luminescence ages from sedimentary archives in Sahul (SahulSED), and luminescence and radiocarbon ages from archaeology in Sahul (SahulArch; Figure 2). OCTOPUS is an Open Geospatial Consortium (OGC) compliant and web-enabled database that allows users to visualise, query, and download data stored in its collections (Figure 2). OCTOPUS v.2 is a relational database which differs from typical flat table databases commonly compiled by Earth science research communities. Flat table databases store all attributes in a single pre-structured column and row matrix. In the relational database structure used for SahulArch, data are organised into multiple flat tables where relations between rows or tables are established with unique key identifiers. This data structure enables efficient data storage and maintenance, querying of the database, and flexibility to create new relationships between data tables without violating the existing database structure. The relational database is organised into multiple themes which encompass information regarding sample type, its ages, context, site location information, and bibliographical information. These can serve a particular collection or a theme (e.g. common dating method, OSL) or can be general and applicable to all collections in the database (e.g. bibliographic information; Figure 2). In the context of this data model, SahulArch has four hierarchical levels:

1. *Metasite* – a broad spatial agglomeration of *sites* which share common properties.
2. *Site* – represented by a geographic point from which *samples* were collected.
3. *Sample* – a tangible material collected from a *site* to be processed and analysed to produce an age estimate.
4. *Observation* – an age estimate obtained from a *sample* (note that multiple observations can be generated from a single *sample* in this data model).

The spatial component is an integral part of the OCTOPUS database (Codilean et al. 2018, 2022b). However, in an archaeological context, spatial information is often culturally sensitive or can place cultural sites at risk of adverse visitation or vandalism. Therefore, site locations in SahulArch are obfuscated within a 25 km radius using a randomising algorithm and are stored as polygons (near circular pentadecagon) that only represent a possible site location within ~1,275 km^2 area. The non-obfuscated site coordinates are still stored in relational attribute tables but are not made public.

SahulArch structure and design

SahulArch is organised into three collections based on dating method: SahulArch OSL (Saktura et al. 2022a), SahulArch TL (Saktura et al. 2022b), and SahulArch Radiocarbon (Saktura et al. 2022c). Each collection is issued with a digital object identifier (DOI) that corresponds to a time-stamped version of the collection (Table 1). Subsequent collection updates will be minted with new DOIs, hence users are urged to specify the version of DOI when using SahulArch in their work.

The structure of SahulArch was designed so it can be easily queried. This was achieved by limiting the number of free text data fields to reduce interference from nomenclatural preferences of the data collectors, alternative spellings, and typographical errors while querying the database.

Observations

Observations (i.e. age records) are the most basic SahulArch entities. Each observation is composed of several components that are either unique or common among related observations from the same archaeological site (Figure 2). A summary of the types of information collected for radiocarbon and luminescence dating methods in SahulArch is shown in Figure 2. Data fields are described in the OCTOPUS data index which can be accessed from supplemental online material (SOM) and from Codilean et al. (2022a), see Table 1. This document tabulates the type of data collected, units and values format, as well as an example and description of each field.

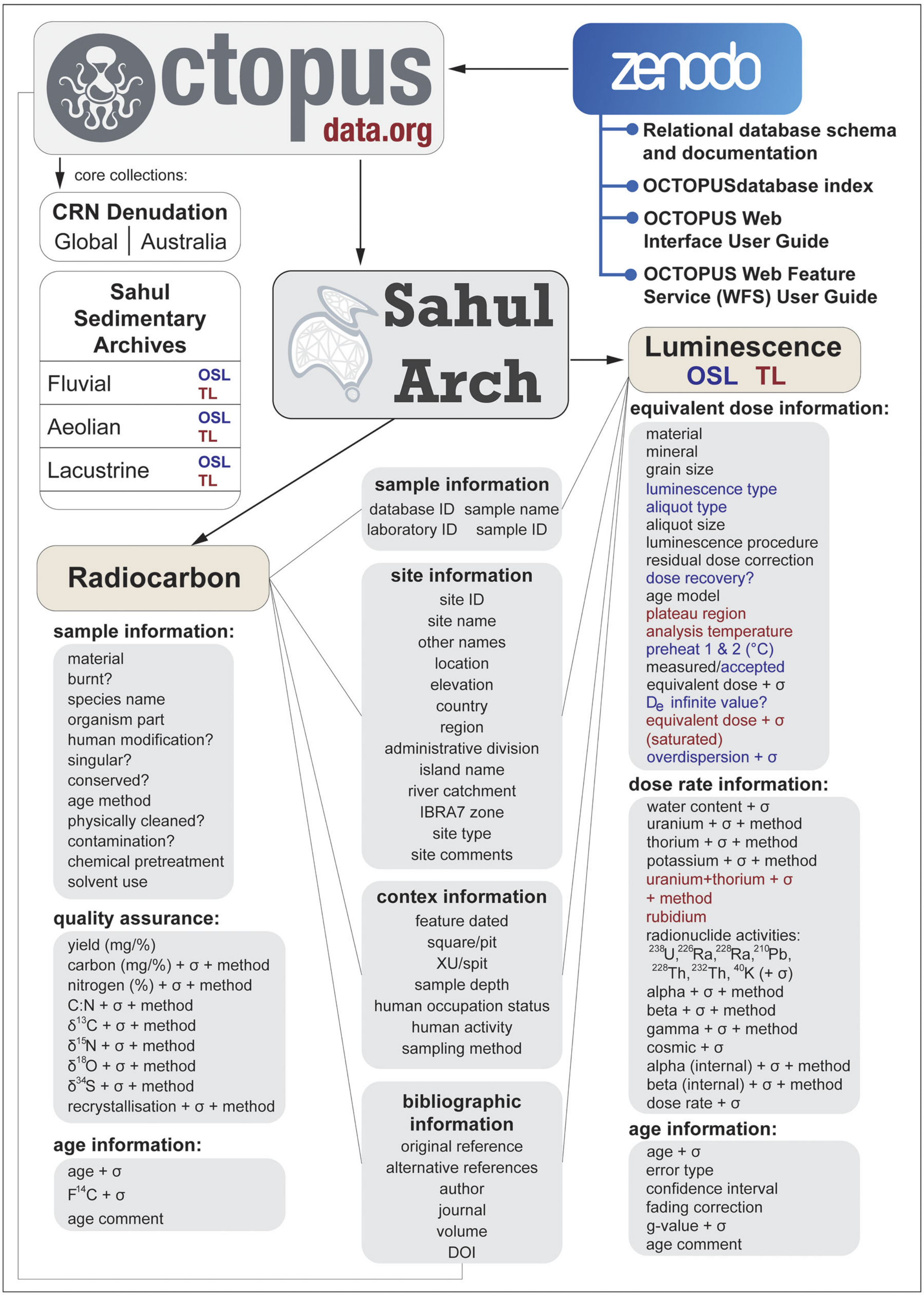


Figure 2. Summary of OCTOPUS v.2 core collections (SahulArch, SahulSed, and CRN) and auxiliary information hosted in the Zenodo online library (see Table 1 for links to these resources). A summary of data compiled for radiocarbon, OSL, and TL collections in SahulArch is also presented. Symbology for luminescence collections: blue = OSL only, red = TL only, and black = both OSL and TL; σ indicates error value is also compiled; 'method' indicates that method information is also compiled.

In SahulArch, samples collected for radiocarbon, OSL, and TL dating from the same site or layer, but published years apart, still share the same site information as well as similar bibliographical and contextual information (Figure 2). Each archaeological site is assigned a 'SITEID' (e.g. ARCH0001 = Madjedbebe). This identifier (ID) links samples (i.e. physical entities) and their observations (i.e. ages derived from samples) to their site of origin and is used to create unique IDs in the database. The first part of the

Table 1. SahulArch core collections and auxiliary database documents.

Database item	DOI	No. of observations	No. of data fields	% complete[a]	Description
Core SahulArch collections:					
SahulArch Radiocarbon	https://doi.org/10.25900/gpvr-ay04.	9504	120	~98%	Collection of radiocarbon ages
SahulArch OSL	https://doi.org/10.25900/9y07-4j77.	973	143	~90%	Collection of OSL ages
SahulArch TL	https://doi.org/10.25900/af67-kh16.	240	141	~94%	Collection of TL ages
Auxiliary items:		**Item description:**			
OCTOPUS database index[b]	https://doi.org/10.5281/zenodo.7352827.	Description of data fields in OCTOPUS collections			
OCTOPUS Web Interface User Guide	https://doi.org/10.5281/zenodo.6469039.	Description of features of the OCTOPUS web interface, and instructions on how to use it			
OCTOPUS Web Feature Service (WFS) User Guide	https://doi.org/10.5281/zenodo.7063533.	Description of how to access OCTOPUS collection through WFS using QGIS and R open-source software			
OCTOPUS Database v.2: Relational database schema and documentation	https://doi.org/10.5281/zenodo.7352807.	Description of the OCTOPUS v.2 relational database structure and operations			
OCTOPUS Database v.2	https://doi.org/10.5194/essd-14-3695-2022.	Original research article presenting OCTOPUS v.2			

Notes:
[a]Based on age observations captured up to June 2022 (cut-off date), known resources captured by previous data compilations, our extensive online and grey literature searches, and data published since our data capture cut-off date.
[b]Also available as supplemental online material.

observation identifier ('OBSID1') consists of the 'SITEID' plus the number of a subsequent observation for a given method. For example, the first radiocarbon age observation from Madjedbebe, the 'OBSID1' will be ARCH0001 + C14001 = ARCH0001C14001. Each method is tallied separately (i.e. OSL001, TL001), and therefore each observation has a unique 'OBSID1'. An important sample/observation discriminant field is the sample identifier code ('SMPID'). The 'SMPID' is a common tally for all observations regardless of method, but it is non-unique and can be duplicated for the same sample where replicate analyses are performed, for example, the same original material is analysed using a different material fraction, analysed with an altered protocol or a different method. For example, in Bird et al. (1999) the charcoal sample 'MM25-XU8(3)' from Ngarrabullgan Cave ('SMPID' = ARCH0016SMP042) was dated using the radiocarbon method 14 times, each with altered chemical pre-treatments and combustion temperatures. This produced 14 different observations from this single sample. All these observations are captured in SahulArch because they are important to methodological studies and research legacy. These observations can be easily identified in the collection with the use of the 'SMPID' code, which is the same for all 14 observations.

Data fields

The number of data fields that constitute an observation in SahulArch vary between each method specific collection (Table 1). The format of data fields includes predefined option selection (lists are presented in the database index; Table 1 and SOM), short text, numerical value, and free text fields for general commentary about an observation or a site. Fields such as 'SITECODE' (type of archaeological site) or 'MATERIAL1'/'MATERIAL2' (type of material used for dating), for example, have predefined options based on expert and user advice to ensure the most commonly occurring options are captured. Less frequently occurring options or subcategories of already existing options have not been included to avoid overcategorisation of data. Hence, some specific/niche terms and those that do not fit available options have been categorised as 'Other'. Free text commentary fields such as 'SITE_COMMT', 'SMP_COMMT', and 'AGE_COMMT' are then used to provide further information for site, sample, and age if necessary.

Data compilation

SahulArch data compilation was undertaken on a site basis. The site information such as alternative

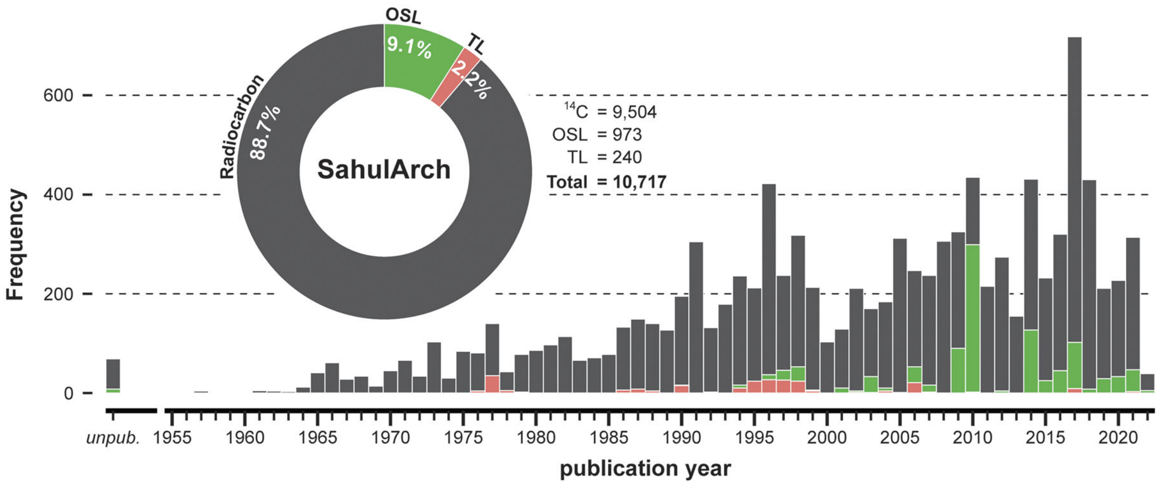


Figure 3. Summary of published age observations per year in SahulArch from 1955 to 2022, the earliest publication in the database to the data collection cut-off (June 2022).

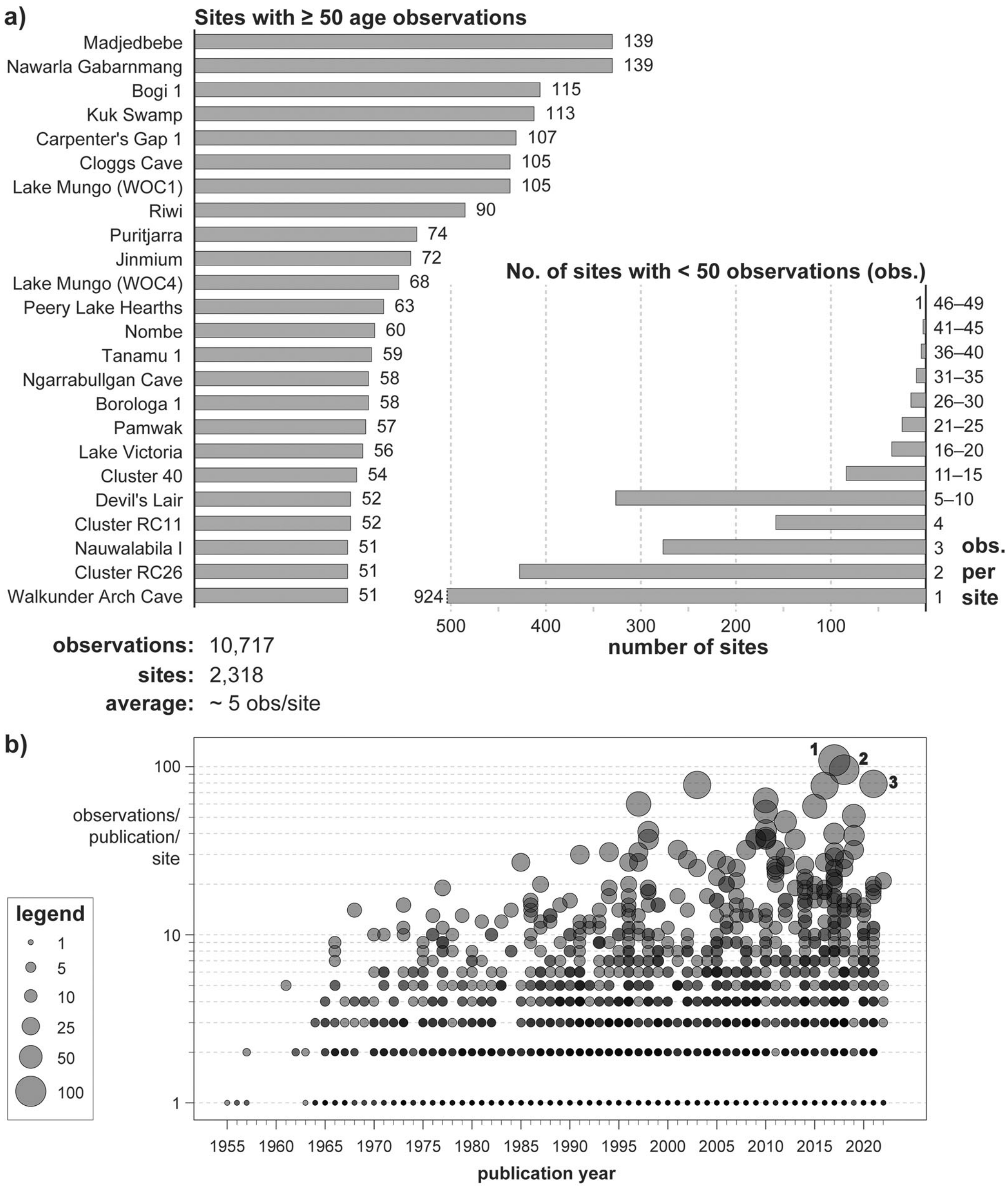


Figure 4. (a) Summary of sites with most age observations (left) and those with fewer than 50 age observations (right); (b) Plot showing number of observations published per study per site (log scale) at yearly interval. The top three largest published datasets: 1 – Clarkson et al. (2017): Madjedbebe; 2 – Maloney et al. (2018): Carpenter's Gap 1; 3 – David et al. (2021): Cloggs Cave.

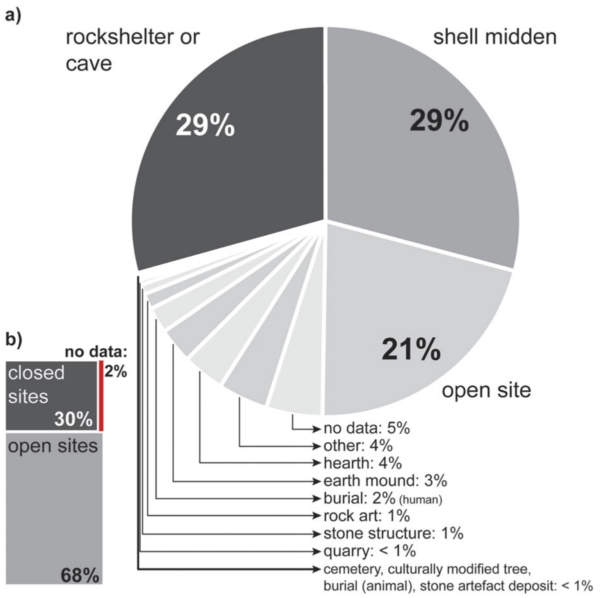


Figure 5. Summary of site type frequency in SahulArch ($n = 2{,}318$). a) All site categories in SahulArch and b) site sub-classification, identified either as 'open' or 'closed'.

names, location or general commentary has been amalgamated from several sources which are listed as alternative references. The AustArch database (Williams et al. 2014) was used to help guide the search for archaeological sites and references in Australia. Obtaining the original information source was always attempted, and was successful for most journal articles, books, and conference proceedings. Most data presented in SahulArch were, therefore, obtained from the original sources. The exceptions were sources listed as personal communications in AustArch, archived reports, reports to government agencies and others like these. These were more problematic to obtain, and data collection in many cases relied on secondary sources. Site location was obtained either from the original source (if available) or interpolated from published maps using Google Earth Pro v. 7.3.4.8248 with varying accuracy depending on map resolution. These locations have been obfuscated in collection output (see above). Data extraction from the literature was done manually, and records were tabulated using the E4 software (McPherron and Dibble 2009) or Microsoft Excel with layout conforming to OCTOPUS indices (Codilean et al. 2022a). The collected data were then processed following the OCTOPUS v.2 relational data model (Codilean et al. 2022b; Munack and Codilean 2022).

Data summary

SahulArch currently contains 10,717 observations across 2,318 sites. It is on average approximately 94% complete (see Table 1) and encompasses a publication period between 1955 and 2022 (Figure 3). Radiocarbon ages are the predominant data type and account for 88.7% of all observations (Figure 3). The number of reported radiocarbon and luminescence ages grew steadily from the 1950s to the 1990s, reaching an average output of ~265 ages per year which has been maintained to the present day, although there is considerable variation (Figure 3). For example, the large number of age observations in 2017 (>700) is due to the synchronous publication of extensive datasets for Madjedbebe, Nawarla Gabarnmang, and the JSARN sites in the Top End, Wadjuru Rock Pool and several sites in the Pilbara, and shell mound complexes in Weipa (Figures 3 and 4). Most TL ages were produced prior to 2000, after which OSL was adopted as the primary method for dating sediment (Figure 3).

More than 900 sites have only one age observation (Figure 4(a)), 9% of all sites have more than 10 age observations, and only seven have more than 100. Figure 4(b) shows the number of age observations published per study per site for every year between 1955 and 2022. There is a general trend that indicates a progressive increase in dataset size and frequency of publishing of chronological data during the last seven decades, although there is little change to the number of studies with fewer numbers (≤5) of age observations (Figure 4(b)).

Sites in SahulArch are classified into 15 site categories (Figure 5). The most common site type categories include rockshelter or cave (29%), shell midden (29%), and open site (21%), whereas the other 12 categories account for the remaining ~21% (<1–4% each; Figure 5). Sites with a single age observation are predominantly shell middens (36%), open sites (21%), rockshelters or caves (19%) and isolated hearths (6%). Sites are additionally categorised either as open (68%) or closed (30%), which refers to their relationship with surrounding landscapes (Figure 5), rather than accessibility.

The radiocarbon collection is dominated by observations from charred material (57%; $n = 5{,}380$) and biogenic carbonate (30%; $n = 2{,}860$), whereas another nine material type categories account for the remaining ~13% of the collection (<1–4% each; Figure 6). Approximately 84% of charred material (Figure 6) was reported as charcoal in original publications but not further identified, hence these were recorded in SahulArch as 'charred', but for the material specific category ('MATERIAL2') 'no data' (ND; $n = 4{,}494$) had to be entered. The remaining 16% of charred materials are dominated by analyses of wood which account for 15% (Figure 6). The biogenic carbonate category is dominated by analyses of marine shell (78%; $n = 2{,}228$) and freshwater shell (13%; $n = 362$; Figure 6).

Shown in Figure 6 (bottom row) is the reporting frequency of contextual and data quality assurance

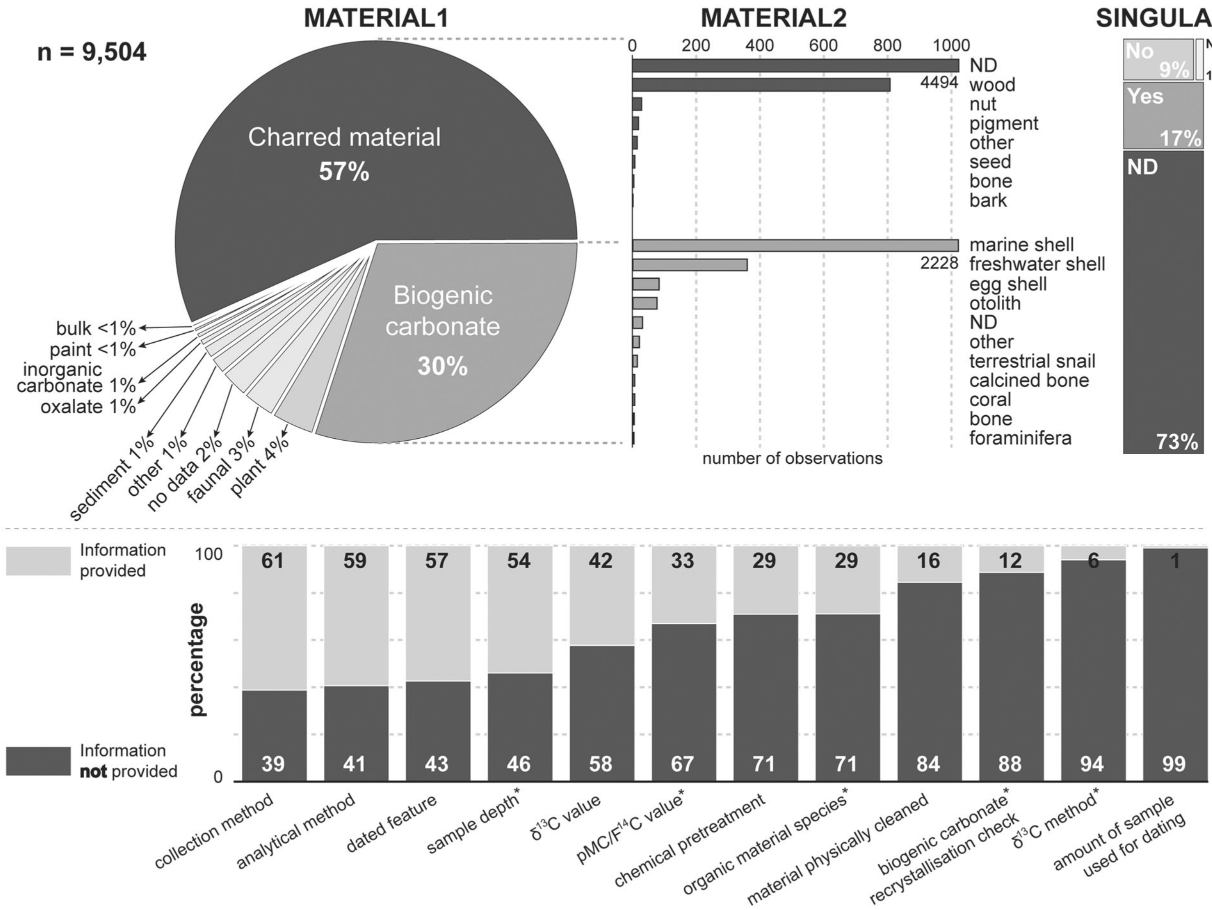


Figure 6. Summary of radiocarbon age observation data showing statistics of material type dated and information reporting frequency. '*' percentage calculation total ≠ 9,504 because lack of information resulted from inapplicability of the category to the sample rather than limited reporting. Totals for each '*' category were changed as follows: sample depth – excludes rock art and wasp nest features; organic material species – excludes non-organic samples where species name is not applicable; $\delta^{13}C$ method – only where $\delta^{13}C$ data were reported; biogenic carbonate recrystallisation check – only biogenic carbonate samples; pMC/$F^{14}C$ value – only samples analysed by AMS.

information. Some categories such as sample collection method, analytical method, type of feature that was dated, or sample depth are frequently reported. However, reporting of other sample and data quality assurance information is less common (Figure 6). In the 53 method-specific data categories in the radiocarbon collection, 80% of the fields contain 'no data'. This large value is partially explained by non-applicability of some categories to every material type (i.e. C:N, ^{15}N, ^{18}O or ^{34}S values are not typically analysed for charcoal samples but are undertaken during the analysis of bone). Also, during earlier decades, reporting standards were less stringent compared to today, and analytical protocols involved fewer additional quality assurance analyses. However, these considerations do not account for all of the 'no data' entries.

The OSL and TL collections, here referred to together as the luminescence collection, consist of data acquired from the analyses of single grains (38%), single aliquots (34%; multigrain), and multiple aliquots (24%; multigrain), where TL dating is performed only with use of multiple multigrain aliquots (Figure 7). The OSL and TL collections have 76 and 74 method specific categories respectively, which contain 52% and 67% 'no data' fields, which is lower than for the radiocarbon collection.

The key parameters that determine a luminescence age are the equivalent dose and environmental dose rate. Figure 7 shows data reporting statistics for some of the components necessary to reproduce a luminescence age. A grain size fraction used for equivalent dose estimation is reported for 81% of the observations, but aliquot size is only reported for 19%. Preheat temperatures (1 and 2) used are not reported for up to 37% of observations (Figure 7); this information is critical for reproducibility of equivalent dose of a sample. For the dose rate parameter, gamma and beta dose rate contributions tend to be non-uniformly reported (Figure 7). In Figure 7 (bottom row) listed in order are several data summaries that follow sample progression from collection method in the field to statistical age model usage for the final age estimation. These

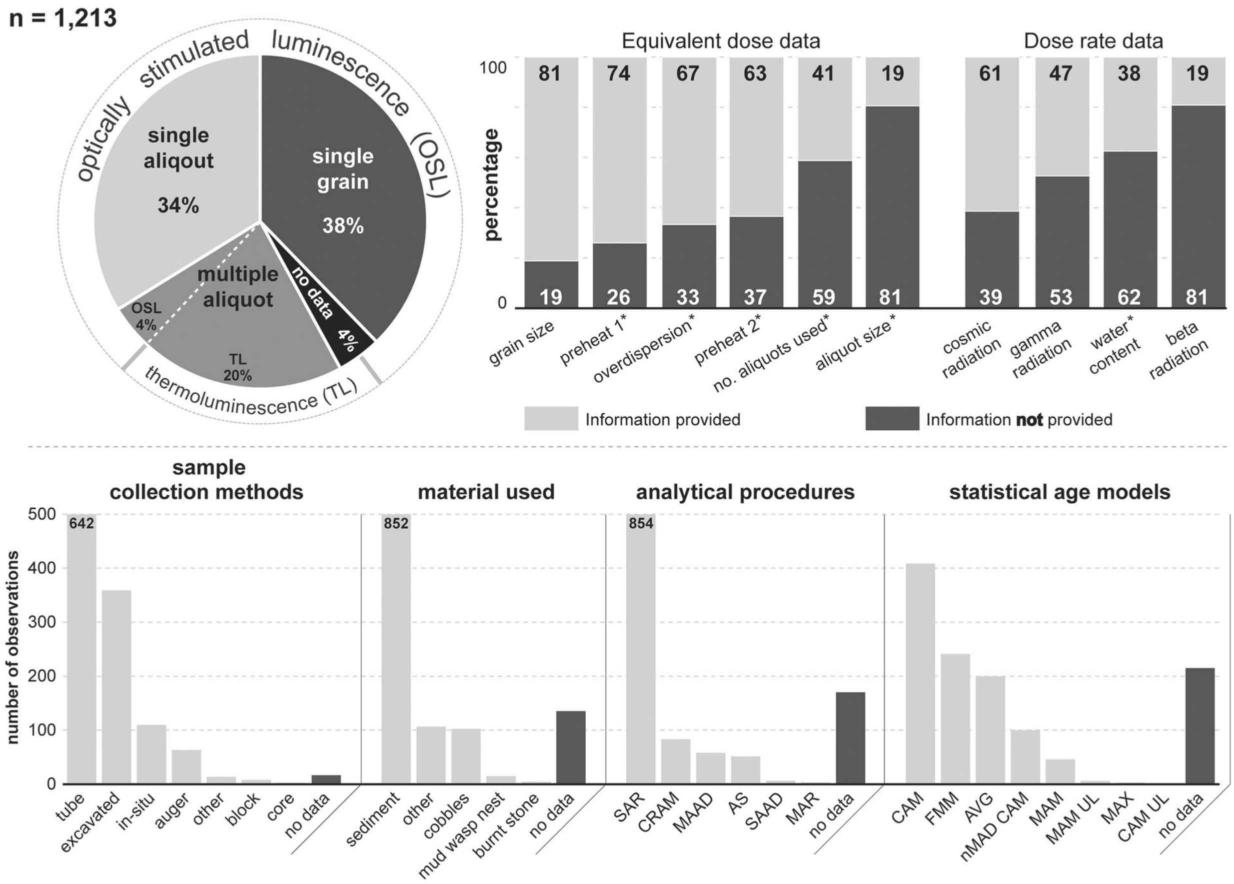


Figure 7. Summary of the luminescence data showing statistics of unit of measurement types and information reporting frequency. '*' percentage calculation total ≠ 1,213 because a condition applies: preheat 1, 2 and overdispersion – applies only to OSL; no. aliquots used – only when both measured and accepted aliquots/grains are reported; aliquot size – single (multi-grain) and multiple aliquot analyses; water content – used water content category (H2O_USED). Abbreviations for **analytical procedures**: SAR – single aliquot regenerative dose; CRAM – combined regenerative and additive method; MAAD – multiple aliquot additive dose; AS – Australian slide; SAAD – single aliquot additive dose; MAR – multiple aliquot regenerative dose; and **statistical age models**: CAM – central age model; FMM – finite mixture model; AVG – arithmetic mean; nMAD CAM – normalised median absolute deviation CAM; MAM – minimum age model; MAM UL – unlogged minimum age model; MAX – maximum age model; CAM UL – unlogged central age model.

show that sampling sediment for dating with a tube is the most common practice, as is usage of the single aliquot regenerative (SAR) protocol (Murray and Wintle 2000) and the central age model (CAM; Galbraith et al. 1999). Figure 7 also shows that the frequency of auxiliary information reported progressively decreases as a sample is progressed from the field, to laboratory, and into an age, as shown by the progressively increasing number of 'no data' fields (from left to right plot; Figure 7 bottom row).

Notes for users

The amount of auxiliary data collected for each observation in SahulArch is significantly larger than in previous archaeological compilations. This broad data collection was undertaken to maximise and diversify database usability; however, it comes at the cost of simplicity. Several mechanisms were implemented to prevent user experience issues, and these are explained below.

Chronological databases cannot be free of replicate (not duplicate) observations, because it is an integral part of the scientific process to re-analyse and verify the authenticity of a result, especially as analytical procedures improve over time. Selection and exclusion of replicate analyses can lead to bias and such actions can make a collection appear subjective or censored. SahulArch is free from selective manipulations and contains many replicate analyses or analyses with multiple results. It is up to the user to screen the data and choose the most appropriate set of observations for their analysis. The 'SMPID' codes and comment fields described above have been introduced (in addition to quality assurance categories) to help users with their data selection and evaluation. Careful evaluation of the data is particularly important for single-grain OSL dating where the equivalent dose datasets are examined with different statistical models (i.e. age models) which can either provide an average age from a dataset or identify multiple age populations and produce multiple ages from a single sample. For example, six OSL samples from Jinmium rockshelter (Roberts et al. 1999) have been analysed with different measurement procedures (multiple and single aliquot and single grain) and different statistical models (central, minimum, and average age) which produced four ages per sample. This is a result of a suspected complex depositional history at the site (Roberts et al. 1999). For such datasets, users are required to study the related observations and consult original publications to avoid misinterpretations of the database records.

Sites such as Madjedbebe have a long history of research, and samples taken three decades ago have been re-analysed multiple times as dating methods improved. For example, sample KTL162 was originally analysed using the TL method (Roberts et al. 1990). Later a sub-sample of the original sample was re-analysed using single aliquot and single grain OSL methods (Roberts et al. 1998). The sample was subsequently sub-sampled again for more single-grain measurements which superseded all previous estimations (Clarkson et al. 2017). These investigations produced seven unique observations for one sample (KTL162) across the SahulArch OSL and TL collections and these share the same 'SMPID' that can be used for easy distinction of replicate analyses. The 'AGE_COMMT' field for such sample sets states: 'This SG age for KTL162 superseded all other reported SG ages - dose rates values were updated (see Clarkson et al. 2017 for details)', and can be used to aid with identification of the most current estimate. However, users are still required to undertake their own careful examination of observations using original sources during their investigation to ensure the most recent and accurate results are used.

Age estimates in the SahulArch radiocarbon collection were compiled as conventional radiocarbon ages (CRA) as initially defined by Stuiver and Polach (1977) and subsequently updated (see Millard 2014 and the SahulArch radiocarbon database index [Codilean et al. 2022a and SOM]). Therefore, users of the radiocarbon collection will need to calibrate ages of interest prior to their use. Several original sources report radiocarbon ages after reservoir correction. In such instances, the correction was reversed during data compilation using the reported correction factor and noted in the 'AGE_COMMT' field.

Internal database operations

The relational database architecture of OCTOPUS v.2 enables efficient data storage and maintenance (Codilean et al. 2022b; Munack and Codilean 2022). Use of this database format required the addition of several key fields that may not have an immediate benefit for the user but serve an important function internally. These include: 'OBSID1', 'METASITEID', 'METANAME', 'REFDBID'#, and 'OAID'# (Codilean et al. 2022a). For similar reasons, codes in 'SMPID' and 'SITEID' fields have letter suffixes. All observations are assigned a letter 'a' suffix; if replicate analyses are present and their report includes some dissimilarities from the first observation, then a subsequent Latin-script alphabet letter will replace 'a'. This is in effect when there is a dissimilarity in information reported for the feature dated, square and spit designation, sample depth, site context and commentary etc. (see *global_SiteMaster* and *archSample* table fields in Munack and Codilean 2022). Hence, letter suffixes can be used for identifying changes or inconsistencies, however, can be ignored for most typical applications.

Future expansions and directions

Radiocarbon and luminescence dating methods were prioritised in SahulArch because they are the most frequently used across Australia and New Guinea. Other dating methods such as uranium series (U–Th) and electron spin resonance (ESR) are also used and will be considered for inclusion in future SahulArch collections. This will require engagement from those communities to help develop a database scheme that will conform with current data reporting standards, and the data repository and reproducibility principles of SahulArch.

Summary and conclusions

A key advantage of SahulArch is its spatial display of stored data in the OCTOPUS v.2 bespoke web-interface (Figure 8). Users can quickly access and assess information about sites and the type of data they contain. SahulArch is easily accessible whether for preliminary inspection online or download and detailed analysis offline. For example, if a distribution of sites with observations older than 30,000 years is sought, the simple application of a 'Filter' will display where these sites are located and how many samples match the criteria (Figure 8). Clicking on interactive datapoints will reveal the basic information for those samples including sample name, author and DOI link to the source article, age, and error (Figure 8). This allows users to assess ages in the context of other archaeological sites in a study region and determine where research gaps may be. More complex, multi-factor filtering schemes are possible in OCTOPUS and are illustrated in Rehn (2022). SahulArch can also be used as a reference database because each observation contains bibliographic information of the study that produced the age as well as additional related

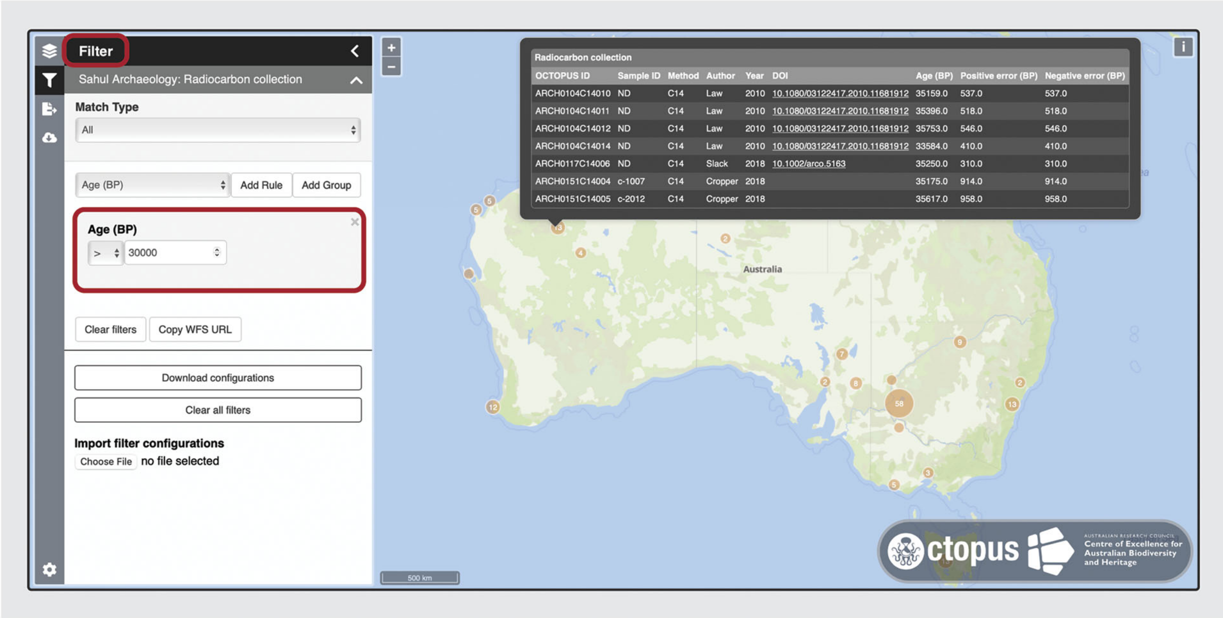


Figure 8. Sample of the OCTOPUS data display showing observation filtering and data summary view.

references. Hence, if undertaking research in a new area, SahulArch can be used as a starting point for identifying existing literature, which can be accessed through DOI links (Figure 8). However, SahulArch is focused on geochronology and is not exhaustive in other aspects of archaeological research.

We are confident that the ease of accessibility of SahulArch collections will make this database a useful tool for Traditional Owner communities and all those interested in archaeological geochronology, spatial distribution of archaeological sites, setting out new research plans, or writing grant proposals. SahulArch will be maintained and updated in the foreseeable future by the ARC Centre of Excellence for Australian Biodiversity and Heritage, and further plans for expansion, maintenance, and data collection are underway.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the ARC Centre of Excellence for Australian Biodiversity and Heritage [CE170100015]. Open access for this article was enabled by the 2022 Open Access High Impact Grants Scheme (OA-HI) offered by the University of Wollongong.

ORCID

Wanchese M. Saktura http://orcid.org/0000-0002-7198-7776
Emma Rehn http://orcid.org/0000-0002-9097-0232
Lauren Linnenlucke http://orcid.org/0000-0001-5915-2859
Henry Munack http://orcid.org/0000-0002-3750-9674
Rachel Wood http://orcid.org/0000-0002-1694-3295
Fiona Petchey http://orcid.org/0000-0002-3308-9983
Alexandru T. Codilean http://orcid.org/0000-0003-1895-5633
Zenobia Jacobs http://orcid.org/0000-0001-5424-5837
Tim J. Cohen http://orcid.org/0000-0003-4081-9523
Alan N. Williams http://orcid.org/0000-0003-4133-4229
Sean Ulm http://orcid.org/0000-0001-6653-9963

Data availability statement

The data that are presented in this study are openly available at: https://octopusdata.org, alternatively, information about each database collection and the data can be accessed through the following DOI links:

SahulArch Radiocarbon – https://doi.org/10.25900/gpvr-ay04; SahulArch OSL – https://doi.org/10.25900/9y07-4j77; SahulArch TL – https://doi.org/10.25900/af67-kh16

References

Bird, M.I., L.K. Ayliffe L.K. Fifield C.S.M. Turney R.G. Cresswell T.T. Barrows and B. David 1999 Radiocarbon dating of 'old' charcoal using a wet oxidation, stepped-combustion procedure. *Radiocarbon* 41(2):127–140.

Bradshaw, C.J.A., K. Norman S. Ulm A.N. Williams C. Clarkson J. Chadoeuf S.C. Lin Z. Jacobs R.G. Roberts M.I. Bird L.S. Weyrich S.G. Haberle S. O'Connor B. Llamas T.J. Cohen T. Friedrich P. Veth M. Leavesley and F. Saltré 2021 Stochastic models support rapid peopling of Late Pleistocene Sahul. *Nature Communications* 12(1):2440.

Brockwell, S., P. Faulkner P. Bourke A. Clarke C. Crassweller D. Guse B. Meehan and R. Sim 2009 Radiocarbon dates from the Top End: A cultural chronology for the Northern Territory coastal plains. *Australian Aboriginal Studies* 1:54–76.

Clarkson, C., Z. Jacobs B. Marwick R.L.K. Fullagar L. Wallis M.A. Smith R.G. Roberts E. Hayes K. Lowe X. Carah S.A. Florin J. McNeil D. Cox L.J. Arnold Q. Hua J. Huntley H.E.A. Brand T. Manne A. Fairbairn J. Shulmeister L. Lyle M. Salinas M. Page K. Connell G. Park K. Norman T. Murphy and C. Pardoe 2017 Human occupation of northern Australia by 65,000 years ago. *Nature* 547(7663):306–310.

Codilean, A.T., H. Munack and W.M. Saktura 2022a OCTOPUS Database v.2: Database Field Description Tables. Retrieved 24 November 2022 from https://doi.org/10.5281/zenodo.7352827.

Codilean, A.T., H. Munack T.J. Cohen W.M. Saktura A. Gray and S.M. Mudd 2018 OCTOPUS: An open cosmogenic isotope and luminescence database. *Earth System Science Data* 10(4):2123–2139.

Codilean, A.T., H. Munack W.M. Saktura T.J. Cohen Z. Jacobs S. Ulm P.P. Hesse J. Heyman K.J. Peters A.N. Williams R.B.K. Saktura X. Rui K. Chishiro-Dennelly and A. Panta 2022b OCTOPUS Database (v.2). *Earth System Science Data* 14(8):3695–3713.

Crabtree, S.A., D.A. White C.J.A. Bradshaw F. Saltré A.N. Williams R.J. Beaman M.I. Bird and S. Ulm 2021 Landscape rules predict optimal superhighways for the first peopling of Sahul. *Nature Human Behaviour* 5(10): 1303–1313.

David, B. L.J., Arnold J.-J. Delannoy J. Fresløv C. Urwin F. Petchey M.C. McDowell R. Mullett J. Mialanes R. Wood J. Crouch J. Berthet V.N.L. Wong H. Green and J. Hellstrom 2021 Late survival of megafauna refuted for Cloggs Cave, SE Australia: Implications for the Australian Late Pleistocene megafauna extinction debate. *Quaternary Science Reviews* 253:106781.

Galbraith, R.F., R.G. Roberts G.M. Laslett H. Yoshida and J.M. Olley 1999 Optical dating of single and multiple grains of quartz from Jinmium rock shelter, northern Australia: Part I, Experimental design and statistical models. *Archaeometry* 41(2):339–364.

GEBCO Compilation Group 2020 GEBCO 2020 Grid. Retrieved 16 November 2022 from https://doi.org/10.5285/a29c5465-b138-234d-e053-6c86abc040b9.

Gill, E.D. 2010 Aboriginal midden sites in western Victoria dated by radiocarbon analysis. *Mankind* 5(2): 51–55.

Godfrey, M.C.S., C.F. Bird D. Frankel J.W. Rhoads and S. Simmons 1996 From time to time: Radiocarbon information on Victorian archaeological sites held by Aboriginal Affairs Victoria. *The Artefact* 19:3–51.

Huxtable, J. and M.J. Aitken 1977 Thermoluminescent dating of Lake Mungo geomagnetic polarity excursion. *Nature* 265(5589):40–41.

Linnenlucke, L.P. 2022 Torres Strait Radiocarbon Database. Retrieved 16 November 2022 from https://doi.org/10.25903/8efv-yw59.

Maloney, T., S. O'Connor, R. Wood, K. Aplin and J. Balme 2018 Carpenters Gap 1: A 47,000 year old record of Indigenous adaptation and innovation. *Quaternary Science Reviews* 191:204–228.

McPherron, S.J.P. and H.L. Dibble 2009 E4 Software. Retrieved 16 November 2022 from https://www.oldstoneage.com/osa/tech/e4/.

Millard, A.R. 2014 Conventions for reporting radiocarbon determinations. *Radiocarbon* 56(2):555–559.

Munack, H. and A.T. Codilean 2022 OCTOPUS Database v.2: Relational Database Schema and Documentation. Retrieved 24 November 2022 from https://doi.org/10.5281/zenodo.7352807.

Murray, A.S. and A.G. Wintle 2000 Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol. *Radiation Measurements* 32(1): 57–73.

Rehn, E. 2022 OCTOPUS Web Interface User Guide. Retrieved 16 November 2022 from https://zenodo.org/record/6469039.

Rehn, E. H., Munack and W.M. Saktura 2022 OCTOPUS Web Feature Service (WFS) User Guide. Retrieved 16 November 2022 from https://doi.org/10.5281/zenodo.7063533.

Roberts, R.G., R.F. Galbraith J.M. Olley H. Yoshida and G.M. Laslett 1999 Optical dating of single and multiple grains of quartz from Jinmium Rock Shelter, northern Australia: Part II, results and implications. *Archaeometry* 41(2):365–395.

Roberts, R.G., R. Jones and M.A. Smith 1990 Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. *Nature* 345(6271):153–156.

Roberts, R.G., N.A. Spooner and D.G. Questiaux 1993 Cautions on the use of extended duration preheats in the optical dating of quartz. *Ancient TL* 11(2):47–54.

Roberts, R.G., H. Yoshida R.F. Galbraith G.M. Laslett R. Jones and M.A. Smith 1998 Single-aliquot and single-grain optical dating confirm thermoluminescence age estimates at Malakunanja II rock shelter in northern Australia. *Ancient TL* 16(1):19–24.

Saktura, W.M., E. Rehn H. Munack A.T. Codilean Z. Jacobs A.N. Williams and S. Ulm 2022a OCTOPUS Database v.2: The SahulArch OSL Collection (v.2). Retrieved 16 November 2022 from http://dx.doi.org/10.25900/9y07-4j77.

Saktura, W.M., E. Rehn H. Munack A.T. Codilean Z. Jacobs A.N. Williams and S. Ulm 2022b OCTOPUS Database v.2: The SahulArch TL Collection (v.2). Retrieved 16 November 2022 from http://dx.doi.org/10.25900/af67-kh16.

Saktura, W.M., E. Rehn L. Linnenlucke H. Munack R. Wood F. Petchey A.T. Codilean Z. Jacobs A.N. Williams and S. Ulm 2022c OCTOPUS Database v.2: The SahulArch Radiocarbon Collection (v.2). Retrieved 16 November 2022 from http://dx.doi.org/10.25900/gpvr-ay04.

Stuiver, M. and H.A. Polach 1977 Discussion reporting of ^{14}C data. *Radiocarbon* 19(3):355–363.

Thomas, D., A. Martin and W. Van Der Stelt 2018 A radiocarbon dating visualisation project for Aboriginal places in Victoria. In C. Spry E. Foley D. Frankel and S. Lawrence (eds), *Excavations, Surveys and Heritage Management in Victoria*. pp.13–17. Melbourne: La Trobe University, Bundoora.

Ulm, S. and J. Reid 2000 Index of dates from archaeological sites in Queensland. *Queensland Archaeological Research* 12:1–129.

Williams, A.N. and M.A. Smith 2012 AustArch 2: A database of ^{14}C and luminescence ages from archaeological sites in the Top End. *Australian Archaeology* 74:146.

Williams, A.N. and M.A. Smith 2013 AustArch 3: A database of ^{14}C and luminescence ages from archaeological sites in southern Australia. *Australian Archaeology* 761: 102.

Williams, A.N., M.A. Smith C.S.M. Turney and M.L. Cupper 2008 AustArch 1: A database of ^{14}C and luminescence ages from archaeological sites in the Australian arid zone. *Australian Archaeology* 66:99.

Williams, A., S. Ulm M. Smith and J. Reid 2014 AustArch: A database of ^{14}C and non-^{14}C ages from archaeological sites in Australia – Composition, compilation and review (Data Paper). *Internet Archaeology* 36(36):1–12.

Williams, A.N., P. Veth W. Steffen S. Ulm C.S.M. Turney J.M. Reeves S.J. Phipps and M. Smith 2015 A continental narrative: Human settlement patterns and Australian climate change over the last 35,000 years. *Quaternary Science Reviews* 123:91–112.