


















DEVELOPMENT OF THE INTCAL DATABASE

Christopher Bronk Ramsey^{1*}  • Florian Adolphi² • William Austin³ •
Edouard Bard⁴  • Alex Bayliss⁵  • Maarten Blaauw⁶  • Hai Cheng⁷ •
R. Lawrence Edwards⁸ • Michael Friedrich⁹ • Timothy Heaton¹⁰  • Alan Hogg¹¹ •
Quan Hua¹²  • Konrad Hughen¹³  • Bernd Kromer¹⁴ • Sturt Manning¹⁵  •
Raimund Muscheler¹⁶  • Jonathan Palmer¹⁷  • Charlotte Pearson¹⁸  •
Paula Reimer⁶  • Ron Reimer⁶ • David Richards¹⁹  • Marian Scott²⁰  • John Southon²¹  •
Chris Turney²²  • Lukas Wacker²³ 

¹University of Oxford, UK

²Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany

³University of St Andrews, UK

⁴CEREGE, Aix-Marseille University, CNRS, IRD, INRAE, College de France, France

⁵Historic England, UK

⁶Queen's University Belfast, UK

⁷Xi'an Jiaotong University, China

⁸University of Minnesota, Minneapolis, USA

⁹University of Hohenheim, Germany

¹⁰University of Leeds, UK

¹¹University of Waikato, New Zealand

¹²Australian Nuclear Science and Technology Organisation, Australia

¹³Woods Hole Oceanographic Institution, USA

¹⁴Heidelberg University, Germany

¹⁵Cornell University, USA

¹⁶Lund University, Sweden

¹⁷University of New South Wales, Australia

¹⁸University of Arizona, USA

¹⁹University of Bristol, UK

²⁰University of Glasgow, UK

²¹University of California Irvine, USA

²²University of Technology Sydney, Australia

²³ETH Zurich, Switzerland

ABSTRACT. The IntCal family of radiocarbon (¹⁴C) calibration curves is based on research spanning more than three decades. The IntCal group have collated the ¹⁴C and calendar age data (mostly derived from primary publications with other types of data and meta-data) and, since 2010, made them available for other sorts of analysis through an open-access database. This has ensured transparency in terms of the data used in the construction of the ratified calibration curves. As the IntCal database expands, work is underway to facilitate best practice for new data submissions, make more of the associated metadata available in a structured form, and help those wishing to process the data with programming languages such as R, Python, and MATLAB. The data and metadata are complex because of the range of different types of archives. A restructured interface, based on the “IntChron” open-access data model, includes tools which allow the data to be plotted and compared without the need for export. The intention is to include complementary information which can be used alongside the main ¹⁴C series to provide new insights into the global carbon cycle, as well as facilitating access to the data for other research applications. Overall, this work aims to streamline the generation of new calibration curves.

KEYWORDS: calibration, database, dating, dendrochronology, radiocarbon.

*Corresponding author. Email: christopher.ramsey@arch.ox.ac.uk



INTRODUCTION

The generation of the IntCal, Marine, and SHCal radiocarbon (^{14}C) age calibration curves (Heaton et al. 2020a; Hogg et al. 2020; Reimer et al. 2020) is only possible because of the research and care which goes into generating and checking the underlying ^{14}C datasets used in their construction. In particular, the assessment of data quality relative to set criteria and the collection of associated metadata (Reimer et al. 2002, 2013) are key elements in the ongoing compilation of these datasets. In addition to their value for ^{14}C calibration, these ^{14}C datasets also have an importance in their own right for a wide spectrum of different areas of research (Heaton et al. 2021) and have routinely been made available to the international research community through the IntCal database (previously available from the Queen's University Belfast webserver). The developments reported here build upon that initiative with the intention of facilitating access to, and enhancing transparency of, the broad range of information used in the construction of the calibration curves.

At present, the data collected and discussed here relate only to the construction of the pre-1950 IntCal, Marine, and SHCal curves that span from 55,000–0 cal yr BP. However, the intention is to broaden this to include the data for calibration in the post-1950 period (Hua et al. 2022).

AIMS OF THE DEVELOPMENT WORK

The main aims of the new database structures are to facilitate the sharing of not only the primary, raw, ^{14}C data used for calibration purposes but also the related data (such as tree ring measurements), metadata (including method statements and images), and to do so in a way that enables the data to be manipulated programmatically using tools such as R (R Core Team 2022), Python (e.g., Van Rossum and Drake 2009), and MATLAB (e.g., Higham and Higham 2016). The specific objectives are to:

1. Make all the primary data (the ^{14}C measurements and accompanying calendar age information with their associated uncertainties) that are used in construction of the IntCal family of calibration curves available in a digitally readable form.
2. Organize this data, as far as possible, by primary record rather than measurement initiative, in order to facilitate comparison of results from multiple laboratories and to put the data into a geographic context.
3. Particularly in relation to dendrochronology, provide supporting data such as ring-width series, and metadata detailing the methods that have been applied (Reimer et al. 2013).
4. Include associated numerical data such as calendar age correlation or covariance matrices for any record where the approach used to generate the timescale introduces dependencies between the estimates of its calendar ages, for example, the annually resolved records such as the Lake Suigetsu record (Bronk Ramsey et al. 2020) with its modified varve-counting chronology, and ancient New Zealand kauri trees wiggle-matched onto ^{14}C variability (Cooper et al. 2021). In a similar way, the Cariaco Basin (Hughen and Heaton 2020) and the Pakistan and Iberian margin records (Bard et al. 2013) which have calendar ages obtained by tuning to closely related climate markers (Heaton et al. 2013), use the proxy data and derived relationships in the construction of the calibration curve (Heaton et al., 2020b). The dating and proxy information for the speleothem records is also fundamental to the timescale for the curves.

5. Associate (or link) the data and metadata with the relevant publications, providing complete DOI information and URL links to the publications. This includes links to complimentary data archives, particularly those relating to the supporting dendrochronological data.
6. Provide methods for visualization of the ^{14}C and dendrochronological data.
7. Include tools for checking and assessing data.
8. Allow for the easy import and export of both data and references to a range of other software.

These aims are intended to help different groups of researchers working with the data: those wishing to use or independently evaluate the data, those working on the preparation of new datasets, and the members of the IntCal group working on curve construction. It is important to remember that research on the underlying records is an ongoing process and that, for example, the timescales for the speleothems and their derived chronologies will change in future iterations (Cheng et al. 2021).

OVERALL DATA MODEL

In order to achieve these aims, use has been made of the pre-existing IntChron framework (Bronk Ramsey et al. 2019). This is specifically designed for sharing linked data and includes elements relevant to chronological data, particularly the ability to handle different timescale units and tools for data visualization. The data schema and associated tools have been updated to include elements required for the IntCal datasets, but the overall data model has been found to cover all the core IntCal requirements.

Data Organization

The data structure is essentially hierarchical but with the ability to link and associate information by searches. At the top level there are three main classes of information, with all other information organized within this structure:

1. Records: These contain all information relating to individual sites or records; each record has a unique short name, or site code, which is used as a key for accessing information. Within the record, there are three types of information:
 - a. Information on the location (latitude, longitude, elevation) and type (e.g., Marine, Terrestrial, Speleothem, ...) of the record.
 - b. Data series lists which are used to hold data for the record. For this application the main data series types are:
 - i. IntCal data: the primary ^{14}C calibration data (containing ^{14}C measurements and quoted uncertainties, as well as the accompanying calendar age information).
 - ii. Dendrochronological sample data: the ring-width or oxygen isotope data associated with the measurement series (if appropriate).
 - iii. Metadata: other descriptive information about the record.
 - iv. Attachments: files (typically images) cited in the metadata.
 - v. Other data, such as ages and correlation or covariance files for records with a level of dependence between the calendar age estimates can be added with specific series types.
 - c. Reference links for the record as a whole and for the series within it.

2. Project data: which refers to data series that do not have a specific link to single records or sites. In the case of the IntCal database, information held at this level includes the calibration curves themselves and an index of data series organized by the IntCal dataset number. This also includes information on the time-variant relationship between the calibration curve and other (e.g., ice-core) timescales (Adolphi and Muscheler 2016; Adolphi et al. 2018).
3. References: which holds full bibliographic details for all references referred to in the database (typically listed under records or data series). These references are usually directly linked to the relevant journal articles via doi.org and can be exported in BibTeX format for use in bibliographic tools.

Data Storage

The underlying storage format within IntChron is JSON because this facilitates easy interfacing with programming tools (most easily using tools such as R, Python, and MATLAB, but also in principle with languages such as C++ and C#), static archiving/storing of data without the need for software installation (Bronk Ramsey et al. 2019), and is commonly used in web-based applications.

The data will be stored and used in three different ways:

- i. There is an active database accessible to members of the IntCal working group which is intended to help develop new calibration curves. This allows for the addition of new records and the updating of existing ones.
- ii. There is a static archive of the IntCal20 datasets at <https://intchron.org/archive/IntCal/IntCal20/index.json>, which holds all the information accumulated for IntCal20 as it was when the curve was constructed. This is open access.
- iii. It is possible for users to make their own copies of this archive (or parts of it) for their research and to prepare new data for inclusion in IntCal. These can be stored on the IntChron server in users' own areas, or on users' own computers.

In the case of the IntCal database itself (point "i" above), a hybrid approach has been adopted. The user interface, archives and file transfer all use JSON, but for the primary IntCal data, an underlying MySQL data table has been retained to minimise the risks of unintended changes. In addition, the JSON data records are stored in a database rather than as files. This has some advantages for a multi-user system and enables the database maintainers to use an automated archiving system on the computer server which holds the database. However, from the perspective of a user of the database or of the derived archives, the data organization is effectively in the form of JSON objects.

Reflecting the data organization, there are only three types of JSON files that are used for data exchange and presentation. These are:

1. Project data: these files hold links to all of the relevant records and project-level data series. In addition, the project data file contains all of the relevant publication information and details of parameter characteristics.
2. Record data: these contain all the information relevant to a specific record, including the data series contained within it.

3. Series data: these are for project-level data series which do not relate to specific records (such as calibration curves).

The overall database model is based on linked data, so it can also include references to files held outside the JSON data structure via attachments (typically images). In general, the aim has however been to avoid putting key information in such attachments because it makes for more difficult data distribution.

As a variant of the above model, it is possible to have all record and series data embedded into a single project data file. We have used this option for the data archive because it means that all of the data (other than attachments) can be retrieved from a single file rather than having to make multiple requests for each record and series. This approach would not be efficient for very large projects but poses no problems for the present IntCal datasets in terms of data handling (the entire IntCal20 archive is < 10 MB).

Attachment files are organized in a hierarchical file structure based on record or series name. A full archive of the IntCal20 data files is included as supplementary online information in this publication, allowing the complete archive to be reproduced without access to the current site. This is an important element in the long-term availability of the data.

Software Overview

In principle, the archive can be worked with entirely using software tools such as R, Python, and MATLAB. However, the IntChron integration tool (Bronk Ramsey et al. 2019) has been further developed to facilitate work with the IntCal database and provide a user interface for some pre-defined data manipulations. This tool can be accessed and preloaded with the open access static IntCal20 data using the url:

<https://intchron.org/tools/integrate/integrate.html?filename=https://intchron.org/archive/IntCal/IntCal20/index.json>

The integration tool has been specifically designed to work with data organized in the format described above, and the user is presented with a list of records and data series. The data can be explored by following links from this level. There are also search facilities built into the system which enable lists of relevant data series or primary data to be extracted. Records can be displayed on a map and the associated information retrieved by selecting the individual location points. Figure 1 shows a screenshot of the application in use.

This tool is available on the IntChron server and the full interface code will also be distributed with future releases of OxCal so it can be used without access to the IntChron site should that be necessary.

DATA ELEMENTS

The overall data schema for IntChron (including the parameters used for the IntCal dataset) is given at <https://intchron.org/schema> with the current version supplied in the supplementary information for this paper. Here we will focus on how the data are used specifically for IntCal, concentrating specifically on elements which might be less intuitive or where this might not otherwise be obvious. Each parameter has a formal name used in the JSON objects and for searches, and a display-orientated name (see Table 1). Parameter, series names and record

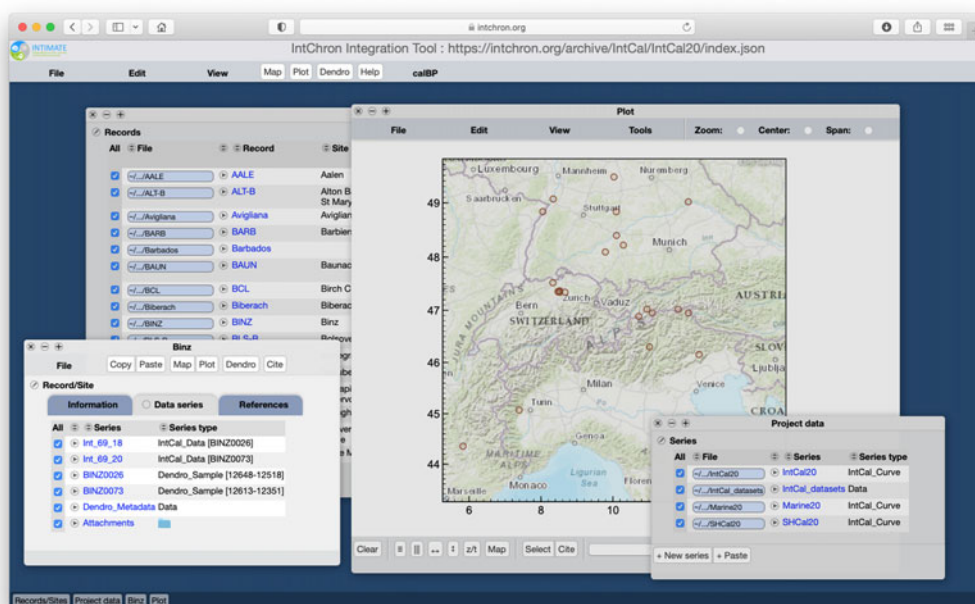


Figure 1 Example screenshot of the database interface showing (from top left, clockwise) windows containing: the list of records, a map of the records in part of Europe, the project data series, and the data series for the “Binz” record (showing the different data types included in this case).

names are all defined so that they are valid JavaScript variable names and can also be used as URLs without escaping.

Record Level Information

At the record level, the key information held is about the location. For dendrochronological samples, the genus and species of wood sampled (e.g., *Quercus robur*, *Agathis australis*) is also included (see Table 1 and Figure 2 for details). There is an optional comment field at this level which can consist of information not included in the standard parameters. The references at this record level should only be the key papers relevant to the sample set within IntCal. More specific references, for example to the dendrochronology, can be included with the dendrochronological series or metadata. The record name is ideally a formal site code (such as SG06) or failing that a short form of the site name (such as Maraa). In the case of records containing compilations of data (mostly older datasets) the lab code, country of origin of the samples, and taxon are used together (as in QL_DE_Oak).

Data Series Types

There are four main data series types included in the records. These are: IntCal_Data (the primary ^{14}C calibration data used in IntCal), Dendro_Sample (dendrochronological data such as ring-widths for the samples), Data (a generic holder used primarily for metadata) and Files (attachments which are used principally for figures which cannot be included in any other way).

Table 1 Main parameter names with special meaning within the IntCal database, organized by data table.

Parameter(s)	Display name(s)	Usage
Record level		
record	Record	Site code – defined without spaces or special characters to enable use in URLs.
site, region, country	Site, Region, Country	Details of the record site. Site names are kept brief since longitude and latitude define the location.
site_type	Site type	For IntCal normally “Marine”, “Terrestrial”, “Speleothen”, “Dendrochronological”.
longitude, latitude	Long, Lat	Location in decimal degrees.
elevation	Elevation	Elevation above sea level (m).
z_type	z type	Defines how z parameters are used for this record. Typically here, “rings” for dendrochronological samples, “depth” for sediments, or “height” for speleothenes.
z_units	z units	Most relevant for depths and heights where the data is stored in m but can be displayed in “m”, “cm”, or “mm”. Set to “rings” for dendrochronology.
taxon	Taxon	Genus and where possible species of tree used for dendrochronological samples.
IntCal_Data series headers		
series	Series	For IntCal series, this is normally defined by the set and division number, so Int_12_3 would be IntCal set 12, division 3.
intcal_set_type	Type	Normally “Int” for IntCal, “SH” for SHCal.
intcal_set	Set	Set number for approved data series. For convenience the IntCal data has been divided into sets that include data from major research projects or groups.
intcal_division	Division	Division number within the set. The sets are subdivided into divisions where this is thought to be useful.
series_sample	Sample	For dendrochronological samples this should normally be the tree sample code (such as ABCD012).
IntCal_Data series data		
z	Ring/Depth/Height	Depending on the context, shows the linear mean position of a sample. It should be the mid-value and so might be non-integer, for example for a decadal block.
z_range	z range	This can either be displayed separately or as a range of values. For tree rings this is zero for a single ring.

Table 1 (Continued)

Parameter(s)	Display name(s)	Usage
t	Date	Floating-point astronomical year time parameter for the sample based on all the available information. Where relevant this is the posterior estimate arising from the curve compilation. It can be displayed in a range of different formats (calBP, AD, CE etc.).
t_sigma	$\pm 1\sigma$	One sigma uncertainty in the age (posterior estimate where relevant).
intcal_data_id	Id	Unique ID number for samples stored in the IntCal database (null for new data).
intcal_seq_no	Seq	An optional sequence number used for a particular dataset. These numbers are used for correlation matrix indices.
sample	Sample	For dendrochronological samples this should typically be the tree sample code and ring number as in ABCD01_r3 or ABCD01_r11-19. In all cases the sample name should be unique to the sample itself within that record and shared by samples measured by different labs. The ages of samples with the same name in the same record are defined to be the same age.
ring_segment	Seg.	Section of tree-ring where required. Possible values are EW (early wood), LW (late wood) or subdivisions EW1, EW2, LW1, LW2 or some combination such as EW/LW. This value is not specified in cases where there was not attempt to subdivide rings as is usually the case for multi-year blocks.
calage	CalAge	The prior estimate of the age of the sample in calBP (following the Schulman convention where appropriate in the Southern Hemisphere; Schulman 1956). Fractional values are used for combinations of even numbers of samples.
calage_range	CalSpan	The number of years of material included: 1 for a single year and 10 for 10 years. This value is typically equal to z_range+1 for tree ring samples.
calage_sigmaI	$\pm 1\sigma_I$	Independent uncertainty in individual sample ages (if dated separately) or if there is an associated correlation matrix.
calage_sigmaD	$\pm 1\sigma_D$	The systematic uncertainty in the age of a set of samples (typically a floating tree ring series).
calage_sigmaC	$\pm 1\sigma_C$	An additional correlated component in the uncertainty in the age of a set of measurements when there is also a calage_sigmaI. Not currently used within any of the approved data series.

(Continued)

Table 1 (*Continued*)

Parameter(s)	Display name(s)	Usage
labcode	Lab code	Unique lab code for this measurement. If required, any pre-treatment and target subdivisions should be included in this code if required. A labcode of ABC-3456.1.2 implies the result was measured at lab “ABC” on sample number 3456 using the first pre-treatment and the second target. The details of any pre-treatment should be given in the linked publications for the dataset.
batch	Batch	This is an optional batch name for the measurements where there is an associated common element to the uncertainty.
r_date	R_Date	The radiocarbon date determination expressed as a Conventional Radiocarbon Age (CRA) in ^{14}C years BP.
r_date_sigma	$\pm 1\sigma$	The one sigma uncertainty in the r_date value.
r_date_sigma_extra	$\pm 1\sigma_+$	Any additional component to the uncertainty (typically used for samples with a reservoir where there is an estimated additional scatter relative to the atmospheric).
r_date_sigma_batch	$\pm 1\sigma_B$	Element of the uncertainty included in r_date_sigma but common to all measurements in the batch.
r_date_sigma_mult	$\pm 1\sigma_x$	An error multiplier applied to the r_date_sigma of some older datasets.
intcal_reservoir	Reservoir	The estimate of the reservoir offset of the data point. Used for marine and speleothem data.
intcal_reservoir_sigma	$\pm 1\sigma$	The uncertainty in any reservoir offset.
Dendro_Sample header		
t_from	From	Date of the first ring in fractional astronomical years displayed in the same format as t.
t_to	To	Date of the last ring in fractional astronomical years.
ring_count	Ring count	The total number of rings in the sample.
dated	Dated	True or false. If true, the t parameter is treated as an age; otherwise it is the group ring number.
Dendro_Sample data		
z	Ring	The ring number, typically from 1 for the oldest ring but in some traditions with 1 as the youngest ring.
t	Date	The fractional astronomical year date for the ring. See the text for more details.
ring_width	Ring width	The ring width in mm.
sample_depth	Sample depth	The number of distinct trees compiled into this dataset (typically 1 for a single sample).
d18O	$\delta^{18}\text{O}$	If needed, the $\delta^{18}\text{O}$ used instead of or in addition to ring_width.

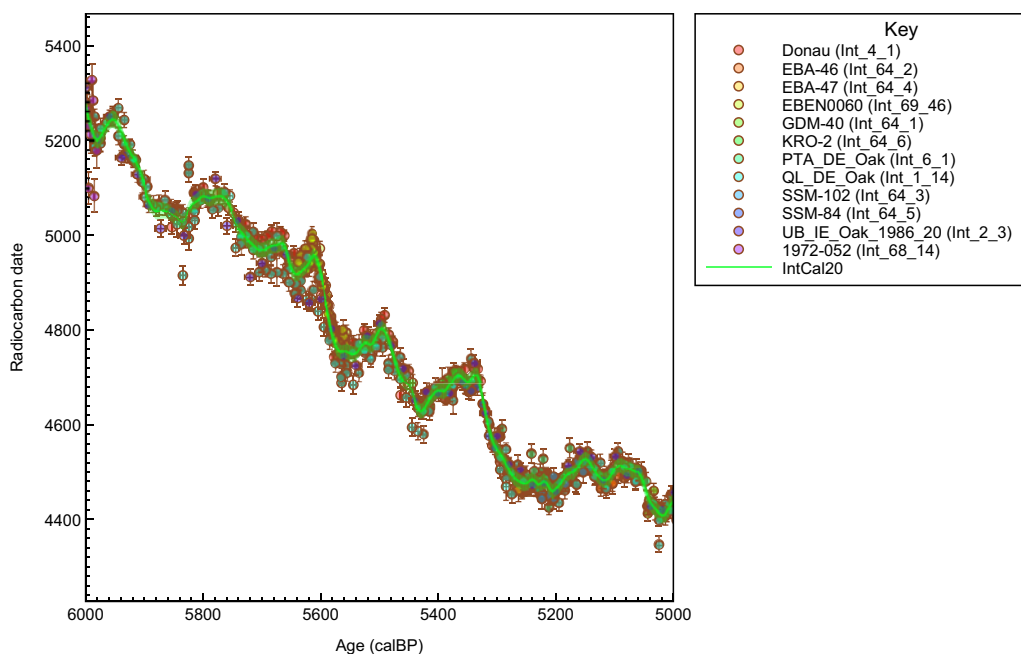


Figure 2 Example of plot drawn from the IntCal database showing the main datasets in IntCal20 over the time period 6000–5000 calBP. For the calibration curves, the series name is given; for all other data the record name is given with the series name in brackets. See section on record level information and Table 1 for details of record and series names.

IntCal Data Series

For IntCal, these are the most important element of the data held in the database. The definitions of the parameters included are given in Table 1 and also in the online schema. For more recent data, each data series (with a set and division number) contains measurements from a single laboratory on a single sample set (environmental record or single tree sample). For some older datasets, multiple samples from a combined chronology are summarised within a single set/division. Ideally, for dendrochronological samples, the sample parameter in the series header will be the same as the series name for the associated dendrochronological data, allowing the two to be cross-checked especially where there are multiple-year block samples used.

Each sample has its midpoint age expressed in two ways. The first is in fractional astronomical year format (t), which takes account of the growing season for tree rings (see next section). For those samples/archives which have uncertain calendar ages when entering IntCal, these t values are the posterior mean calendar age of the sample (obtained *after* the calendar ages have been updated during IntCal construction, Heaton et al. 2020b). The second is “calage” which gives the calendar age of the mid-point. The parameter calage refers to the year from which the sample derives (in calBP), so for NH wood this would be 0 for wood from AD1950, which would have a t value of 1950.5. For SH wood, the calage parameter assumes the Schulman convention (Schulman 1956), and thus calage = 0 would imply wood which started to grow in the austral spring of AD1950 (ending in the austral autumn of 1951), so it should have a t value of 1951.0. For samples dated by means other than dendrochronology, the calage is simply the

prior age estimate before AD1950 with an associated uncertainty. For those samples/archives which have uncertain calendar ages when entering IntCal, the calage values provide the prior calendar age estimate before curve construction. Consequently, in the construction of the calibration curve the calage parameter is used as the inputs for curve construction, while the t parameter provides the posterior estimate of the age after curve construction has been completed and is the measure used for plotting purposes.

Dendrochronological Sample Data Series

These can be imported from various dendrochronological formats directly, or from the NOAA International Tree Ring Data Bank (ITRDB: NOAA NECI 2022) exported as Tucson.rwl files. The internal format of these files has some unique features needed for this global dataset. Ring numbering is typically from old to young but can be reversed where this is the case for the primary data. The date (t) parameter is a floating-point astronomical year and should reflect the growing season for the wood. As previously described, NH wood grown in 1950 will be stored as 1950.5, whereas SH wood that starts to grow in 1950 will be stored as 1951.0. In the interface you can choose whether to use the Schulman convention for display purposes. If this is selected, 1951.0 will be shown as AD1950– showing that it is the end of this year, whereas if the convention is not applied it will show as AD1951+ indicating the start of the year. Alternatively, all dates can be shown in fractional format. The purpose of the internal format is that samples can be plotted on an absolute timescale that takes account of the different growing seasons between the two hemispheres.

Metadata

In order to properly understand the background to the dendrochronology underpinning the datasets used in IntCal, the database includes metadata. Such metadata are of greatest importance when the associated data are not already published elsewhere. The data included have been selected to include key information needed as outlined in Reimer et al. (2013). The metadata are usually structured to address key questions (see Table 2) and are in the form of a plain text file. Where tables are required, these can be incorporated within the main notes field using tab characters. A “code” field can also be added if needed (for example, COFECHA output extracts). The metadata should be kept succinct and not include unnecessary detail published or deposited elsewhere. Some of the information in the metadata is also held in a more structured way in the main database structure.

Attachments

In addition to the text information included in the structured database, attached files can also be included. These are mostly intended to be figures referred to in the metadata but can consist of longer datafiles and pdf reports if essential. It is however preferable that such information is published elsewhere and only referenced/linked in the IntCal database itself. Such attachments should not be seen as an alternative to the provision of structured information within the database.

Other Data

The calendar age correlation or covariance matrices, for those records with timescales that have been obtained using approaches that introduce a level of dependence between the calendar age estimates, are the main other data required for the IntCal calibration curve

Table 2 Main topics covered in dendrochronological metadata included where this information is not otherwise available in linked publications, and where this information is duplicated elsewhere in the database.

Note	Content/question	Field	Location
a	Sample identifiers	series series_sample	Dendro_Sample IntCal_Data
b	Sample derives from a single tree (Y/N)		
c	Site	site etc.	Record
d	Object type (fossil wood/standing building etc.)		
e	Preservation (dry/waterlogged etc)		
f	Sample type (core/slice etc)		
g	Taxon /Genera/Species	taxon	Record
h	No of rings	ring_count	Dendro_Sample
i	Pith (Y/N)		
j	Bark (Y/N)		
k	Waney edge (Y/N)		
l	Sapwood (Y/N)		
m	Methodology used for dating (text & tables)		
n	Dated individually (Y/N)		
o	Raw ring widths available (Y/N) (values are stored in a separate table)	ring_width	Dendro_Sample
p	Remainder of sample stored in which archive?		

generation. These are stored in a particular IntCal_Correlation dataset which holds the matrix as a simple tab-delimited text matrix that can be read by standard software packages. See for example the SG06 record. This Lake Suigetsu record has an adjusted varve-counted chronology (Bronk Ramsey et al. 2020) whereby calendar age uncertainties at any individual depth within the core are propagated to the other depths (due to both the necessary depth ordering and limitations on changing sedimentation rates).

Age-depth models, proxy data and other types of dating information can also be included as outlined in Bronk Ramsey et al. (2019).

TOOLS FOR DATA VISUALIZATION

In addition to allowing the data to be explored, the IntChron integration tool user interface has functions that enable the data to be plotted and visualised in several different ways.

Mapping of Records

The first of the display methods is a mapping interface. This allows the records (either as a whole, or a selected subset) to be shown on a map (as in Figure 1). The map is dynamically linked, so hovering over the points will give their site name, and clicking on them will bring up the relevant record and associated data. A single site can also be mapped to check its location.

Plotting Radiocarbon Data

The ^{14}C calibration data can be plotted in several different ways. To select data to be plotted, you can either navigate to the records and add them to an accumulating plot, or you can use the top-level [Plot] function to select a period and the types of data that are to be included. The plotting can be against any of the main time-scale measures, overlay the appropriate calibration curves, and can use ^{14}C age, $F^{14}\text{C}$ or age-corrected $\Delta^{14}\text{C}$ as the plotted value. All errors are stored, displayed and plotted at 1σ . The plotting routines collate the associated publications, so selecting the [Cite] option for a plot will give a reference listing.

Plotting of Dendrochronological Data

It is also possible to plot the tree ring data included in the database using either the raw ring widths or filtered values. This is most useful when there are multiple sets from the same chronology or the user imports additional series either from the ITRDB or by uploading files.

TOOLS FOR DATA PREPARATION AND ASSESSMENT

One of the main reasons for the choice of the underlying data model for the IntCal database was to make the task of preparing, adding, and assessing data easier for both submitters of data and those involved in the compilation. Clearly, the central database itself cannot be open for modification, but by enabling users to make their own copies of the database, this allows additions and changes to be tested with full access to the associated tools. It is also hoped that by enabling this, data providers (who normally understand their data best) will be able to get the data ready for submission themselves.

There is a help facility within the software, available through the [Help] menu, and this contains specific information for IntCal data which will be kept up-to-date with any developments in the interface. Dedicated functions to extract values from the `intcal20.json` file have also been added to the `rintcal` R package, available on R's CRAN repository (through the R command `install.packages('rintcal')`).

This paper aims to indicate what is possible without being a complete guide as to how it is done. Workshops and recorded videos will be made available to explain this in further detail.

Creation of New Records

Within the IntChron framework, creating a new record is simply done by selecting the [+New record] option; this will prompt the user for a record name and then allow all the main record data fields to be filled in (see Table 1 or examples in the existing archive). The location of the site can be checked on a map using the [Map] option.

Creation of New Data Series

Once the record has been created new data series can be added by selecting "Add to data series"; again the user will be prompted for a series name and can then choose the type of data (typically `IntCal_Data` or `Data` for this purpose). An [Import/Export] function allows data to be imported from or exported to a spreadsheet (see Figure 2).

Working with Dendrochronological Data

Dendrochronological data can be most easily added in one of two ways. Within the record using [File > Import] will bring up the option of importing “Dendro” data (which will allow Heidelberg, Tucson or some other formats to be imported) or “NOAA NCEI study” which allows importing of a study already available within the ITRDB. The latter option will first add a link to the study, and then it is possible to import the raw ring data for the relevant sample from that link.

The main advantages of having raw ring width data within the database are twofold. Firstly, it allows the ^{14}C data to be directly related to the ring width data by the shared number of the rings; if the dendrochronology is ever revised, this will allow for the correction of the dataset. Secondly, it enables the chronology of the dendrochronological series to be directly related to the ^{14}C dataset. Suppose the dated tree-ring series is included in the database and the ^{14}C samples have the ring numbers listed. In that case, it is possible to set the calage and t parameters of the ^{14}C series directly without having to work these out independently. This should avoid problems arising with different conventions, timescales and growing seasons, all of which can be handled within the interface. Even if this is not used directly, it is still possible to check the dendrochronological and ^{14}C series against one another, as in Figure 3. Such a plot has proved to be very useful for checking for internal consistency between dendrochronological reports and ^{14}C datasets.

INTCAL WORKFLOW

Figure 4 shows the intended workflow for using the IntCal database. Users will typically start with the static archive, add their own data and check them using the visualization tools provided. Once checked, they would then send their data for possible inclusion in the main IntCal database and use for calibration curve generation.

Making Partial Copies of the Archive

Starting with the online archive (see above), users can save the whole archive or parts of it. Assuming they only wish to save part of it, they can select records either one by one— or most easily by using the plotting function described above:

1. Use the [Plot] option to select the time range and sample type.
2. Use the [Edit > Deselect all] to deselect the whole archive.
3. Use the [Select] option in the plotter to select all plotted records.

The user will then have the subset of records required. They can then save this in one of two ways. They can create their own project on the IntChron server by using [File > Save as] and then give a project name, or they can use [File > Download] to download the data (choosing only the selected data). The downloaded file can be uploaded into another project running on the server or the user’s computer.

Addition of New Data

Adding new data involves the creation of the record (as described above) and then the input of the key data. This will always include IntCal_Data and if relevant Dendro_Sample and

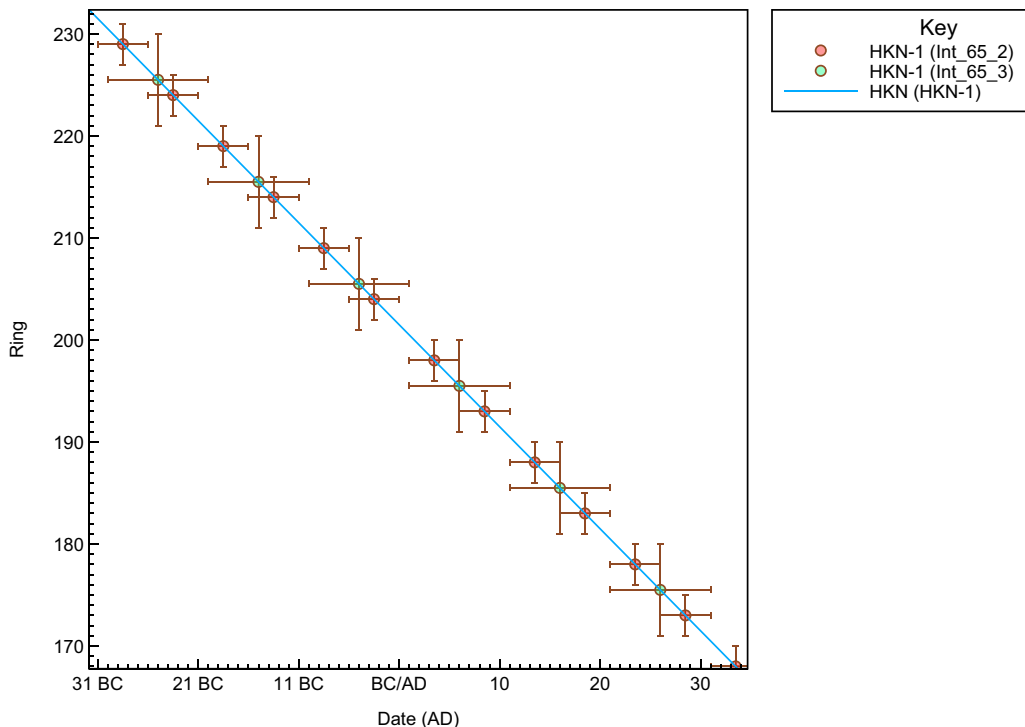


Figure 3 Example of a plot of ring number against age for sample HKN-1 from Hakone town, Kanagawa, Japan. The x-axis is the dendrochronological age (error bars are the span) and the y-axis is the ring number (error bars are the range). The blue line shows the dendrochronological sequence and the Int_65_2 and Int_65_3 datasets are two radiocarbon datasets measured at different resolutions. In this case ring number goes from young to old and the coincidence of the line with the datapoints shows that the radiocarbon and dendrochronological datasets are consistent. (Please see online version for color figures.)

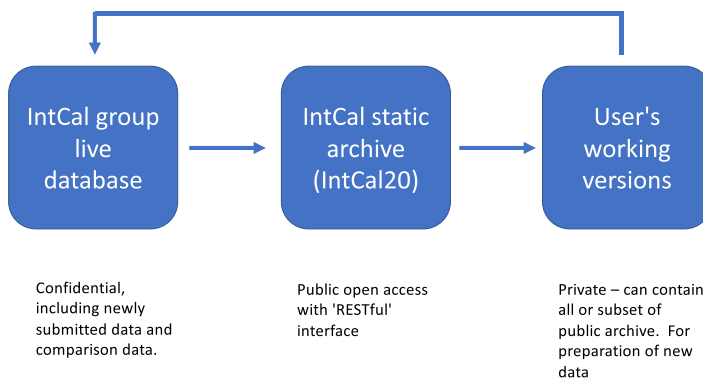


Figure 4 This shows how different versions of the IntCal dataset can be used by researchers wishing either to use the data, or to submit new data for the calibration curve.

metadata using the Data type to reproduce what is seen in the archived datasets. The `intcal_set`, `intcal_division`, and `intcal_record_id` can be left as null.

Submission of Data to the IntCal Group

The new record and dataset can then be sent to the relevant members of the IntCal group by using the [File > Download] option under the new record. This will download a JSON file for that record which can be considered for inclusion in the central database once all the appropriate checks have been made.

CONCLUSIONS

It is hoped that the provision of this comprehensive database for the IntCal project will facilitate the work of the ^{14}C research community wishing to use the IntCal data and those wishing to contribute to it. The intention is to make the tools and all of the data used within the group open for all researchers. We consider it particularly important that all of the extensive work which underlies the datasets is fully referenced and that these references are easy to use for those accessing the data. Work is underway to prepare for the next update of the IntCal calibration curves, and loading data into this updatable format is intended to be the first stage in that process.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/RDC.2023.53>

ACKNOWLEDGMENTS

The authors of this paper would like to acknowledge all of the work which has gone into generating the datasets included within this dataset, and all of those who have helped in their compilation for IntCal20 and earlier iterations of the calibration curves. In addition, we would like to acknowledge a number of young researchers who have tested and worked with the IntChron interface through its development over the last few years, particularly Rebecca Kearney, Lorena Becerra-Valdivia and Jakov Mlinaravic. Finally, we would like to acknowledge the support of the NERC NEIF facility (PR19008-NEIF-Oxford) for the underlying computing infrastructure and support for the IntChron interface development.

REFERENCES

- Adolphi F, Muscheler R. 2016. Synchronizing the Greenland ice core and radiocarbon timescales over the Holocene—Bayesian wiggle-matching of cosmogenic radionuclide records. *Climate of the Past* 12(1):15–30. doi: [10.5194/cp-12-15-2016](https://doi.org/10.5194/cp-12-15-2016)
- Adolphi, F, Bronk Ramsey C, Erhardt T, Lawrence Edwards R, Cheng H, Turney CSM, Cooper A, Svensson A, Rasmussen SO, Fischer H, Muscheler R. 2018. Connecting the Greenland ice-core and U/Th timescales via cosmogenic radionuclides: testing the synchronicity of Dansgaard-Oeschger events. *Climate of the Past* 14:1755–1781. doi: [10.5194/cp-14-1755-2018](https://doi.org/10.5194/cp-14-1755-2018)
- Bard E, Ménot G, Rostek F, Licari L, Böning P, Edwards RL, Cheng H, Wang Y, Heaton TJ. 2013. Radiocarbon calibration/comparison records based on marine sediments from the Pakistan and Iberian margins. *Radiocarbon* 55(4):1999–2019. doi: [10.2458/azu_js_rc.55.17114](https://doi.org/10.2458/azu_js_rc.55.17114)
- Bronk Ramsey C, Blaauw M, Kearney R, Staff RA. 2019. The importance of open access to chronological information: the Intchron initiative. *Radiocarbon* 61(5):1121–1131. doi: [10.1017/rdc.2019.21](https://doi.org/10.1017/rdc.2019.21)
- Bronk Ramsey C, Heaton TJ, Schlolaut G, Staff RA, Bryant CL, Brauer A, Lamb HF, Marshall MH, Nakagawa T. 2020. Reanalysis of the

- atmospheric radiocarbon calibration record from Lake Suigetsu, Japan. *Radiocarbon* 62(4):989–999. doi: [10.1017/rdc.2020.18](https://doi.org/10.1017/rdc.2020.18)
- Cheng H, Xu Y, Dong X, Zhao J, Li H, Baker J, Sinha A, Spötl C, Zhang H, Du W, et al. 2021. Onset and termination of Heinrich Stadial 4 and the underlying climate dynamics. *Communications Earth & Environment* 2(1):230. doi: [10.1038/s43247-021-00304-6](https://doi.org/10.1038/s43247-021-00304-6)
- Cooper A, Turney CSM, Palmer J, Hogg A, McGlone M, Wilmshurst J, Lorrey AM, Heaton TJ, Russell JM, McCracken K, et al. 2021. A global environmental crisis 42,000 years ago. *Science* 371(6531):811–818. doi: [10.1126/science.abb8677](https://doi.org/10.1126/science.abb8677)
- Heaton TJ, Bard E, Bronk Ramsey C, Butzin M, Köhler P, Muscheler R, Reimer PJ, Wacker L. 2021. Radiocarbon: a key tracer for studying Earth's dynamo, climate system, carbon cycle, and Sun. *Science* 374:eabd7096. doi: [10.1126/science.abd7096](https://doi.org/10.1126/science.abd7096)
- Heaton T, Bard E, Hughen K. 2013. Elastic tie-pointing—transferring chronologies between records via a Gaussian process. *Radiocarbon* 55(4):1975–1997. doi: [10.2458/azu_js_rc.55.17777](https://doi.org/10.2458/azu_js_rc.55.17777)
- Heaton TJ, Köhler P, Butzin M, Bard E, Reimer RW, Austin WEN, Bronk Ramsey C, Grootes PM, Hughen KA, Kromer B, et al. 2020a. Marine20—the marine radiocarbon age calibration curve (0–55,000 cal BP). *Radiocarbon* 62(4):779–820. doi: [10.1017/RDC.2020.68](https://doi.org/10.1017/RDC.2020.68)
- Heaton TJ, Blaauw M, Blackwell PG, Bronk Ramsey C, Reimer PJ, Scott ME. 2020b. The IntCal20 approach to radiocarbon calibration curve construction: a new methodology using Bayesian splines and errors-in-variables. *Radiocarbon* 62(4):821–863. doi: [10.1017/RDC.2020.46](https://doi.org/10.1017/RDC.2020.46)
- Higham DJ, Higham NJ. 2016. *MATLAB guide* (Vol. 150). Siam.
- Hughen K, Heaton T. 2020. Updated Cariaco Basin ¹⁴C calibration dataset from 0–60 cal kyr BP. *Radiocarbon* 62(4):1001–1043. doi: [10.1017/RDC.2020.53](https://doi.org/10.1017/RDC.2020.53)
- Hogg A, Heaton T, Hua Q, Palmer J, Turney C, Southon J, Bayliss A, Blackwell P, Boswijk G, Bronk Ramsey C, et al. 2020. SHCal20 Southern Hemisphere calibration, 0–55,000 years cal BP. *Radiocarbon* 62(4):759–778. doi: [10.1017/RDC.2020.59](https://doi.org/10.1017/RDC.2020.59)
- Hua Q, Turnbull JC, Santos GM, Rakowski AZ, Ancapichún S, De Pol-Holz R, Hammer S, Lehman SJ, Levin I, Miller JB, et al. 2022. Atmospheric radiocarbon for the period 1950–2019. *Radiocarbon* 64(4):723–745. doi: [10.1017/rdc.2021.95](https://doi.org/10.1017/rdc.2021.95)
- NOAA NECL. 2022. International Tree Ring Data Bank (ITRDB) Online Resource: <https://www.ncel.noaa.gov/products/paleoclimatology/tree-ring>
- R Core Team. 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Reimer PJ, Hughen KA, Guilderson TP, McCormac G, Baillie MGL, Bard E, Barratt P, Beck JW, Buck CE, Damon PE. 2002. Preliminary report of the first workshop of the IntCal04 radiocarbon calibration/comparison working group. *Radiocarbon* 44(3):653–661. doi: [10.1017/s0033822200032100](https://doi.org/10.1017/s0033822200032100)
- Reimer PJ, Bard E, Bayliss A, Beck WJ, Blackwell PG, Bronk Ramsey C, Brown DM, Buck CE, Edwards LR, Friedrich M, et al. 2013. Selection and treatment of data for radiocarbon calibration: an update to the international calibration (IntCal) criteria. *Radiocarbon* 55(4):1923–1945. doi: [10.2458/azu_js_rc.55.16955](https://doi.org/10.2458/azu_js_rc.55.16955)
- Reimer PJ, Austin WEN, Bard E, Bayliss A, Blackwell PG, Ramsey CB, Butzin M, Cheng H, Edwards RL, Friedrich M, et al. 2020. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* 62(4):725–757. doi: [10.1017/RDC.2020.41](https://doi.org/10.1017/RDC.2020.41)
- Schulman E. 1956. *Dendroclimatic changes in semiarid America*. Tucson (AZ): University of Arizona Press.
- Van Rossum G, Drake FL. 2009. *Python 3 reference manual*. Scotts Valley (CA): CreateSpace.