**Student Guide: How to Collect Radiocarbon Dates**

**PART I: Collection**

Student assistants: First, you need to know a few things about radiocarbon, or 14C dates. Radiocarbon dating was developed in 1950 by the chemist Willard Libby. Put simply, radiocarbon dating works by calculating the amount of 14C (carbon-14) in a sample relative to other isotopes of carbon, 12C and 13C. Anything organic can be dated with radiocarbon dating, although some materials are better than others; this will become important below. While alive, plants take in 14C (along with two 13C, and 12C) through respiration, and animals by eating plants and other animals. When the organism dies, it stops taking in 14C, which, since it’s radioactive, begins to decay. The amount of 14C can then be converted into an age through a mathematical formula. Dates always come as a mean age with a statistical standard error, e.g., 1010 BP +/- 60 (columns R and T on the CARD sheet call this the date’s “sigma”). BP is “before present,” and since the present is always moving ahead, researchers fixed it at AD 1950.

There are two basic kinds of radiocarbon dates: **Radiometric** and **AMS** (accelerator mass spectrometry); there is also **radiometric-plus**, a radiometric date with a longer counting time, usually done for small samples. Radiometric and AMS techniques calculate dates in different ways: for radiometric dates, and put very simply, a purified sample is placed into a Geiger counter which then counts beta emissions – those are the particles that are released as 14C “decays” into a stable isotope. In AMS dating, the sample’s 14C, 13C, and 12C atoms are counted through a accelerator mass spectrometer (often the 13C/12C ratio is, and should be, calculated on a regular mass spectrometer). AMS dating is usually more accurate (though tests show Beta Analytic’s radiometric and AMS dates are very close); its real advantages are that it produces a smaller standard error and can be used on very small samples (allowing us to date, e.g., individual seed). Some labs only run AMS dates (and in my opinion they are the only dates archaeologists should be using today).

Radiocarbon decays with a half-life of 5730 years. By half-life we mean that however much 14C an organism contained when it died, only half of that amount remains after 5730 years; and half of that half remains after another 5730 years and so on . . . until the amount because too small for us to measure (at about 45,000 years of age). However, Libby started the method with 5568 years as the half life. Although researchers knew that the half-life was 5730 years by the early 1960s, they elected to keep using 5568 years to keep everything in the same system and in anticipation that the half-life would be further refined (it has not). This is not a big problem, because the process of calibration (see below) takes this fact into account. HOWEVER, some researchers decided to “go rogue” and use the 5730 year half life. This does not create a problem if a report tells us that the 5730 yr half life was used (we can convert the date to a 5568 yr half life by dividing it by 1.03). Unfortunately, reports often don’t tell us which half-life was used. We assume it was 5568 (which is actually a fairly safe assumption).

* Note: WSU may have been using the 5570 year half-life for a time; we treat WSU dates as using the 5568 yr half-life unless it is reported otherwise; the error is miniscule.

In addition, we divide radiometric dates into **measured** and **normalized** (or conventional). The measured age is the age of the sample derived from the radiometric counter. But organic materials take in 14C in different ways that results in more or less of it being retained by the organism. For example, corn and oak trees have different photosynthetic pathways (known as C4 and C3, respectively, there is a third pathway as well, CAM). The result of these different pathways is that C4 plants retain more 14C relative to an oak tree, with a C3 pathway, and that means a maize plant that dies the same year as an acorn will appear to be younger – because it began with more 14C than the oak’s acorn. Fortunately, we can compensate for this with the aid of the sample’s **δ13C** (“delta 13-C) value (column U), a measure of a sample’s ratio of 13C/12C. δ13C values are measured as parts per mill (‰) always negative (because they are deviations from a geologic standard), and normally run between -27 and -8, but they can go much higher and lower. The δ13Cvalue allows us to convert a measured age into a normalized age. Such conversion is automatic for AMS dates, hence, all AMS dates are normalized ages. However, not all radiometric dates are normalized. In fact, many dates generated between 1950 and 2000 are measured ages. Not all labs made the conversion because they were not equipped with the necessary equipment (which became more widely available after 1980). By the 1990s, many labs would make the conversion, but they charged extra for it (and since archaeologists are stingy, many did not use that service). Today, such conversion is standard, although sometimes a “standard” δ13C value is used (e.g., -25.0 ‰ for charcoal samples – just hope they are not maize because maize has δ13C values around -10.0 and they alter the measured age significantly). δ13C values that are more negative than -25 (e.g., -27) make a date younger, and those less negative (e.g., -12) make a date older. The age increases or decreases at ~16 years/1‰.

The difference between measured and normalized ages matters because of calibration. Calibration takes into account the fact that the amount of 14C in the atmosphere has fluctuated over time (Libby provisionally assumed it was constant). During times where there was less atmospheric 14C than at other times means that plants that grew during those times had less 14C to incorporate, and thus they appear to be older than they actually are. The good news is we can fix this through “calibration” – *but the calibration process assumes that the date has been normalized* – corrected for different photosynthetic pathways (it also corrects for the fact that the half-life is 5730, not 5568 yrs, which is why we can continue to use the “wrong” half-life).

Here’s a few tips for dealing with measured ages:

If a source does not tell you if a sample is measured or normalized, then it is probably measured (this is especially true for older reports because converting the sample was not even done in the early days of 14C dating, and it’s only been done routinely since the late 1990s). However, if a report provides a δ13Cvalue, then the date is probably normalized. That’s the only reason for providing the δ13Cvalue. There is a column in the spreadsheet for δ13Cvalues (Column U).

There are many 14C labs around the world; many of these are now defunct (e.g., Washington State University, University of Michigan) and the defunct ones did not frequently normalize dates. For some we can figure out if the date is measured or normalized:

* Beta dates with lab numbers > 70783 are assumed to be normalized whether they have a delta 13C value or not; Beta tells us that after Beta-70783 all samples were normalized. So, if a sample has a Beta number > 70783, it can go in the normalized column even if there is no associated delta 13C value. This is good because Beta-Analytic (aka “Beta”) produces many of the dates in North America. Note also that the correct abbreviation for Beta-Analytic is “Beta” not simply “B” (which is actually the abbreviation for the Bern lab in Switzerland; other reports use “β”, which is also incorrect).
* All University of Georgia lab dates (UGa-) >5470 are normalized. (Note: Do not confuse these with UGAMS- dates (University of Georgia AMS) – as AMS dates, they are all normalized.)
* All Geochron (GX-) dates >GX-8633 can be considered normalized; Geochron was actually among the first labs to acquire their own mass spectrometer.
* All ISGS dates are normalized (Hong Wang, email of January 2017)
* All AMS dates are normalized, even if there is no δ13C; this includes
	+ AA dates (Arizona Accelerator)
	+ EZV dates (Note: EZV dates are very rare– I’ve only seen them in California).
	+ UCIAMS
	+ CAMS
	+ D-AMS (Direct AMS)
	+ UGAMS (U Georgia)
	+ PRI (PaleoResarch, Inc.)
	+ ETH
	+ NSRL (INSTAAR AMS lab, U. Boulder)
	+ OxA (Oxford Accelerator)
	+ Beta often ran AMS dates through ETH so they might have a lab number such as” Beta-43055/ETH-7729; they also ran dates through other AMS labs, e.g., Beta-XXXX/CAMS-YYYY
	+ Beta switched to running all AMS dates about mid-2012; however, we have seen Beta lab sheets with RadiometricPlus dates after this. Looking at the lab numbers, we assume any Beta date >400000 is an AMS date unless it is listed as Radiometric – in which case it is almost certainly RadiometricPlus (or an error!).
	+ UCR ran AMS dates – these are usually labeled with two numbers, UCR-XXX/CAMS-YYY.
	+ PSUAMS
	+ ICA
	+ MAMS
	+ NZA (Rafter lab, New Zealand)
	+ ISGS (Illinois Geological Survey) ran AMS dates, indicated by an “A” prefix to the number: ISGS-Axxxx.
	+ Any date otherwise categorized as AMS (there are a few other, rarely used, AMS labs) in the “Additional Information” column are normalized dates even if there is no δ13C.
* Dates from older labs are assumed to be measured ages, unless a δ13C is present. This includes:
	+ M (Michigan)
	+ WSU (Washington State University, which opened in 1962; WSU rarely did the correction as it did not have a mass spectrometer on campus (April 7, 1983 letter to Stuiver; a 1982 document states that “samples requiring 13C determinations are sent elsewhere” but that most samples don’t need correction – i.e., they are charcoal with an assumed δ13C of -25, which has no effect on the mean measured age. Last WSU sample is WSU-5058, run on January 7, 1999. δ13C correction was an add-on to the end.).
	+ SMU (Southern Methodist University)
	+ TX (U Texas) (note: we have seen dates in a partial TX lab database, that were not reported in the publication with δ13C values when in fact they had those values in the lab database – those were published as measured ages. We have the δ13C values for many (though not all) TX dates and Kelly will make the correction. The Texas lab closed in 1997.
	+ WIS (Wisconsin) dates are measured; no corrections were ever done. I’ve been in touch with James Stoltman, but there is little information available about this lab.
	+ GaK (Gakushuin, Japan) dates are mostly measured ages. The lab did not use a delta 13C correction until 1998; the lab shut down in 2001. So, most GaK dates are measured. δ13C corrections were made beginning with Gak-20067. Any GaK lab number < 20067 is measured (Postscript: our largest GaK date is 9719 so all our GaK dates are measured ages.)
	+ I dates (Teledyne Isotopes)
	+ L dates (Lamont-Doherty)
	+ UM (Univ. of Miami)
	+ FSU (Florida State): Note: FSU and UM, both defunct, were started by Jerry Stipp (who after Miami closed started Beta Analytic with Murray Tamers in 1979) FSU: Began in 1965. Professor A. Leroy Odom at FSU was told that the FSU dates run after Jerry Stipp left were “questionable” and then department head George DeVore closed the lab and sent the equipment to Stipp in Miami.
	+ C (Chicago, the old Libby lab)
	+ RL (Radiocarbon, Ltd)
	+ SI (Smithsonian)
	+ SL (Sharp Labs)
	+ UCLA
	+ W (USGS, National Center)
	+ UW (Univ. Washington)
	+ QL
	+ Other rare lab numbers (e.g., O – Humble Oil Co.; PL – Purdue; SM – Mobile Oil Corp.; TBN [or TBNC] — Kaman Instruments; TO – Isotrace Lab) are almost certainly measured ages.

The data table we use is taken from CARD (Canadian Archaeological Radiocarbon Database), which is where all our data will eventually reside. We aim to collect data that future researchers will need in order to decide to include a date in their analysis or not. We can’t anticipate everything, but some metadata are (a) hard to get and (b) not very useful in deciding whether to use a date or not. In those cases, I’d rather not spend time tracking them down.

In addition:

1. I created data files for each state based on the CARD template; the “Additional Information” is used to note if dates are AMS, Radiometric or RadiometricPlus (aka, extended count). It is important to at least note which are AMS dates (the others are assumed to be radiometric, but we try to label them as such if they are labeled as such in a report; Beta ran AMS dates early in its history – although these are often listed as Beta-xxxx/ETH-yyyy, authors did not always add the AMS lab that actually ran the sample).
2. Note that the CARD template uses dropdown menus for some columns (e.g., Material). If you put something in the wrong column and copy-and-paste to the right one, the dropdown menu will be moved to the cells that you pasted data into; likewise, the reverse is also true. We should try to avoid this.
3. The same goes for some of the data validation codes used in some columns, e.g., the “sigma” columns won’t accept “0” as a valid entry. Copy and paste and you move those validations for the affected cells.

Spreadsheets Data columns:

Note that the Green-highlighted column labels are ones that CARD considers mandatory. Some of the other columns are not important and we shouldn’t waste time on them; bold-faced ones below are the most important.

1. **lab number**: important; some are blank or ?; we need to try to clear them up – it is not always possible. We always write lab numbers as the “lab code hyphen number,” e.g., Beta-24135, or GX-6670, or D-AMS-6509, etc. Some sources will leave out the hyphen, but we always put it in so that the duplicate search mechanism (see below) works properly (because to a computer Beta-24125 is not a duplicate of Beta 24125). For the same reason, if there is a modification to a date, e.g., Beta-24125A, we add the “A” without a hyphen. This is important: if dates are not put in correctly you will think you have no duplicates when, in fact, you do. We do not use leading zeros on lab numbers as I find it makes them hard to read when running through a long list. This does mean that Excel sorts the lab numbers “alphabetically” rather than numerically.
	1. Use the correct lab number (see Radiocarbon’s annual listing). For example, Beta Analytic is “Beta-,” but frequently appears as “B” (which is the Bern lab’s designation) and “𝛽.” The Lawrence Livermore National Lab is not LLNL but CAMS, and the University of California-Irvine Keck lab is not KECK but UCIAMS. Woods Hole website states its prefix is OS- but it has appeared in Radiocarbon’s lists as NOSAMS and WHAMS (we change these to OS). Likewise, the INSTAAR lab in Boulder uses CURL, an in house tracking number, and NSRL; dates should appear as NSRL (unfortunately, the CURL and NSRL numbers are different, so we have to leave them as we find them in reports).

B. Field number (occasionally given in reports)

**C. Material dated**: this is a drop-down menu, so make the appropriate choice; choose “unknown” if the material is not given

**D. Taxa:** if you can provide specifics here, do so, e.g., if the Material Dated is charcoal and the source indicates it was maize, or some other specific plant, add that here; likewise, if the Material Dated was bone, give the animal (“Homo sapiens” for human) if the source provides it, and whether it was from collagen or apatite (those are two different fractions of bone that are dated separately). For bone, also indicate the sample preparation if given, e.g., ultra-filtration, XAD.

**E. Type of date**: another drop-down menu, choices are cultural, paleontological, or geological. Most of ours should be “archaeological”; any date on a Pleistocene animal should be “Paleontological” and any dates on sediments should be “Geological” unless something in the context indicates otherwise (e.g., there are sometimes sediment dates on hearth sediment – that would be archaeological).

**F-J Locational Data**:

* F. Locality: usually this is just the county; we can fill these in using the site number, since county is encoded there (a few states don’t use the Smithsonian site number that many of you may have seen elsewhere, e.g., 48BH3457 – the BH standing for Bighorn County). Arizona is a special case; Kelly will handle these.
* **G, H and I**: locational data is decimal latitude/longitude and elevation (in meters!). To avoid breaking the law (ARPA), we are using county centroids as location. Kelly will add these later.
* Arizona relies on 1 x 1 degree map blocks, labeled A through FF, and divided into 16 numbered blocks with sites numbered sequentially therein. We treat these blocks as counties.

K – P Provenance (not to be confused with “provenience” – provenance is similar to “chain of evidence”): these six columns are less important, although we can add a year (2016 or 2017) under the “date updated” column for any date that you add or modify. If the data are easily available, by all means include them; otherwise, don’t waste time digging for them.

**Q – V Age**: Obviously, the age columns are important; so is the delta 13C column. Note that most dates are calculated with the “Libby half-life” of 5568 +/- 30 years. This is actually incorrect – the half-life is 5730 +/- 40 years (known as the Cambridge half-life). However, the field elected to continue using the Libby half-life since many dates had already been generated using it, and since the field assumed the half-life would change again in the future (it hasn’t). This is not a large problem as the mean age of a Cambridge half-life date can be converted to a Libby half-life date by dividing by 1.03. We make that conversion where a report tells us that the date is reported using the Cambridge half-life; virtually all labs use the Libby half-life, but some researchers in the past altered the date to the Cambridge half-life. This is not necessary today with calibration. Unless an author tells us otherwise, we assume the date uses the Libby half-life and can be reported as written.

* + See above comments about measured and normalized ages; we put the date’s mean and sigma (standard error) in separate columns. Some sources will provide the measured and normalized ages – if so, record them both.
	+ Column U, δ13C: see comments above; if given record it; it won’t be there often for dates obtained prior to the late 1990s.
	+ The δ13C *source* column (V) is not important and usually hard to get. Sometimes, these values are estimated and this column is there to record that. However, I’ve found that it is rare for anyone to report how the δ13C value was obtained; and if it is -25 or -10, it is almost certainly an estimate (used to convert a measured to a normalized age, for unidentified charcoal or maize, respectively; also, normalization only affects the mean, not the standard deviation).
	+ Early (pre-1963) Michigan lab dates sometimes appear in one publication with standard errors twice that which appears in another publication. Bill Lovis cleared this mystery up: Michigan was the first lab after Libby’s Chicago lab, and there were no rules back then. So, the early dates were published with errors that were twice the standard error plus a little extra to compensate for assumed contamination (we suspect they rounded up to the next 50 years). This primarily affects Michigan dates whose lab numbers are < M-1000; we have halved the errors so they are comparable to other dates in the file.
	+ Note in the Comments if the method used was the initial “solid carbon” method (these dates are not considered to be trustworthy)
	+ Earlier dates were frequently reported as AD/BC; if we can determine whether this conversion from radiocarbon to calendar years was a simple matter of the difference between the date and AD 1950, we convert the age and record it under “Measured Age”. If not, we cannot record the date.
	+ Some Texas (Tx) lab dates (primarily those with numbers > Tx-6200) had δ13C values in the partial database we obtained but the lab did not normalize those ages, leaving that to investigators (who apparently rarely made the conversion). Consequently, we normalized those dates assuming 1‰ = 16 years.
* **W – AB Provenience**: some of these are important:
	+ Significance refers to the site’s phase or culture – I don’t regard this as important, and it’s sometimes hard to get; if it is easy to get, put it in; if not, move on – it can be quite controversial, anyway.
* **Site identifier**: this is the site’s official state number; it’s important because it provides a site’s county, and that is used as the site’s “location.” Also, it is crucial in sorting out duplicates, especially where one is lacking a lab number. Normally, this is the Smithsonian trinomial (e.g., 48BH1234), but some states (Connecticut, Delaware, Maine, New York, and Rhode Island use different systems, or no discernible system. Arizona relies on 1 x 1 degree map blocks, labeled A through FF, and divided into 16 numbered blocks with sites numbered sequentially therein. We treat these blocks as counties. Unfortunatel, site numbers are assigned independently by ASU and ASM, and by some agencies, so there can be, e.g., AZ K:14:123 (ASM) and AZ K:14:1234 (ASU), and sometimes practitioners leave off the suffix (we drop the AZ in our data – Arizona extended the system to neighboring states and Mexico). The other system uses sequential numbers assigned by the Museum of Northern Arizona, NA-xxxx; where possible, we assign these to the first systems’ map blocks to give the sites a “county” location.
	+ **Site Name**: record if you can (not all sites have names); it helps to sort out duplicate values and – this is important – if the site is cave, shelter or rockshelter, include one of those terms in the site name, e.g., Danger Cave, Gatecliff Rockshelter, Triple-T Shelter. This helps researchers pull dates only from “closed” sites, as opposed to “open” sites, e.g., on a valley floor or floodplain.
	+ Stratigraphic component: record if you can, but it is rarely easy to get.
	+ **Context:** any information about the date’s context is also important, especially if it comes from a feature such as a hearth; normally, this will be in a reports data table; this will include things like “post in House C” or “feature 1, hearth” etc. Record that if you can. Feature is abbreviated in many ways (F., Fea., Feat.); we’ve chosen to write it out: “Feature”.
	+ Associated Taxa: this column reflects the paleontological interests of CARD’s creator; I ignore it as it would take too long to list it accurately.
* **AC: labeled Additional Information:** has a drop-down menu (AMS, Radiometric, RadiometricPlus – the last of these is rarely used); important, but you can’t always get it; normally, if a date is an AMS date, the report will state this, or the lab number will give it away (see above); still, we list dates as radiometric if we know the lab did not run AMS dates (see above) or if the report tells us a date is radiometric (this is a problem with Beta dates and we cannot assume Beta dates are radiometric or AMS). For dates run prior to the invention of AMS dating (about 1977), they don’t tell you a date is “radiometric” because that was the only kind of dating there was. We’re trying to figure out which labs only ran radiometric dates so that we can use a date’s lab number to categorize it. See comments above about how to use lab/number to determine if a date is AMS. Some newer labs only produce AMS dates (D-AMS, UCIAMS, CAMS, MAMS, UGAMS; and also AA, ICA, OS, ETH, EZV, NSRL, PRI, NZA, and OxA). ISGS labels AMS dates with an A prefix, e.g., ISGS-Axxxx. Beta began processing only AMS dates in 2012; dates with lab numbers >Beta-327191 are probably AMS but as we found a few extended count exceptions, we only assumed Beta dates >Beta-400000 are AMS dates.
* AD: Comments: usually there is nothing to say here, but if a comment is included in a source’s data table, then include it here. Sometimes a date is reported, but the excavator thinks it is “bad” (contaminated, out of context – if the author states that, it can be included here)
* **AE: References**, use at least the last name and year: Smith 1990; or Kelly et al. 1964. Where you can give an abbreviated title, please do so. (The objective is not to give the complete reference, as that is time-consuming, but to give a user enough information to easily Google a reference.) In some cases you might pull dates from a table that the author compiled from other sources. In that case, please enter the reference that you are using, as well as the references cited in the table. It’s important for someone to know, in the case of an error, which source we were using.

I will give you a state data table; this contains dates taken from CARD and perhaps other sources. You will learn a lot about working with spreadsheets in this task. The files are kept on a Dropbox account and are automatically saved as you make changes or add information. (If Dropbox tells you someone else is working on a file when you open it, close the file – Dropbox will create two files, which can then be combined, but it’s an onerous process and we’d rather avoid it.)

Your job is to (a) fill in the holes in the state data table provided to you (using the guide above), check for duplicates and resolve them, and add more dates. I can show you how to check for duplicates (block out column A, the lab number, then click on Condition Formatting under the Home tab, then click on Highlight Cell Rules, then on Duplicate Values, and click OK on the default formatting – duplicate lab numbers will be highlighted in red). You can do this for any column, but we normally work off the lab number to look for duplicates.

I can also show you

* how to convert a PDF into an Excel spreadsheet so that data from tables can be cut and pasted from a source to the appropriate column in the spreadsheet. The conversion sometimes results in an Excel table that has to be cleaned up (sometimes a lot), but eventually this results in fewer errors than re-entering data by hand.
* how to take a hard copy of a table, convert it to a PDF, and then convert it into a spreadsheet (it’s a little clunky and messy, but it will save time, and reduce errors, over entering data by hand).

**Duplicate Entries**

If duplicate entries contain the exact same information, your job is easy: simply delete one of the lines.

If one duplicate contains information that the other does not, simply transfer the data from one record to another, i.e., consolidate all data into one record and then delete the other. The problem arises when duplicates contain conflicting information, because then you have to figure out which one is right. You can start by searching on the web using the lab number, the word “radiocarbon” and maybe the site number or name, or some combination of those. Sometimes you can get to the information on line; sometimes, you can only get to a source that you’ll have to get from the library or interlibrary loan. Generally, we find that if you can’t get the information or a source within 10-15 minutes, then you’re probably not going to get to it (leave it for Kelly or a post-doc to figure it out). And it rarely pays to look at more than 2 pages of Google hits, and mostly only the first page. (Note: the website tDAR, the Digital Archaeological Record, has PDFs of some reports, but it has far more “citation records” i.e., a bibliographic entry, but nothing more, of the reference you might be searching for. Because of this, some students have given up checking tDAR: do not do this, check tDAR. The report might be there!)

Special note: I have found that if two entries contain different dates then only sometimes is the problem a simple recorder error. Other times, the problem is that one source recorded the measured age and the other recorded the normalized (conventional) age. You have to figure out if that is the case for duplicates. This is one reason we like to record measured ages even if a normalized age is available.

**Finding New Dates**

Where can you find new dates? Here in Laramie, we do it through the library and the web. For the library, simply look up your state and archaeology, e.g., “Wisconsin, archaeology.” That will bring up a bunch of sources – and they will generally be located near one another in the stacks – so, pick a call number, go find it, and then look around that number for other books, monographs, etc., and start paging through them for dates.

You can also try googling a state’s name, “archaeology” and “radiocarbon” and see what comes up – you never know. If you find a particular paper or monograph, trying googling it through Google Scholar – it might be available electronically.

We will also go through a state’s local journals and publications; try to find these, but they may be hard to get electronically (World Cat can tell us which institutions hold which journals). If you put together a list or relevant sources. Sometimes you can find these digitally on the web, e.g., I purchased all the back issues to the Wisconsin Archaeologist (some were made available as PDFs, the others were mailed on a flash drive). If you need to spend money, just see me.

The post-docs and I will obtain permission to access SHPO files in those states where such access is possible.

**Dates without lab number**: many dates lack lab numbers in published reports (they have a lab number, the author just didn’t publish it), and these are almost impossible to find (reporting standards were lax in the past, but I have seen post-2000 reports that don’t list lab numbers, or raw dates). One possibility is that they are actually included elsewhere in the table, but since we search for duplicates using the lab numbers, we can’t recognize them. One way around this problem is to sort the dates by age (using first the measured and then the normalized mean date), look for duplicate ages, and then see if the rest of the information for two entries is the same – if so, the line with the unknown lab number can be deleted. I admit that this strategy does not resolve many dates with unknown lab numbers. Note: Be careful when you sort an Excel file – use the sort button with the column header to make sure the entire file is sorted.

**Site numbers/names**: sometimes we find that site numbers get entered without a site name, and vice versa. To check this, simply sort by site numbers. If the same site number has a site name in some cases, but not others, then add the site name; likewise, if some sites with names, lack of number, add the number from an entry that does have one. It is common to find sites with numbers but no names; less common to find sites with names, but no site number.

Please let me know of any other questions or concerns and I’ll update this set of instructions.

**Defunct Lab Record locations**

Gakushuin: I have the spreadsheet of >20,000 dates

WSU: paper records are in boxes at Washington State University, anthropology department

TX: U Texas, Austin, Vertebrate Paleontology Lab, paper (from Ernie Lundelius); we received a partial database

SMU: Director Herb Haas, deceased; tried to locate lab assistant, Tim Dalbey, but have been unsuccessful. (Note: was located by M. Boulanger after end of project; paper files at SMU, DeGolyer Library Special Collections)

WIS: University of Wisconsin (was at Institute for Environmental Studies, but that no longer exists, file location unknown: information from Stoltman).

Michigan: records are archived at Bentley Historical Library, but these dates are fairly well-published so we did not visit the archive.

UCLA: fairly well published, we did not check for records.

SI: fairly well published, we did not check for records.

FSU: dead end

Miami: dead end

Radiocarbon Dates Association, Inc.: We were scheduled to receive cards from Berkeley, but Covid prevented their transmittal until after the project was completed.

There are a few early labs run by individuals, including a 15-yr-old teenage. These include Grey and Schwartzman; they produced very few dates.

Note: active labs cannot share data without PI permission.

**Penultimate Check of 14C State Files**

* Sort database by site number: are there any from other states? This should be clear by the state number, e.g., a site number beginning with 10 (Idaho) doesn’t belong in Wyoming (beginning with 48). If so, check to make sure the date is in the other state file -- if not, move it.
* Sort database by site name: add site number for those entries from the same site that are missing a site number. Note: site names may be misspelled for some entries; or sometimes they are entered with the site number before or after them – they should be removed so they do not affect future site name sorts by users.
	1. Sort database by site number. Our goal is to give site numbers to every entry possible so that locational information (county centroids) can be associated with as many dates as possible.
* Sort database by lab number: check for site numbers written improperly, e.g., Beta-24411 will not be recognized as the same number by Excel if it is written with a space, e.g., Beta- 24411. This can be checked with a global search of the lab number column, converting “ –“ and “- “ to simply “-“.
	1. Likewise, some numbers are entered with leading zeros, e.g., Beta-0024411; these also are not recognized as the same as Beta-24411, so we need to do a global search and replace, converting “-0” or “-00” to simply “-“.
	2. Other errors can creep in, e.g., occasionally there are suffixes to dates, e.g., Beta-24411-A; I’ve been converting these to simply Beta-24411A (no hyphen for the suffix).
	3. Sometimes there is a space at the end of a lab number – Excel will not recognize this as a duplicate. Therefore, do a search for spaces in the lab number column, looking for those at the end of numbers; removing them often results in a duplicate that then has to be resolved.
	4. And then there are some odd lab designations – sometimes these are simply rare labs (see *Radiocarbon*’s official lab code sheet–a PDF is in the dropbox), and sometimes they are codes that an author made up, e.g., L-n/a-UT1 (not knowing that there are actually official lab designations); leave these for me to resolve: they are rare.
* Sort database by “Taxa dated” (column D). Move any contextual information from this column to the Context column (AA).
* Sort database by Material Dated (C). We use CARD’s preferred entries here (see below), with additional information going into the Taxa Dated (D) column. In particular, I’ve been finding “Maize” in column C when it should be in column D, and listed as Zea mays; maize and other cultigens should be listed in column C as “Seeds”; if the item is carbonized, add that information to column D, e.g., “Zea mays kernel, charred”.
	1. I’ve also been trying to remove all the French translations; many reflect some older categories that CARD apparently no longer uses, they will also make r-code checking of data difficult.
* Sort database by Material Dated and look at the Type of Date field (E); hide other fields to the right of E (F through Z), so you can look at Context entry. Sediment dates should probably be listed as “geological” unless something in the Context column suggests otherwise; all charcoal dates should probably be listed as “archaeological” unless something in the context says it should be “geological”. Most Pleistocene fauna dates, or things like sloth dung boluses, and everything from packrat middens should be “paleontological.”
* Sort database by lab number (after making fixes as noted above).
	1. Run another check for duplicates and resolve them as needed; if you cannot, leave them for me.
	2. Follow guidelines in “How to Collect Radiocarbon Dates” to check that dates are under the correct Measured or Normalized columns. If a date has a δ13C value, it should be under the Normalized columns (or if the context column states that the date was normalized). See list above for labs whose dates are assumed to be measured only.
	3. Follow the same guidelines to make sure that AMS lab dates are listed as AMS under column AC.
	4. Sort dates listed as “?” under lab number by their age; sometimes there are duplicates but we can’t recognize them as such based on lab numbers, obviously.
		1. Do the same for dates listed as “Beta-?”.
* Scroll through the file to see if information under “Stratigraphic Component” (Z) should be moved over to “Context” (AA).
* Make sure all data are aligned similarly, either right or left (and bottom aligned), depending on data (text should be left-aligned, numbers right-aligned).
* Remove all commas and backslashes from the date columns, e.g., 10,450 should be 10450.
* Remove all quote marks from lab numbers, dates, etc. We don’t know why but these sometimes show up.
* Check spellings of lab dates for consistency, e.g., BETA- should be Beta-.

PART II: FINAL CHECK

1. First do the following as it makes it easier to spot errors:
	1. Make sure all data are aligned similarly:
		1. All data should be bottom aligned
		2. Do not wrap text
		3. All text should be left aligned (except site numbers, right align them; lab numbers are text)
		4. All numbers should be right aligned (including Date Updated): easiest to do this and the above by blocking out the relevant columns and setting the alignment – but reset any column headings – they should all be left aligned.
	2. Remove all commas and backlashes from date columns (e.g., 10,450 should be 10450): do this by searching for “,” and “/”
	3. All sheets should be in Calibri size 12 font (block out data rows, then check typeface and font at upper left while in the Home tab)
	4. Make sure that dates and 13C (columns Q-U) are read as numbers by Excel, not text. If there is a tiny green triangle in the upper left corner, then Excel is reading it as text. To fix it all simply block out all the data lines in columns Q-U, then click on the tiny exclamation point that will appear to the left of the blocked entries, click on it, then click on “convert to number”. If no exclamation point appears, then all entries are already numbers.
2. Clear any duplicate Conditional Formatting rules to ensure all ID duplicates are highlighted
	1. Highlight column A and open ‘Conditional Formatting’ (under the Home tab) > ’Managing Rules…’
	2. If there are multiple rules present delete the duplicates. Only one rule should remain which says ‘=$A:$A’ under ‘Applies to’; if that rule is not there, just delete them all and reset the rule under ‘Conditional Formatting’ > Highlight cell rules > Duplicate Values.
3. Checking Lab Numbers
	1. Sort by Column A (remember – sorting is easiest to do by highlighting row 4, with column headings, then click Data tab, then Filter. With this, you can click on the arrow that appears in each column heading to sort the entire spreadsheet by the values in that column).
	2. Search Column A for any spaces and remove them (this is a real problem as it prevents the program from detecting duplicate dates, e.g., AA-4331 is not read as a duplicate of AA-4331 ; i.e. AA-4331 with a space after it).
	3. Search for ‘-0’ and removed any zeros after the dash (but not PRI dates, e.g, PRI-09 since that refers to a year)
	4. Remove any suffix dashes (e.g., Beta-2411-A becomes Beta-2411A).
	5. Check that each lab number prefix is spelled correctly. Misspelled entries will appear at the beginning or end of each lab’s section.
		1. Also make sure labs are consistently presented (BETA- should be Beta; Uga should be UGa); you can do this most easily by blocking out the relevant portion, then use the search and replace button.
	6. Check that there are no prefixes that do not exist
		1. E.g LLNL (Lawrence Livermore) is ‘CAMS’; Keck should be UCIAMS.
		2. If unsure check the Radiocarbon Laboratory abbreviation list (<http://radiocarbon.webhost.uits.arizona.edu/node/11>)
		3. NOSAMS and WHAMS should be OS (there is some confusion in the field over which prefix to use – Woods Hole has OS on their website so we use that).
	7. Sometimes there are two lab numbers, e.g., UCIAMS-127308/SR-8253; you need to check if these appear independent of one another elsewhere. Check the immediate rows above and below, and run a search for the second ID. If it is present, resolve the duplicates (keeping the entry with two lab numbers)
		1. Dates with two lab numbers should appear in column A separated by a “/”, as above.
		2. CWRU- dates (only ~100 were ever created) were produced by the DIC- lab, and sometimes they are reported as CWRU dates and sometimes as DIC dates; fortunately, they use the same number; only the prefixes are different. So, check for CWRU dates and see if there is a DIC- date with the same number; if so, treat them as duplicates and resolve as always (most likely you can simply delete the CWRU date). If there is no DIC duplicate, then change the CWRU lab number to DIC-xyz/CWRU-xyz; the number following the prefix can be the same.
4. Radiometric vs AMS
	1. While sheet is sorted by Column A check Column AC for AMS and radiocarbon designations; easiest to hide the columns between column A and AC temporarily.
		1. The following lab prefixes are all AMS: AA, ACRF, CAMS, D-AMS, ETH (sometimes with a Beta- number as well), ISGS (with ‘A’ suffix only, as in ISGS-A4451), ICA, NSRL, NZA, OZA, OxA, PRI, UCIAMS, UCR/CAMS, UGAMS, OS, CAIS (which are Georgia dates, but they apparently used CAIS instead of UGAMS for a while)
		2. Some labs (e.g., Beta, GX) are not distinguishable as AMS or radiometric. If AMS/radiometric was not specified during data collection leave them blank. The default assumption for old labs, however, is radiometric, including M-, UM-, FSU-, WSU-, GaK-, I-, DIC-, W-, S-, SI- and others.
		3. Beta switched to all AMS between 2010 and 2012; we have a sample submitted to Beta in July 2012, with lab number Beta-327191; however, they apparently still ran some RadiometricPlus dates, until at least 2013 (Beta-346686; these Beta numbers are from Beta lab sheets). Beta could not tell us when they switched entirely to AMS; To err on the side of caution, list all Beta dates with lab numbers >400000 as AMS unless already entered as RadiometricPlus.
5. Measured vs Normalized (Conventional)
	1. While the sheet is sorted by Column A check measured vs normalized dates (hide columns B-P)
	2. Assume the following prefixes are measured UNLESS there is a delta 13 C value (or unless a report says the date has been corrected; this might show up in the comments column): Beta (<70783), C, FSU, GaK, GX (unless it is listed as AMS), I, L, M, O, PL, QL, RL, SI, SL, SMU, TBN (TBNC), TO, Tx, UCLA, UCR, UM, UW, W, WIS, WSU
	3. All AMS dates should be listed in the normalized column
	4. In addition to the AMS labs, the following prefixes can be assumed to be normalized: Beta >70782, UGa >5470, ISGS, SR, GX>8633.
	5. Any dates with δ13C values should be put in the normalized column
	6. If “modern” or “?” is listed as the date in the measured and/or normalized columns, delete it. For modern dates put “date is modern” in the comments column. A “?” in the Measured or Normalized column probably means a modern date, so treat it similarly but list it as “date is modern?” in Comments column.
6. Check for missing site numbers (this takes time as it requires visually inspecting the entire list)
	1. Sort by ‘Site name’
	2. Add site number for entries from the same site but where only one or a few have a site number
	3. Make sure site names are consistent for the same site numbers (if you are unsure, ask us)
	4. Then sort by “Site Identifier” and check to see that a site identifier is not missing a name (plenty will have no site names, leave them as is, but if there are multiple dates from a site and most of those site identifiers have a site name, provide the site name for those similar site numbers that lack a name)
	5. Check that there are no site numbers from other states (if so, cut and paste those entries into the correct state – if they are duplicates, resolve them as before)
7. Check for unidentified duplicates (e.g., states with “?” in ID numbers, typos)
	1. Sort by ‘Measured Age’, and then ‘Site Identifier’ (then hide columns B to P)
	2. Scan the Lab Number column for “?” or any lab number with a “?” attached and ensure they are different from other dates found on that site.
	3. Then scan each date by site and see if there are any duplicate dates and errors. One way to do this is to highlight the Measured Date column and check for duplicates. There will be many, of course, but scan down the file, looking for duplicates that are next one above the other – these might be duplicate dates. After the next step, do the same for the normalized dates (i.e., Sort by ‘Normalized Age’, and then ‘Site Identifier’ and proceed to the following step)
	4. If there are duplicates check that they are actually two separate dates using the context information and the Lab Number. If year, standard error, and context information is the same and the Lab Number has a “?” it is safe to assume that it is a duplicate. If the Lab Numbers are different check, see if the difference can be explained by a simple typo (e.g., missing number, number off by one digit, transposed digits, an errant space). Keep the ID number that is in sequence with other IDs from the site or, if there are no other numbers, check the original source if possible (if not, add a “?” to the possibly problematic date lab number).
8. Fill in county name ‘Localities’ column
	1. Sort by site number and use them to identify counties (hide columns G-W)
		1. Note: County names and abbreviations are on the “Instructions” tab at lower left of spreadsheet; there are occasional errors, bring problem children to us.
	2. Add county to “Localities” column where needed
		1. But if there is already different text in the locality entry, place the county name in front of it so we can later use the locality column to sort by county.
		2. Don’t worry about geological/paleontological entries – if we can get the county, great; if not, let it go
		3. Kelly will fill in county centroids in the latitude/longitude columns later from another file.
9. Sort ‘material’ column
	1. Make sure every entry has a value filled in. Change blanks or “?” to ‘Unknown’
	2. Check that only CARD preferred categories are used.
		1. Any French translations can be removed
		2. If the material is “charcoal” and “charcoal” or “charred material” is also present in the Taxa Dated column, delete the text in the Taxa Dated column; no need to repeat. But “charred wood” (or anything else charred) can stay (as it implies the material was not entirely carbonized)
		3. Any materials listed that is different from the CARD categories can be moved to the ‘taxa dated’ column – replace with the appropriate CARD category text in the Material column; e.g., if it says “wood, charred” in Material Dated, Change it to Wood, and put “charred” in Taxa Dated column
		4. Make sure spelling is correct and consistent throughout the column
			1. Any misspelled words should appear at the beginning or end of sorted material type
	3. Make sure all maize dates are listed as ‘Seeds;’ if burned then “charred” can be added to “Taxa Dated” column description. All maize dates should have ‘Zea mays’ in column D (can be accompanied by “kernel,” “cob,” etc. You can do this by searching for “maize” and “corn” – changing these to Zea mays, then search for Zea mays (or sort by Taxa Dated – but that won’t catch everything if Zea mays is not the first words in the entry); check to make sure all Zea mays is listed as “Seeds” in Material Dated (after changing corn and maize to Zea mays, sort by the Taxa Dated column and check the Zea mays entries)
	4. Likewise, highlight column D and do a global search and replace “human” with “Homo sapiens”
	5. Make sure all nut/nutshells (or specific plants: e.g., acorns, hickory nuts, walnuts) are listed as ‘Seeds’ – put nut/nutshell in Taxa Dated column (where they should be now); again you can do this with a search and sort
10. ‘Type of Date’
	1. Leaving the columns sorted by materials check each ‘type of date’ (hide columns F to

P)

* 1. Check the context entry to identify if each date is ‘Archaeological’, ‘Geological’, or ‘Paleontological.’ Leave blank if you can’t make a decision
	2. If the date is really old (>14,000) then it should probably be geological or paleontological depending on the material dated; I admit there is a controversial grey area here for dates >~13,500.
	3. Most charcoal dates will be archaeological unless the context suggests it is a stratum date (in which case it should be Geological)
	4. Anything listed in Taxa Dated as human bone, Zea Mays, an artifact (e.g., atlatl, fish weir stake), or human feces should be “Archaeological”
	5. Anything listed in Context column as coming from a feature should be “Archaeological”
	6. Anything listed in Material Dated column as Textile, Leather or Pottery Residue should be “Archaeological”
	7. Pleistocene fauna (unless it is clearly culturally associated) should be listed as ‘Paleontological’
	8. Check that all packrat (search for packrat, pack rat, or Neotoma) midden dates are listed as ‘Paleontological’ in ‘Type’ column
	9. “Sediment” dates (under Material Dated) should probably be listed as ‘geological;’ this is mostly, but not universally true so check the context column
	10. If you are really uncertain, leave the entry blank; we’ll deal with them
1. Final formatting
	1. In the ‘Updater’ column list the ‘Wyoming Team’ and fill in the ‘Date Updated’ column with 12/31/2020; do this for the first entry, then do a blanket copy and paste
	2. Make sure that in the ‘Context’ column that the word ‘Feature’ is spelled out (e.g., no F, F., Fea, Fea., Feat., or Feat); you can search for those abbreviations and replace them (using search and replace), and search for F1, F2, F3…F9, replacing them with “Feature 1”, “Feature 2” etc. Be careful – because sometimes F might refer to a unit or strata. Make sure you highlight the Context column before you do this! Even then, don’t do a blanket replacement – go through hit by hit.
	3. There should be vertical lines between columns: E:F, J:K, P:Q, V:W, AB:AC, and AE:AF
	4. Move any current lat long designations into two columns after the references column (we will remove these before sending to CARD; we have such data for very few sites and some of it is inaccurate)
2. Sort the database by ‘locality’ or ‘site number’
	1. Copy and paste the county centroid decimal lat longs from the master file (on dropbox, filename “County Centroids” – if you are not comfortable doing this, leave it for Bob and post-doc) into the latitude and longitude columns in the state file; be careful, I’ve discovered that the county abbreviations file with the CARD excel sheet (tab marked instructions at lower left) are not always correct. I prefer to use the County Centroids file as a guide – even then, I’ve spotted a few that have me scratching my head).
	2. Repeat for each county for the state
	3. The following states do not use Smithsonian numbers:
		1. Arizona: Map Grid (AMS, ASU) system- Bob has the centroids
		2. Connecticut: By municipality -assign to counties
		3. Delaware: Two letter system- first is County second is a block within that county
		4. Maine: 15 min map grid system- we used reports to assign to counties.
		5. New Mexico: LA sequential numbers- Send to NM SHPO to get county information
		6. New York: multiple systems in use, or none. Link sites to counties as best we can from reports.
		7. Rhode Island: Sequential numbering so we will treat as one county – easy, all sites have the same location – state centroid
3. Column headings have to match those in the CARD template exactly (major offenders: Column AC: AMS or Radiometric should read “Additional Information” and two Location columns should read simply “Latitude” and “Longitude.” Bob uploaded the county centroids and the official CARD excel template to Dropbox. Headings should be left justified.
4. Run r-code check for duplicate lab numbers between states. Resolve duplicates.
5. Run r-code check for incorrect entries in columns