Background

Radiocarbon dating is one of the most important methods for absolute and relative chronological reconstruction of cultural development in prehistoric archaeology (Taylor & Bar-Yosef, 2016). This is true for the intrasite level, for regional comparisons and also in cases, where processes that have a large spatial and temporal reach are to be investigated. Prominent examples for the latter include the ‘neolithization’ of Europe or the ‘Bantu expansion’ in sub-Saharan Africa. They have been extensively analyzed with radiocarbon data (e.g. Ammerman & Cavalli-Sforza (1971), Garcin et al. (2018), Jerardino, Fort, Isern, & Rondelli (2014), Lemmen, Gronenborn, & Wirtz (2011), Oslisly et al. (2013), Pinhasi, Fort, & Ammerman (2005), Russell, Silva, & Steele (2014), Weninger et al. (2009)).

Data selection for such models is complex and requires a thorough understanding of the archaeological questions. Most of the time it is not sufficient to include every date that vaguely fits into the context. Some dates have to be deliberately omitted. In order to make this important process of data filtering as transparent and reproducible as possible, the criteria for selection and especially the original data set must be generally accessible and well contextualized. Otherwise, peers can not evaluate the results in a meaningful way.

Fortunately, many archaeological institutions and individual authors are sharing their radiocarbon collections online (e.g. Hinz et al. (2012), Kneisel, Hinz, & Rinne (2013), Courtney Mustaphi (2016), Seidensticker (2016)) – some of them are archives with a long tradition of quality control and maintenance. Also, the boom of the Open Data movement in recent years has led to an increase of publications with raw data in archaeology (e.g. Palmisano, Bevan, & Shennan (2017)). These collections are an important archive for future research questions.

However, the entire data basis is currently highly decentralized and lacks basic standardization. That results in an effective loss of the possible added value that could result from the intersection of data sets in terms of searchability, error checking and further analysis. The creation of a world-wide and centralised database of radiocarbon dates which could solve these issues is not to be expected for the near future.

Code Summary

c14bazAAR is an R package that attempts to tackle the problem at hand by providing an independent interface to access radiocarbon data and make it available for a reproducible research process: from modelling to publication to scientific discourse. It queries openly available $^{14}$C data archives, but not those behind pay- or login-walls.
The package includes download functions (accessible with the main interface c14bazAAR::get_c14data()) that – first of all – acquire databases from different sources online. They then reduce the tables to a set of common variables and store them in a dedicated R S3 class: c14_date_list. The c14_date_list is based on tibble::tibble to integrate well into the R tidyverse ecosystem. It also establishes standardised data types for the most important variables usually defined to describe radiocarbon data.

Beyond the download functions, c14bazAAR contains a multitude of useful helpers that can be applied to objects of class c14_date_list. These include methods for bulk calibration of radiocarbon dates with the Bchron R package (Haslett & Parnell, 2008), removal of duplicates, estimation of coordinate precision, or conversion to other useful R data types (e.g. sf::sf (Pebesma, 2018)). For the classification of sample material c14bazAAR provides a manually curated reference list that maps the inconsistent attributions in the source databases to a standardized set of material classes. Such a reference list exists as well to fix the country attribution value of dates – which is especially important in case of missing coordinate information. Methods to determine the source country based on coordinates fail on such dates.

c14bazAAR was already used for data acquisition and preparation in at least one research paper: Schmid (2019).

Acknowledgements

The package got valuable code input from several members of the ISAAKiel group (Initiative for Statistical Analysis in Archaeology Kiel), most notably: Daniel Knitter, David Matzig, Wolfgang Hamer, Kay Schmütz and Nils Müller-Scheeßel.

References


