XRONOS: an open infrastructure for archaeological chronometric data

Joe Roe ⬤
University of Bern

Clemens Schmid ⬤
Max Planck Institute for Geoanthropology

Setareh Ebrahimiabareghi ⬤
University of Bern

Caroline Heitz ⬤
University of Bern

Martin Hinz ⬤*
University of Bern
martin.hinz@unibe.ch

ABSTRACT   XRONOS (https://xronos.ch) is an open data infrastructure for the backbone of the archaeological record – chronology. It provides open access to published radiocarbon dates and other chronometric data from any period, anywhere in the world.

KEYWORDS   open data; chronometry; radiocarbon dating; dendrochronology; typological dating; chronological modelling

Introduction

- Chronometric modelling depends on ready access to chronometric data!
- Luckily, we do have a lot of it.
- Moving from 'open data' (uploading a CSV) to 'open infrastructure' (Kintigh 2006)

Chronometric data has been one of the principle domains in which open data and open science approaches have been applied over the last ten years. Comprehensive compilations of radiocarbon dates have become available for many parts of the world in the last decade and, as natural next step, there are now several initiatives to collate this data globally, including the retrieval tool c14bazAAR (Schmid, Seidensticker, and Hinz 2019), the IntChron exchange format (Bronk Ramsey et al. 2019), and the synthetic database p3kc14 (Bird et al. 2022). This move from archiving individual datasets to producing a living open data infrastructure for archaeological chronologies raises a number of challenges. Radiocarbon datasets are still sorely lacking for many parts of the world and, even in those regions with good coverage, the quality of data is highly uneven and largely undocumented. There exists no central repository ensuring the long-term sustainability and completeness of these datasets, and the potential of placing other sources of chronometric information

---

*Corresponding author.

(e.g. dendrochronology, typological dating) in an open data framework has hardly been realised at all.

Here, building on and complementing these initiatives, we present XRONOS: a new web-based platform for chronometric data from archaeological contexts worldwide, combining an open data repository with tools for importing, curating and analysing chronometric information from diverse sources. The principal design goals of the software are to combine all available sources of radiocarbon and other chronometric data in single database; develop robust tools for the continuous ingestion and refinement of this data; and disseminate this data within an open and FAIR framework, embedding it in the wider world of Linked Open Data in archaeology and beyond.

This paper outlines the conceptual and technical infrastructure developed to realise these goals in XRONOS' initial phases of development (2019 and 2021–2024), including a generalised data model for site and radiocarbon information, extendable to other chronometric data; an R- and Ruby-based pipeline for continuous ingestion of data from a variety of sources; continuous, semi-automated data cleaning protocols; a Ruby-on-Rails application providing a web-based frontend to the data and a REST API for programmatic access; and an R package for interfacing with the API. We believe the XRONOS framework provides more open, more reliable, and more comprehensive access to chronometric data than previously available, as well as a foundation for its continuous expansion and refinement.

In this paper we:

- Review the state of the art in radiocarbon data compilation (and other chronometric data, such as they are), highlighting the need for XRONOS
- Describe XRONOS' design goals and software architecture.
- Describe the implementation of some representative features

  - With a focus on the choices we made in representing/modelling chronological data

- Assess the current state of the software and database, and give case studies of how it can be used
- Conclude with our future plans

State of the art

We live in interesting times: Today, archaeological research can no longer complain that too few absolute dates exist for the phenomena of the past. Large-scale dating projects such as [example] and [example] are producing extensive data series in a short time, which make it possible to situate individual sites in time with unprecedented precision. But also due to the dating methods that have become affordable even for smaller budgets in the meantime and the integration of natural scientific dating that is almost mandatory, such large initiatives are accompanied by a great amount of individual and scattered chronological information that represents a rich source of data for supra-regional investigations. Based on that, data modelling can be used as a method to draw conclusions from a variety of individually information-weak data.

The problem, however, is that these data are very scattered. Traditionally, radiocarbon data were disseminated by the laboratories, with special attention paid to publication in the form of "date lists" in the journal *Radiocarbon*. Today, unfortunately, this has become very rare, on the one hand because the quantity and complexity of the data limit the possibilities of publication on paper. On the other

hand, data in printed form are hardly suitable for reuse today. As a result, access to chronological information, especially a comprehensive overview, is made much more difficult (Bronk Ramsey et al. 2019). Furthermore, data is presented in a variety of different formats and granularities, and critical information is often missing, although standards exist (Millard 2014).

Without data repositories, most comprehensive analyses are no longer feasible. So far, there has been no overarching database that constantly collects both 14C and dendrochronological data and makes them accessible in the long term. In the spirit of Open Science, it is the goal of XRONOS to establish such an Open Access database. Our primary ambition is to collect 14C and dendrochronological data for Switzerland and Europe as extensively and comprehensively as possible. However, the project is designed in such a way that an expansion beyond the originally defined spatial focus is already a reality.

Currently, all scientific disciplines face a revolution in the way data is disseminated and used (Bronk Ramsey et al. 2019). The access to these data, and methods integrate individual data into larger scientific questions of social relevance are both are decisive, but not yet sufficiently implemented prerequisites to reach an understanding of the temporal phenomena and dynamics in past societies (Kristiansen 2014). This novel, cross-disciplinary circulation and utilisation makes it possible to deal with completely new scientific questions and makes archaeological data relevant in thoroughly new contexts. For example, analysis of past climatic changes and their long-term effects rely on information from archaeologists to estimate the impact on human and biotic activities (**KintighEtAl2015**). An important source of such information is absolute chronological data. However, the basis of any empirically based modelling is the availability of data.

Compilations of radiocarbon dates

- Where do radiocarbon dates come from?
- Where are they published?

Although 'open data' is often thought to be a new phenomenon in archaeology (**CITE**), it has a long history in radiometric dating. Starting in the 19XXs, radiocarbon laboratories shared and compiled as 'date lists' published on a regular basis in the journal *Radiocarbon* (**CITE**). However, as the number of labs and volume of radiocarbon dates being produced grew, this paper-based format became impractical: the last 'date list' appeared in 19XX, and unfortunately was not replaced by another form of systematic data-sharing or dissemination. Additionally, because the date lists were sourced from radiocarbon laboratories directly—not those who collected the sample—they typically included only very limited contextual information. Thus, even though radiocarbon data comes from a relatively limited number of sources (**IntCalLabList**) and has relatively standardised reporting conventions (Millard 2014), in practice the only way to produce aggregated datasets recently has been to manually search through relevant literature for dates reported by the submitter of the sample. This already laborious process is further hampered by a significant inconsistency in how much authors adhere to measurement reporting conventions (Millard 2014), the lack of any convention on the reporting of contextual information, and weak or nonexistent disciplinary norms regarding the responsibility to publish results openly in a timely fashion.

- Graph of number of dates/databases through time
- Table of databases

Despite these difficulties, there have been many efforts to systematically compile radiocarbon dates in this way over the last few decades (**FIG**).

E.g. RADON: At the time, this was one of the largest 14C data collections for absolute chronological information in Europe. RADON and RADON-B can be considered a complete success, as they have been the basis for several of our own publications (Hinz, 2015; Hinz, Feeser, et al., 2012; Krossa et al., 2017; Weinelt et al., 2015) and as a data repository for high-level publications by other research groups (Bevan et al, 2017, 2019; Dolukhanov and Shrukov, 2004; Henkner et al., 2018; Lemmen and Wirtz, 2014; Manning et al, 2016; Pinhasi et al, 2005; Puchol and García, 2017; Sanders, 2012; Shennan et al, 2013; Timpson et al, 2014), and inspired other projects (Schmid, 2018; IsoMemo; Schmid et al, 2018).

We are aware of XX existing radiocarbon databases or published datasets, comprising an estimated XXXX individual records (**TABLE**) – though this is likely an underestimate. The majority of these resources have a specific regional (XX%) or temporal (XX%) scope. Others are associated with individual radiocarbon labs (e.g. KIK-IRPA, ORAU, BANADORA) and a handful aim at global coverage (XX%, discussed further below).

- How are they published? As databases, SI, etc.?
- How many are open?

While these compilations have greatly advanced access to chronometric data and enabled new modes of analysis (e.g. 'dates as data', **CITE<empty citation>**), there are some obvious limitations to the fragmentary way in which they have been compiled.

- And patchy: highly uneven spatial and geographic coverage, different data collected, different conventions for recording data

- Regional databases are put together *ad hoc* and rarely continuously maintained over long periods

- The coverage of these databases is uneven, systematically biased, and duplicative of each other

- Not all are open

- Not all are machine readable

- Recent moves towards a global synthesis: c14bazAAR (Schmid, Seidensticker, and Hinz 2019), IntChron (Bronk Ramsey et al. 2019), p3k14c (Bird et al. 2022)

  - Each taking a different approach, pros and cons
  - We see a need for something that is comprehensive, unified, living

- And we don't see why it should be limited to radiocarbon: what about other radiometric dates? Dendro? Typologies?

- Some databases, such as KIK-IRPA, ORAU and BANADORA, which are maintained by 14C laboratories, only contain their own measurement data and are therefore of very limited use for overarching research projects.

- Others, such as CONTEXT, RADON, RADON-B, 14SEA and CARD, have a limited regional or temporal focus, and the rich data available e.g. in Switzerland and surrounding countries is not sufficiently represented in these databases. In addition, some of them are project-specific and their continued existence can be described as uncertain due to funding constraints.

- Recently, (Bronk Ramsey et al. 2019) introduced IntChron, a data structure designed to ensure easy availability of radiocarbon data from published articles. We think this approach is very commendable and it inspires us in terms of reproducibility of dating. But we want to go beyond that and put more emphasis on the evaluation and dissemination of relevant information. "IntChron is not intended to be a stand-alone database" (5), but XRONOS is. An overarching database is critical for any research application in terms of data retrievability. Nevertheless, interoperability with other data services, including IntChron in particular, is very important to us.

Beyond radiocarbon
- There are lots of other types of chronometric data!

  – Dendro - in certain places (e.g. Switzerland!), way more important than radiocarbon
  – Other radiometric/absolute dating methods, OSL, etc.
  – Typology - still the backbone of many chronologies
  – Application-specific: sea level dating, rock patina, etc.

- Compilation of other types of dates has been much more limited.

  – Exceptions: …?

For European dendrochronological data, the situation is even less favourable for archaeological purposes:

- NOAA stores some global datasets, but has its main focus in the United States and, with its only 34 European prehistoric datasets, is not useful for the study of the European past.

- The ADS database is compiled and maintained by the UK-based Vernacular Architecture Group and is limited to data published annually in the journal Vernacular Architecture, containing only medieval or later dates for the UK.

- DendroDB has a European focus but only provides data for historical periods.

- The "Dendrochronological Database" of the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL) has a worldwide focus on collected natural wood samples and has different goals than a database for archaeological wood material (furthermore, although announced a long time ago, this database is still not functional).

- Although "Digital Collaboratory for Cultural Dendrochronology (DCCD). An international digital data library for dendrochronology" is well connected in the context of archaeological data services (e.g. ARIADNE), it is focused on the Netherlands (more than 2/3 of the data, while 0.08% of the Quercus data refer to Switzerland) and contains very few prehistoric data (about 2.5%). Furthermore, the overall activity in this database has decreased from 3846 new project records in 2010-2014 to only 83 since then (until the end of 2019).

- Typology
  - perio.do
  - Often embedded in other types of database (e.g. gazetteers)
  - But nobody has treated it as a form of chronometric data in its own right?

What is completely missing worldwide so far, is a database that encompasses both, 14C and dendro data. However, to efficiently conduct data driven modelling of human-environmental dependencies and dynamics on the past, such a data base seems indispensable for future research.

Concept

XRONOS builds on the structure of RADON (Hinz et al. (2012)), one of the earliest comprehensive databases of radiocarbon dates.

Design goals

Our concrete goals in the design and implementation of XRONOS are the following:

1. Provision of access to 14C and dendrochronological data with meta information on archaeological contexts via web-based Open Access frontend
2. Web frontend with processing functions (filters and queries, export to various data formats, 14C calibration & chronology formation) for in place analytical functions
3. machine readable interface — API — for direct access e.g. by open source statistical environments (eg. R) for analytical purposes or dendroanalytical software
4. licensing and versioning of individual data using eg. DOI
5. Storage in a repository with user and role-specific access rights as well as encryption for data privacy
6. Ensuring sustainable data security and usage through compliance with standards, appropriate backup solutions and high-availability technologies

Points 1-3 are now largely implemented and ready to be used. The results of such a search is easily exportable in formats that are simple and generally usable, and represent common, open standards (CSV, JSON). Such formats can be conveniently imported into common evaluation environments (R, Matlab, Python, OxCal, Chronomodel), or even better, directly accessed via the API. This can serve to establish a single gateway to the different sources of information and eliminate the necessity for data publishers to maintain information in multiple formats. With regard to the 14C data, this analytical functions include the live calibration of the individual dates, allowing the most recent calibration curve to be considered. In addition there are extended calibration possibilities like sum, as known from the standard software solutions (e.g. OxCal). Furthermore, the Bayesian calibration in relation to phase models (sequential calibration under consideration of stratigraphy) may be implemented as far as possible, whereby this will take place in close connection with the endeavours of the further development of corresponding R packages. For the dendrochronological data, especially the implementation of visualizations as well as the possibility to create mean curves and to identify impact phases will represent extended possibilities of immediate analysis. This portfolio is to be expanded in consultation with the user group, especially the state (cantonal) dendrochronological laboratories of Switzerland. More elaborated functionality should be made available primarily via an R Package, since this would be too complex for the user to be channelled via an web interface.

An increasing number of research funders legitimately demand that the data that is used in the course of the research they fund become available to other researchers. This disclosure enables new knowledge to be extracted through comparative studies, data mining, etc.; it also enables better transferability and reproducibility of the research process, which increases the quality of research. All data in XRONOS should be uniquely referenceable. Common publication practices, which so far led to ambiguities regarding the data, must be overcome. Therefore, each individual data set should be permanently archived in its current version, and changes resulting from updated information or changed boundary conditions (updated calibration curves) should be provided in a new version with versioned unique identification. In this way, it can be ensured that the exact state of a reference to a date can be unambiguously stored. Only in this way can the research based on these data truly become traceable. One means for this is a Digital object identifier (DOI), which combines a unique and machine-readable as well as persistent referencing with easy retrievability. This functionality is still to be implemented. In order to make such a sustainable data referencing possible, XRONOS strives for a cooperation with longterm data repositories, so that a permanent storage of the relevant information can be guaranteed. This ensures that the collected information will be available to the research and public community beyond the lifetime of XRONOS and thus beyond a possible end of the project.

The ultimate goal of XRONOS is to create an open data structure for the exchange and evaluation of absolute chronological data by the entire research community. However, the data management with regard to dendrochronological data in particular is associated with different interests of the institutions involved, so that compromises have to be made here. For all dendrochronological laboratories control over data disclosure is indispensable. These are mainly publication-related reasons and data is used as "exchange capital" between laboratories. However, in order to provide an incentive for central data management, it must be ensured that users retain full sovereignty over their data. To this end, this project plans to keep the data fully encrypted in such a way that even the database administrators have no insight into the data. For this purpose, methods of asymmetric cryptography (public/private key encryption) based on passwords will be integrated into the database, which make it possible to make data available to individual other users or to the general public.

We hope that the combination of easy and open data access, the provision of analytical functions within the web interface and by means of R packages, the sustainable, transparent and unambiguous data management, the long-term archiving already considered in the initial phase, together with the possibility for individual users to have full sovereignty over their data, will distinguish XRONOS as an outstanding data source and repository, and that we will thereby contribute to the sustainable provision and use of big data in archaeology and advance the use and improvement of standards in this field.

- Inspired by GBIF, which manages to integrate large amounts of data from diverse sources into a signal, consistent database

Data model

TODO: Figure: conceptual scheme, centred on chrons (i.e. an idealised form of the database).

At the base of the XRONOS data model are sets of spatiotemporal coordinates or, as we call them, *chrons*. In an archaeological context, we conceptualise a chron

as an assertion linking human activity with a particular point in space and time. Our data model currently encompasses three types of chron: radiocarbon dates, typological dates (e.g. 'Early Neolithic') and dendrochronological dates. However we anticipate that the concept will accommodate other types of absolute and relative dating techniques, as the scope of the database expands.

Chrons are conceptually useful because they emphasise that different types of archaeological 'dates', drawn from different sources, have essentially the same information content: the location of an event in space and time. We thereby avoid privileging certain sources of chronological data over (as might be the case if, for example, we treated 'period' as a fixed attribute of a site) and can accommodate contradictory (e.g. differences of opinion on typological classification). This is important given that XRONOS aspires to be an authoratative 'backbone' with a global scope, so we cannot realistically impose a single chronological scheme or resolve conflicting information provided by specialists. They are useful practically because they expose a common interface for attributes that all types of chronological information share, such as a *terminus post quem* (TPQ), *terminus ante quem* (TAQ), and midpoint estimate. This allows applications that use XRONOS' data model (including XRONOS itself) to collate chronological data from multiple sources, without necessarily having to be aware of the pecularities of each type of dating.

In order to unify chronological information in the form of a chron, we need a common chronological 'coordinate system'. The natural choice is a *calendar probability distribution*, which expresses the probability that an event occurred as a function of time on a calendric scale. Most archaeologists are familiar with working with this kind of representation in the form of calibrated radiocarbon dates, but it can be extended and generalised to essentially any kind of chronological information (**CITE**). For example, in aoristic analysis (**CITE**), a periodic time estimate (e.g. the event occurred in the Neolithic) is conceptualised as a uniform probability distribution over the timespan between the known start and end dates of that period. A similar model is used in OxCal (**OxCal**) to integrate prior chronological information from diverse sources. In practical terms, this model means that the canonical representation the time component of any chron in XRONOS', regardless of source, is a probably distribution over the set of calendar years (arbitrarily measured in years Before Present) in which it could have plausibly occurred. Further statistics, e.g. a midpoint estimate or TPQ/TAQ range, can be derived from this distribution using well-known methods. In this way, we can support many different types of date and much of the implementation of XRONOS can be agnostic to the source of chronological information.

Chrons are located in space through association to a *sample* – the physical object from which a chronological determination was made. The location of samples is represented with geographical coordinates and an associated coordinate reference system (CRS), though since in practice the precise location of single samples is rarely available, this property is usually inherited from the site. We also record relevant metadata on the nature of the sample. For radiocarbon dates, for example, we follow established conventions (Millard 2014) in recording the type (e.g. charcoal, charred seed) and, where applicable, taxonomic designation (e.g. TODO, TODO) of the organic material used for dating. For typological dates, an ideal scenario would be for the sample to represent the particular object from which an inference was made (e.g. 'Natufian' might be inferred from 'lunate-type microlith'). In practice, the best we can glean from most published datasets is the type of material used (e.g. 'pottery', 'lithics'). The same sample can be associated with multiple chrons, including different types of chron. This is useful, for example, for repre-

senting replicate radiocarbon dates on the same sample, or radiocarbon dates and dendrochronological made on the same section of wood for wiggle-matching.

Further contextual information is associated with *contexts* and *sites*. The site is the primary geographic container for chronological information. As already mentioned, we typically record the spatial location of chrons using this entity, though it is possible to modify this by providing specific coordinates at the sample level. Sites also have attributes describing their conventional name or names in different languages and are associated with a flexible 'site type' typology that combines information on their form and function.

A context represents the specific find-context of a sample, e.g. an architectural feature, stratigraphic unit, or phase. Since the units and conventions for recording such information vary greatly between different regions and archaeological traditions—and XRONOS is designed with global data in mind—we leave the question of what a context precisely represents open, and only record an unstandardised, free text label for it. Crucially, however, contexts can have a self-referential association to other contexts belonging to the same site. This allows it to encode arbitrary relational structures between contexts, whether they be hierarchical (e.g. phases and sub-phases) or graphical (e.g. stratigraphic). In this way, it can serve as a foundation for chronological modelling.

The series of relations `[chron] > sample > context > site` links the chronological and contextual sides of the XRONOS data model. Each step is a many-to-many association, meaning for example that it is possible to attach multiple chrons to the same sample (e.g. replicated radiocarbon dates on the same material), multiple *types* of chrons to the same sample (e.g. radiocarbon dates on tree-rings for wiggle-matching). Since this kind of information is rarely systematically recorded in our source databases, there are currently few actual records that make use of this feature of the data model. However, we hope it will provide a foundation for … ?

- Bibliographic references

- Paper trail

- LOD??

Software architecture

The XRONOS data model is implemented as a relational database using the free and open source database management system PostgreSQL. However, apart from backups and other routine maintenance procedures, all interaction with the database is via a web application, which thus forms the core of XRONOS' architecture. The XRONOS web application is written in Ruby and uses CRUD (Create, Read, Update, Delete) and MVC (Model, View, Controller) patterns as implemented in the Ruby on Rails framework. The choice of this somewhat traditional or 'boring' architecture—as opposed to the non-relational/semantic technologies that are more *à la mode* in the digital humanities (Fan 2018; Hyvönen 2020; Schloen and Prosser 2023)—was purposeful, motivated by a sense that it offers a better chance of longevity and maintanability. Our aim is for the software architecture to be as boring as possible so that the scientific contents can be as interesting as possible.

The XRONOS web application exposes two distinct user interfaces: a graphical user interface accessed through a web browser; and an application programming interface (API). Both interfaces follow a REST (Representational State Transfer) pattern (Verborgh et al. 2015), where each resource (e.g. a single radiocarbon

date, a single bibliographic reference, or a single user) is statelessly mapped to a single address. Users can then interact with resources at these addresses using a preditable and uniform interface based on HTTP verbs. For example, the radiocarbon date RTD-8904 is represented by the address https://xronos.ch/c14s/156205. Users can view information on this resource by sending a GET request to that address, regardless of which interface they are using, and authorised users can modify it using POST, PATCH, or DELETE requests. The bibliographic reference associated with this date (Richter et al. 2017) is similarly represented at the address https://xronos.ch/references/17778, and can be accessed at that address using the same interface as the radiocarbon date. These uniform REST interfaces are another example of a boring architectural choice that make it easier for us to enrich the scientific contents of XRONOS, by adding new types of modular resource that represent new scientific entities.

This basic REST pattern is augmented by seven 'actions' (following the standard pattern in Rails application) that express different ways of interacting with a resource: index, show, destroy, new, create, edit, and update. The 'show' action represents interaction with a single resource, as described above. The 'index' action, which lists resources of a given type (e.g. https://xronos.ch/c14s for radiocarbon dates), is worth special mention because it is through this that the filtering logic at the core of XRONOS' two interfaces is implemented. By passing a query as HTTP GET parameters to the index action of a resource, the list returned the user is modified to only include records that match that query. For example, https://xronos.ch/sites?site%5Bcountry_code%5D=CH (the part of the URL after the ? character encodes the SQL WHERE clause `country = 'CH'` as a GET parameter) lists sites in Switzerland . More complex queries can be executed using nested parameters. For example, https://xronos.ch/c14s?c14%5Bsample%5D%5Bmaterial%5D%5Bname%5D=charcoal (encoding that the `c14` table should be joined to the `material` table via `sample`, followed by the WHERE clause `material.name = 'charcoal'`) lists radiocarbon dates obtained from charcoal samples . Uniquely, index actions can also respond with the result in a tabular data format (i.e. `.csv`).

The graphical user interface (GUI) to XRONOS, accessed through a web browser (e.g. at https://xronos.ch), uses REST resources and actions as the building blocks for various interfaces through which users can browse, search, retrieve, and analyse chronometric data. Each action on each resource is represented by a page, though not all of these are publicly accessible. Pages representing REST resources directly are supplemented by a number of synthetic interfaces, for example the 'data browser' (https://xronos.ch/data), which facilitates more complex filtering, or the search interface (https://xronos.ch/search). The GUI also includes several resources which are not part of XRONOS' scientific data model, for example documentation pages, user profiles and news articles; these are currently not exposed in the API.

Access to various 'backstage' interfaces for creating, editing, and deleting data, and monitoring data quality is managed using a user permissions system. Currently only authorised users affiliated with the XRONOS project can access these, but in the future we intend to support open registration and expose editing interfaces to all authenticated users. For this reason, there is no sharp division between a 'public' and 'private' areas – viewing/querying data and editing/curating data share the same architecture and interface patterns.

The XRONOS API uses the same addresses as the web-based GUI (with the exception of some of the synthetic interfaces mentioned above) but responds with machine-readable data in JSON format, rather than a HTML page. This response

can be triggered by appending `.json` to the address or by including a HTTP `content-format` header in the request. Though users can make such requests manually and parse the data with one of several off-the-shelf tools, the primary intended uses of this interface is to provide access for 1) programmatic clients to XRONOS and 2) other web services. The XRONOS R package (discussed in more detail below) is an example of the former use, which could be extended to other programming environments used for scientific computing, e.g. Python or Haskell.

An overarching principle of this software architecture is that all interaction with XRONOS' data store, and as much of the data processing and 'business logic' of responding to REST requests as possible, is directed through the same server-side routines. First and foremost, this allows us to provide multiple interfaces (i.e. the GUI and API, perhaps more in the future) without duplicating these elements of our codebase. Avoiding reliance on client-side processing, e.g. with Javascript of Web Assembly (WASM)—which would be the other option—also allows us keep our client interfaces simple (in most cases plain, semantic HTML pages and self-contained stylesheets) and therefore, we hope, sustainable in the face of constantly-evolving client-side technologies and standards. It does have the weakness that, in practical terms, XRONOS is difficult to run and relies on the continued existence of a maintained external server. We have however tried to mitigate this by providing clearly-documented source code and regular data dumps so that, if our instance of XRONOS disappears, or if one simply does not want to use it, it is possible for others to host a XRONOS server of their own. As the sustainability of scientific software and data infrastructures is a pressing problem (a point we will return to in the conclusion), in the future it may be desirable to support further decentralisation through, for example, a federated server-server model (**CITE**).

Implementation

The first prototype of XRONOS was developed in 2019. It has been under sustained development 2021–2024 with the first public release in July 2021.

The current database is modular and, in its current stage of development, contains all the structures necessary for storing 14C data. In addition, the integration of dendrochronological information into the structure is already prepared, and a security and data protection architecture has been prepared for the final stage of development. The software is available as open source on Github (https://github.com/xronos-ch), and an R package for easy access, data retrieval and evaluation is also already available. XRONOS currently holds 92,129 14C data from 22,225 sites. The primary focus is on Switzerland and Europe, but the database also contains information from Africa, South America, Asia and Australia. The collection is subject to continuous addition.

Due to space limitations, and because we hope and expect XRONOS to continuously evolve, we don't attempt a comprehensive description of its implementation here. Instead we offer a selection of representative features which illustrate our design goals and key aspects of our approach.

- Data import / assessment of XRONOS' coverage
- The data browser
- XRONOS R package
- Tale from the trenches: cleaning species and countries?
- Tale from the trenches: duplicate processing?
- Case study of use?

Conclusion

- Chronology is the backbone of the archaeological record
- XRONOS could be a backbone for a universe of linked open data resources…
- …but sustainability is a challenge.

  - We're coming to the end of our project funds
  - We'll try to get more, but you never know (grant-based science and long-term investments in digital infrastructure don't work so well together!)
  - Future sustainability

Data availability statement

TODO

Funding statement

Acknowledgements

References

Bird, Darcy, Lux Miranda, Marc Vander Linden, Erick Robinson, R. Kyle Bocinsky, Chris Nicholson, José M. Capriles, et al. 2022. "p3k14c, a synthetic global database of archaeological radiocarbon dates." Number: 1 Publisher: Nature Publishing Group, *Scientific Data* 9, no. 1 (January 27, 2022): 27. Accessed March 23, 2023. https://doi.org/10.1038/s41597-022-01118-7. https://www.nature.com/articles/s41597-022-01118-7.

Bronk Ramsey, Christopher, Maarten Blaauw, Rebecca Kearney, and Richard A Staff Staff. 2019. "The Importance of Open Access to Chronological Information: The IntChron Initiative." Publisher: Cambridge University Press, *Radiocarbon* 61 (5): 1–11. Accessed April 20, 2019. https://doi.org/10.1017/RDC.2019.21. https://www.cambridge.org/core/journals/radiocarbon/article/importance-of-open-access-to-chronological-information-the-intchron-initiative/5D76092C90DED5500B2512E0AC287398.

Fan, Lai-Tze. 2018. "On the Value of Narratives in a Reflexive Digital Humanities." *Digital Studies / Le champ numérique* 8, no. 1 (1 2018). https://doi.org/10.16995/dscn.285.

Hinz, Martin, Martin Furholt, Johannes Müller, Dirk Raetzel-Fabian, Christophe Rinne, Karl-Göran Sjögren, and Hans-Peter Wotzka. 2012. "RADON - Radiocarbon dates online 2012. Central European database of 14C dates for the Neolithic and the Early Bronze Age." *Journal of Neolithic Archaeology* 14:1–4. https://www.jna.uni-kiel.de/index.php/jna/article/view/65/116.

Hyvönen, Eero. 2020. "Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-Analysis and Serendipitous Knowledge Discovery." *Semantic Web* 11, no. 1 (January 1, 2020): 187–193. https://doi.org/10.3233/SW-190386.

Kintigh, Keith. 2006. "The Promise and Challenge of Archaeological Data Integration." Publisher: Cambridge University Press, *Am. Antiq.* 71, no. 3 (July): 567–578. Accessed August 15, 2019. https://doi.org/10.1017/S0002731600039810. https://www.cambridge.org/core/journals/american-antiquity/article/promise-and-challenge-of-archaeological-data-integration/037763619B121990D59B585523826A03.

Kristiansen, Kristian. 2014. "Towards a New Paradigm? The Third Science Revolution and its Possible Consequences in Archaeology." *Current Swedish Archaeology* 22:11–34.

Millard, Andrew R. 2014. "Conventions for Reporting Radiocarbon Determinations." Publisher: Cambridge University Press, *Radiocarbon* 56 (2): 555–559. Accessed February 2, 2018. https://doi.org/10.2458/56.17455. https://www.cambridge.org/core/journals/radiocarbon/article/conventions-for-reporting-radiocarbon-determinations/E4077EC6F1EE1C90C5170309E2C0CF9B.

Richter, Tobias, Amaia Arranz-Otaegui, Lisa Yeomans, and Elisabetta Boaretto. 2017. "High Resolution AMS Dates from Shubayqa 1, Northeast Jordan Reveal Complex Origins of Late Epipalaeolithic Natufian in the Levant." *Sci. Rep.* 7, no. 1 (December 5, 2017): 17025. https://doi.org/10.1038/s41598-017-17096-5.

Schloen, Sandra R., and Miller C. Prosser. 2023. "The Case for a Database Approach." In *Database Computing for Scholarly Research: Case Studies Using the Online Cultural and Historical Research Environment,* edited by Sandra R. Schloen and Miller C. Prosser, 25–73. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-46696-0_2.

Schmid, Clemens, Dirk Seidensticker, and Martin Hinz. 2019. "c14bazAAR: An R package for downloading and preparing C14 dates from different source databases." *Journal of Open Source Software* 4, no. 43 (November 25, 2019): 1914. https://doi.org/10.21105/joss.01914. https://joss.theoj.org/papers/10.21105/joss.01914.

Verborgh, Ruben, Seth van Hooland, Aaron Straup Cope, Sebastian Chan, Erik Mannens, and Rik Van de Walle. 2015. "The Fallacy of the Multi-API Culture: Conceptual and Practical Benefits of Representational State Transfer (REST)." *Journal of Documentation* 71, no. 2 (January 1, 2015): 233–252. https://doi.org/10.1108/JD-07-2013-0098.