ABSTRACT. Radiocarbon data are the most commonly used chronometric measurement technique in archaeology. The introduction of the radiocarbon method offered new potential for independent, internationalized research projects. Today millions of radiocarbon measurements exist globally. However, the many strengths of radiocarbon for research in archaeology have also created an internationally significant challenge in heritage practice. How can we attempt to curate huge volumes of radiocarbon “legacy” data in systematic ways that facilitate interdisciplinary, international research? How can we contend with a dataset that is rapidly scalable, and needs to be kept live—updated, validated, curated, and related to existing national archives and data systems—beyond the timescale of any individual project? In this paper we introduce an international project, “Project Radiocarbon; Big Data, integrated cross-national heritage histories”, working across the historic environment sector in Ireland and the United Kingdom, that is developing a solution to these issues. We argue that we need to think critically about how we classify and curate radiocarbon data, to render them interoperable and findable. Such work requires inter-sector approaches to ensure sustainability and scalability, and to anticipate the increasing value of these data into the future.

INTRODUCTION: BACKGROUND TO THE PROJECT

The introduction of the radiocarbon dating technique represents one of the most significant developments in the history of archaeological thought; with this innovative method archaeology really started again in the 1950s (cf. Griffiths 2022), and the method has been hailed as a revolution in understanding for archaeology (cf. Bronk Ramsey 2009; see Griffiths et al. 2023 for further discussion on the “revolutionary” nature of the technique).

Radiocarbon measurements are the most routinely used form of chronometric data in archaeology. The technique’s ubiquity derives from the wide range of sample types that can be measured, and the production of results with precision that is routinely useful for archaeological research questions. Innovations in measurement (including through the development of Accelerator Mass Spectrometry) have led to reductions in required sample size that have furthered the use of the technique, and relative costs of measurements have decreased over time. In contrast, for example, dendrochronology, which was developed prior to radiocarbon and can give much more precise measurements, can only be applied in a more restrictive set of circumstances. Radiocarbon is the international scientific dating technique for archaeology, and there has been a proliferation of measurements over the last 70 years.

*Corresponding author. Email: seren.griffiths@mm.ac.uk
Today millions of radiocarbon measurements probably exist globally, with measurements commissioned by many different organisations, for a variety of purposes, at many different laboratories. Radiocarbon measurements may be produced by professional archaeological organisations working in developer-led or Cultural Resource Management contexts, or by museums as part of their collections management policies. Government or other national historic environmental agencies may also have programmes of radiocarbon measurements to support the management of their estates or sites of national significance, as a result of community-led archaeology projects, and as part of police and forensic work. Academic research projects may produce measurements as part of synthetic archaeology and palaeoenvironmental research projects. Private individuals may commission results because of personal interests, or perhaps in order to authenticate artefacts prior to sale. Because radiocarbon is the global chronometric technique, the scale of production of radiocarbon data is directly related to national and international heritage management policy, as well as global economic issues. For example, the number of radiocarbon data significantly increased in the United Kingdom after changes in Governments’ planning policies (e.g., the introduction of Planning Policy Guidance 16 and Planning Advice Note 42; e.g., Bayliss et al. 2008). In the Republic of Ireland, the construction boom associated with the “Celtic Tiger” economy had a significant impact on the production of radiocarbon data (e.g., Eogan and O’Sullivan 2009).

Recently, developments in Big Data approaches (coupled with advances in the computer processing power) have resulted in renewed interest in analysis of populations of radiocarbon data, especially when combined with innovative new scientific approaches (like aDNA research; cf. Griffiths et al. 2023 and discussion below).

The many strengths of radiocarbon for research in archaeology have also created an internationally significant challenge in heritage practice. How can we attempt to curate huge volumes of radiocarbon “legacy” data in systematic ways that facilitate interdisciplinary, international research? How can we contend with a dataset that is growing rapidly, and a repository that needs to be kept live and scalable—updated, validated, curated, and related—to existing national archives and data systems—beyond the timescale of any individual project? Because of the scale of radiocarbon legacy datasets, the range of stakeholders producing and consuming these data, and the significant potential of these data in future research, the challenges for the curation and reuse of radiocarbon legacy data are global. While some of these issues can be understood in relatively common “FAIR” terms of data reuse (Crosas et al. 2016; Wilkinson et al. 2016), we also argue that there are specific and significant implications for heritage data science in using chronometric legacy data in archaeological research undertaken at the international level (cf. McKeague et al. 2020). We identify these in terms of the utility of Big Data approaches within archaeology, and in terms of the practice of archaeology as a global, integrated discipline (cf. Griffiths et al. 2023). There are also specific issues in terms of the creation of additional value in heritage data science (which may increase from the point in time when physical archives are deposited in an archive), and the curation of relationships between physical archives and data in order to ensure discoverability (some of us discuss these issues in more detail elsewhere; Johnston et al. 2023).

In this paper we introduce an international project—“Project Radiocarbon; Big Data, integrated cross-national heritage histories,” working across the historic environment sector in Ireland and the United Kingdom, that is developing a solution to these issues. There is considerable variation in government and state heritage agency policies across these jurisdictions. This project is a collaboration between people with specialisms in archaeology, data science, and scientific dating working in academia, in government
historic environment agencies and other government agencies, and in museums. The project will create digital infrastructure for the curation and transformation of extant radiocarbon data from Ireland and the United Kingdom. The results will be available as open access and reported to maximize “FAIR” principles. Beyond the project lifespan, “live” updates and data cleaning will be provided by national government partners to ensure sustainability. We will create new map-based and time-based search potential, to increase data reuse. We will increase value in two significant ways, by transforming data in terms of international interoperability and findability (creating persistent identifiers; see discussion below) and repopulating archaeologically-important attribute data to extant legacy data (see discussion below). We want to change how we curate radiocarbon data, how we think about the digital legacy of radiocarbon data, and how we create change through interactions across the heritage sector.

We outline our approach using a series of key themes. Firstly, we argue that we need to address some fundamental issues with the practice of archaeology to work at this international level. As chronological data, radiocarbon measurements have an important role in creating archaeological knowledge; working to achieve interoperability across diverse datasets requires us to address some fundamental aspects of archaeological practice. Secondly, we argue that in working with these legacy data we can contribute to more considered and creative practice in Big Data research, that goes beyond reproducing our disciplinary knowledge structures and historic interpretations. Thirdly, we want to highlight some under-recognized potential in such datasets in terms of research practices in the historic environment. Finally, we emphasize the importance of imaginative approaches to sustainability and scalability, and to recognize the need to foster creative ways for heritage stakeholders (including members of the public, professional archaeologists, researchers, curators and so on) to interact across the sector to maximize the value of extant data.

DISCUSSION
The Research Potential and Challenges
Radiocarbon data need to be correctly reported if their research value is to be curated and effectively leveraged in subsequent research (cf. Millard 2014; Bayliss 2015). As Johnston et al. (2023) have argued for radiocarbon data specifically—and as part of the “organics revolution” in archaeology more broadly—the value of such data and associated archives have significantly multiplied over the last few decades. However, despite the longevity of the method and the centrality of radiocarbon to archaeological practice, there is a crisis in reporting data and therefore in reproducing these data. Radiocarbon measurements cannot be used in isolation—each measurement is produced with associated data that are essential for its interpretation (including site details (e.g., “the archaeological context”), and measurement details (e.g., “the measured sample”). If attribute data that are produced during the processes of radiocarbon measurement are lost or dissociated, the utility of radiocarbon data become compromised and their value lessened.

As we noted above, there is a global issue in terms of radiocarbon data preservation and curation. Across the United Kingdom and Ireland, for example, there is no single functioning radiocarbon archive that is publicly accessible, that ensures quality and accurate data, and is live to allow continuous data deposition. Although both Historic England and Historic Environment Scotland maintain databases for their jurisdictions, there remain issues with public access to these repositories and ensuring repositories are up-to-date, legacies of pre-digital approaches to data curation. In the case of Ireland and Wales to collate and curate data...
Despite the structured nature of scientific data, there is no agreed approach to collecting and sharing data across custodians. Across these jurisdictions there are significant issues with making chronometric data from different nations interoperable (see discussion below). Because of this, in these countries alone, data which cost millions of euros/pounds for the radiocarbon measurements alone (excluding for example the cost of excavation and curation) are being made rogue—with inaccurate, incomplete, or otherwise compromised attributes of radiocarbon measurements not reported in the literature or reported incorrectly.

Globally, this picture is replicated. As a result, research is significantly impoverished, despite many attempts to curate data from across national resources. These have included an initial International Radiocarbon Database (e.g., Kra 1989) along with more recent international, regional or thematic repositories (e.g., Gajewski et al. 2011; Loftus et al. 2019; Bird et al. 2022; Kelly et al. 2022) as well as laboratory-led initiatives (Bronk Ramsey et al. 2019). A full review of the existing approaches to data curation is not the purpose and beyond the scope of this paper. Generally, however, many of these initiatives face fundamental challenges: outside laboratory-led projects, this work is often not designed to create “live” datasets; datasets may not be related to government heritage agency identifiers (so data risk being orphaned—that is data may become disassociated from the context of production); it may not be possible to render datasets interoperable or findable because of an absence of internationally-defined persistent identifiers. Internationally, projects like ARIADNE provide digital archive infrastructure for the archaeological community worldwide which is updated (Aloia et al. 2017; Meghini et al. 2017; cf. Wright and Richards 2018), however, such projects may not currently include specialist data standards to ensure chronometric data are reusable to their full potential or that the maximum value of these data are retained.

All such initiatives face common challenges when working internationally, including challenges in political terms—international cooperation is required across heritage agencies to ensure long term sustainability—and epistemological terms—there remain fundamental challenges to achieving international interoperability for archaeological data.

**Transformation and Interoperability—Why Data Mining Approaches are Insufficient**

As we noted above, huge numbers of radiocarbon data are available on numerous publicly accessible data repositories. These have been curated to various standards and include many different fields, which will be of varying utility to researchers depending on their interests. This might include, for example, ensuring that common geospatial data systems are associated with measurements in order to undertake temporal-spatial analysis. It might involve ensuring that all data are associated with accurate laboratory codes so that subsequent researchers can cross-reference data. Researchers using legacy data often “clean” datasets to address very specific project needs, and in these cases scientific data can become orphaned from national and regional inventories of sites, or from details of the archaeological interventions, or lose spatial definition more generally (cf. McKeague et al. 2017).

Data cleaning approaches create datasets that are more accurate and more useful than they would otherwise be. However, the internationalized research potential of radiocarbon data may only be achieved with additional data transformation, including one of the fundamental aspects of radiocarbon datasets—how archaeologists use temporal descriptions to categorize these data.
Across archaeological research, radiocarbon data are defined and categorized using local temporal schemes. These might include classifying data according to “periods”, “culture historic terms”, “peoples”, “pollen zones” and so on. Such classifications can be formalized internationally or by national heritage agencies (e.g., Forum for Information Standards in Heritage or “FISH” (Historic England n.d. a, http://purl.org/heritagedata/schemes/eh_period). These terms are however, heuristic devices, foundational concepts—that now seem “natural”—because they were often developed early in the history of archaeology when research did not have the benefit of chronometric measurements. These kinds of localized temporal terms present fundamental issues for undertaking research at an international level because these definitions can be specific to individual nation states or regions within states. In this case, interoperability in international research requires us to go beyond data collation and cleaning; we sometimes need data transformation.

The approach we are taking in in Project Radiocarbon to make these terms interoperable is to create persistent identifiers defined temporally for our culture historic labels in order to render them findable and interoperable (https://perio.do/; Rabinowitz 2014). This will allow international research across traditional culture historic and national boundaries (cf. Griffiths et al. 2023; Feinman and Neitzel 2020).

Interoperable international persistent identifiers exist in the PeriodO resource (https://perio.do/en/), where local culture historic terms are defined temporally in years BC/BCE and AD/CE, as part of the international gazetteer of data schemes (cf. Rabinowitz 2014; Rabinowitz et al. 2018). In the case of Project Radiocarbon, we were able to create the first formally defined temporal classificatory system for archaeological practice on the island of Ireland (Carlin et al. 2022; http://n2t.net/ark:/99152/p06hps8). By relating these terms in this way from across the United Kingdom and Ireland, we can for example, compare evidence from our legacy dataset for “Pictish” (https://client.perio.do/?page=period-view&backendID=web-https%3A%2F%2Fdata.perio.do%2F&authorityID=p0xxt6t&periodID=p0xxt6tghvr; Historic Environment Scotland 2018) activity in fourth century CE/AD in modern-day southern Scotland, with contemporary “Roman” activity in fourth century CE/AD activity in modern-day northern England (Historic England n.d. b, http://www.heritage-standards.org.uk/chronology/). A further challenge in international research may be presented in the requirement to create “fuzzy” temporal search boundaries, where the timing and magnitude of social changes apparent in the archaeological record are—in themselves—the foci of research; in these cases, it may be possible to incorporate temporal uncertainty into searches semantically (e.g., Binding 2010), so that ambiguity in the temporal definition of local culture historic terms can be incorporated into searches (e.g., if we are unsure about the temporal definition of culture historic terms like “the Neolithic” in different regions or nations). In our approach, we are also able to use temporal units (e.g., the 23rd century cal BC/BCE) as the basis for data interrogation, an approach which provides a means to directly counter the persistence of culture historic approaches in archaeological research (cf. Griffiths et al. 2023; Feinman and Neitzel 2020).

**Big Data and Dates-as-Data Approaches**

We suggest that the temporal description of archaeological data is especially important given recent trends in Big Data in archaeology. Interest in the analysis of populations of radiocarbon data has grown since the early development of the technique (e.g., Deacon 1974 in southern Africa; Rick 1987 in north America), and as part of these approaches we have seen repeated protocols developed for the rejection or inclusion of measurements in any given analysis.
(e.g., Ashmore 1999; Pettitt et al. 2003). In Big Data approaches, analyses generally take as their starting points a “dates as data” approach, whereby patterns in the distribution of a population of radiocarbon measurements are argued to be archaeologically meaningful, representing variously human population size, activity levels and so on (e.g., Shennan et al. 2013; Chaput and Gajewski 2016; Zahid et al. 2016; Bevan et al. 2017). There are tensions in these approaches, in terms of what numbers of radiocarbon measurement mean, and an extensive literature has made comparisons between these analyses (cf. Contreras and Meadows 2014; Bronk Ramsey 2017; Crema and Kobayashi 2020 and references therein).

The intellectual ambitions of Big Data approaches in archaeology more broadly provide an important context for these studies. Big Data approaches are not “simply” defined by the number of datapoints in any given analysis. Kitchin (2014) recently defined “Big Data” as: huge in volume, exhaustive in scope, and striving to capture entire populations or systems. While archaeological Big Data work may not be comparable to studies in the social sciences or biomedical fields (for example), some archaeological research has some of the qualities of Big Data work. We suggest this is especially true in recent aDNA population genomics research which often—by its nature—inherently operates at a population-level (regardless of the actual size of the dataset). Similarly, research that use dates-as-data are “Big Data” in their outlook; when the attempt is to achieve a completist record of the population of radiocarbon measurements, the sample is the population. As noted above, the real question is what this population represents, for example whether patterns reflect research history rather than anthropogenic or palaeoenvironmental changes in the past (cf. Contreras and Meadows 2014; Crema and Bevan 2021); it becomes very important that these distributions are not biased in unexpected ways.

We suggest that at this level of Big Data analysis there is an important interplay between the quantitative sample of radiocarbon data and the qualities which archaeologists attribute to radiocarbon measurements, where bias can become incorporated into space-time analyses regardless of which temporal analytical methods are employed. This bias originates in the archaeological heuristic devices (“the Bronze Age”) that are used to classify data; these can create spurious structures in datasets. When temporal heuristic devices—like “the Bronze Age”—come to be regarded as inalienable aspects of radiocarbon measurements, these devices can become the foci of knowledge production as if these concepts actually existed as entities in societies in the past (Griffiths et al. 2023).

The utility of radiocarbon data is determined by the processes of knowledge production or data transformations from the field up, what Chapman and Wylie (2016) have called the “contingent logical scaffolding”. Any interruption of this process means that data cease to transport truth, as Latour (1999:69) put it. If data are transformed so that they become identified by their nature using archaeological heuristic devices, some of the contingent logical scaffolding is also stripped away; no radiocarbon measurement is by its nature “Bronze Age”, but these kinds of rationalising transformations—from radiocarbon measurements into theoretically-laden knowledge claims—routinely happen in Big Data approaches. We argue that this process of rationalisation and the chronological tension it entails is in part why some recent aDNA work has been so contentious (cf. Booth 2019; Frieman and Hofmann 2019; Furholt 2019a, 2019b; Sykes et al. 2019). In Big Data studies where radiocarbon data are essentialized in theoretically-laden terms, we risk creating interpretations that become “coherent myths” (Oldfield 2001:123), “just so stories” or “Panglossian Paradigms” (cf. Gould and Lewontin 1979). If space and time are the fundamental axes against which we identify patterns in data, then the co-location of radiocarbon measurements using both radiocarbon
ages and archaeologically-situated terms can create tension and bias in our narratives. Internationally, in order to challenge these biases we need interoperability between different time structures.

Curating the Historic Environment—identifying richness and research patterns

Beyond dates-as-data analysis of evidence for human activity in the past, there are some areas where dates-as-data approaches to radiocarbon populations can provide us with incontrovertible and rich research potential.

Radiocarbon data distributions can directly inform about histories of archaeological research practice (including variation in research priorities and funding in different regions and nations). These details can in turn inform in the management of the historic environment. For example, arguably the most important research priority for the Stonehenge World Heritage Site should be the period between 800 BC/BCE–43 AD/CE (the local culture historic “Iron Age” period), for which we have the least chronometric data, rather than the research emphases placed on the local “Neolithic” and “Bronze Age” cultural historic records (cf. Leviers and Powell 2016).

We argue that the research potential of radiocarbon data for exploring and understanding our own research history in these terms is an under-explored aspect of management of the historic environment. Part of our analysis from the dataset that we will be undertaking is engaging with the research foci of radiocarbon measurements—over time and space, and over different local culture historic entities—from across the historic environment in Ireland and the United Kingdom.

Inter-sector approaches, sustainability and scalability

We argued above that in creative chronology projects it is not enough to collate and clean datasets; we need to transform data to make them interoperable. We further argue that solutions for the crisis in radiocarbon data curation also need to be sustainable and scalable. Our resource will be hosted by the Archaeology Data Service and will be openly and freely accessible, with discoverability routes from national heritage partners. The platform will have long term sustainability beyond the project lifecycle, and the resource will be scalable. New data will continue to be validated and deposited in the resource by our historic environment partners, making it a live research tool going forward, with a lasting high-quality legacy, and with significant impact for historic environment research quality and open access. We are achieving this by working closely with our stage heritage partners to achieve for example new service agreements to update the repository.

We argue that this kind of scalability should be essential to ensure that the value of chronometric data is maintained into the future. Indeed, we have seen as part of the “organics revolution” in archaeology more broadly that the value of archaeological organics in archives has grown significantly with the recent advent of a range of techniques including radiocarbon measurements, but also aDNA, stable isotope, biomolecular and other analyses (cf. Johnston et al. 2023). Scalability and sustainability are essential in order that we can safeguard the current value of these dataset, and anticipate the potential for growth in value of both digital data and the associated parent archives. Globally, the heritage sector needs to be able to account for the value of our archives, and this requires more creative solutions to digital data.

Because of the fundamental importance of chronometric data to archaeological practice in all our nations, and because of the importance for future legacy, collaborations working across national jurisdictions and organisations were required to develop Project Radiocarbon. Work at this level can be challenging, addressing issues that touch on institutional and national
politics. We suggest that sometimes creative and playful solutions—which harness the resources and skills from inter-sectional approaches, and which emphasize the collaborative experience of such work—are the most effective means to create the ambitious changes such as those we are attempting to build in Project Radiocarbon.

CONCLUSION

Radiocarbon measurement created a fundamental change in the practice of archaeology. As a globally applicable technique, with utility over some 60,000 years of the archaeological record (and c. 75 years of archaeological research) has been hugely important in the development of modern archaeology. This ubiquity also creates a fundamental challenge in ensuring the value of “legacy” data is safeguarded. Our approach in Project Radiocarbon has included not only collating and cleaning data, but also transforming and making interoperable data in order to create the potential for new internationalized research, including new approaches in Big Data projects. If we do not address some of the fundamental issues in the way data are curated, we suggest that simplistic interpretations may reproduce archaeological heuristic devices, rather than critically engaging with evidence.

In order to create international change in heritage data science we need ambitious and creative solutions, working internationally, across agencies and sectors. This kind of collaboration is key to accessing funding, and to ensure the scalability and legacy of our hard-won data.

ACKNOWLEDGMENTS

Dr Steve Burrow (Burrow and Williams 2020) and Robert Chapple (Chapple 2019) have independently curated and catalogued radiocarbon data from Wales and the island of Ireland for many years. Ian Ralston and then Patrick Ashmore (Ashmore et al. 2000) compiled radiocarbon data in Scotland until 2005, with additional radiocarbon data collated and published by Alison Sheridan, Ian Shepard and others (https://canmore.org.uk/project/919374). This project is indebted to their work. We would like to thank two anonymous reviewers for their comments on the paper, which helped improve it.

FUNDING STATEMENT

This project was funded under the joint Irish Research Council and the United Kingdom Research and Innovation, Arts and Humanities Research Council scheme “Collaboration in Digital Humanities Research” (AH/W001942/1; IRC/W001942/). No new data were generated for this publication. This research has been supported by contributions in kind from Archaeological Data Service, Department for Communities (Northern Ireland), Department of Housing, Local Government and Heritage (Republic of Ireland), Historic England, Historic Environment Scotland, National Museum of Ireland, Royal Commission on the Ancient and Historic Monuments of Wales, Transport Infrastructure Ireland, and The Heritage Council.

REFERENCES


Historic England n.d. a. Available at: http://purl.org/heritagedata/schemes/eh_period


