

Creación de un Archivo Mundial de Datos de Radiocarbono en Arqueología

Andrew Martindale y Thomas J. Brown

Creación de un Archivo Mundial de Datos de Radiocarbono en Arqueología¹

Andrew Martindale
Thomas J. Brown

*Department of Anthropology, University of British Columbia
6303 N.W. Marine Drive, Vancouver BC Canada V6T1Z1
andrew.martindale@ubc.ca*

Introducción

Existe una desconexión entre la necesidad arqueológica de contar con una base de datos de radiocarbono correcta y completa y la capacidad colectiva de la comunidad arqueológica de desarrollar plataformas digitales de archivo para esos recursos. A pesar del evidente beneficio colectivo de una plataforma mundial singular y de la relativa facilidad para crear la infraestructura digital, la mayoría de los esfuerzos actuales resultan básicamente regionales, lo que fragmenta y reduce las posibilidades de búsqueda. Esto dificulta la localización de datos relevantes actualizados, la compilación de cronologías locales o la recopilación de grandes conjuntos de datos para análisis paleodemográficos, que requieren de gran cantidad de dataciones (por ejemplo, Vermeersch 2005; Peros et al. 2010; Gajewski et al. 2011; Woodbridge et al. 2014; Chaput et al. 2015. Cf. la discusión completa en los capítulos 24 y 25 de este libro).

En los últimos años hemos intentado aumentar la utilidad de una de estas instalaciones regionales, la Base de Datos de Radiocarbono Arqueológico Canadiense (CARD) (<http://card.anth.ubc.ca/>), para hacerla más útil en áreas más allá de Norteamérica. Como parte de este proceso, comparamos grandes bases de datos ¹⁴C de todo el mundo en un esfuerzo por identificar el rango de variación en los estándares de datos, la gestión de datos y la utilidad. El resultado de estas comparaciones ha ayudado a identificar un conflicto estructural significativo en la construcción de bases de datos arqueológicas ¹⁴C. La estandarización de los campos de datos que requiere una base de datos mundiales, que debe ser utilizada por investigadores e investigadoras con formas de trabajo muy diferentes, crea limitaciones que hacen que un archivo sea menos valioso para usuarios/as locales. Dado que los archivos mundiales dependen en gran medida de la colaboración de investigadores individuales que trabajan principalmente en contextos regionales,

¹ Traducción al castellano de Juan A. Barceló

no es mucha la motivación para contribuir a una base de datos mundial, que puede que no satisfaga las necesidades locales.

Una base de datos mundial de radiocarbono requiere dos componentes principales: 1) infraestructura digital y programación para proporcionar acceso y funcionalidad a los usuarios, y 2) un diseño de gestión de datos que estandariza el alcance y el contenido de los campos de datos. Inesperadamente, las limitaciones que hemos encontrado no están relacionadas con la programación o la infraestructura digital, sino que se derivan de cuestiones más ontológicas y epistemológicas. Argumentamos que las bases de datos globales, incluso para datos tan consistentes como los resultados del radiocarbono en contextos arqueológicos, producen discordancia entre la coherencia de los datos (estandarización de los campos) y la amplitud de los datos (riqueza de los datos). Esta cuestión a menudo se pasa por alto en el desarrollo de bases de datos regionales, pero sigue siendo un obstáculo crítico para el desarrollo de una plataforma mundial de ^{14}C . Aquí revisamos las características clave de los sistemas existentes y evaluamos las opciones para llevar a cabo esta ambición.

¿Qué es una base de datos?

Una base de datos no es sinónimo de un contenedor de datos. Más bien se trata de un repositorio estandarizado y sistemáticamente organizado de campos de datos relacionados y definidos, diseñados para un propósito específico. El diseño de la base de datos plantea tres preguntas relacionadas:

1. ¿Por qué construirlo? Los objetivos y propósitos de una base de datos definen su diseño. Diferentes propósitos pueden entrar en conflicto y competir por recursos limitados de programación e infraestructura. Inevitablemente, si no se establece de antemano, surge una jerarquía de prioridades.
2. ¿Para quién es? Los usuarios esperados de una base de datos ayudan a definir su propósito y características, pero también guían la gestión de datos. No todos los usuarios tienen requisitos similares, y el diseño de la base de datos para usuarios diversos a menudo equilibra la funcionalidad específica con la utilidad general.
3. ¿Cuál es su escala? Las bases de datos tienen alcance (la gama de contextos de los que se extraen los datos), capacidad (la amplitud de la cobertura de los datos), precisión (el nivel de detalle en los campos de datos), exactitud (la tolerancia a la incertidumbre en los campos de datos) y duración (el período de archivo del sistema). Las decisiones relativas a estos parámetros definen la estructura de la base de datos.

Por lo tanto, una base de datos no es una cosa singular, sino un conjunto de relaciones que surgen de decisiones que se toman mejor de antemano con respecto a su propósito, función y escala. Aunque existen bases de datos ^{14}C con alcance global, incluyendo CARD, nadie ha diseñado todavía una base de datos verdaderamente global. La historia de CARD ilustra esta distinción.

3. La historia de CARD

La Base de Datos Arqueológica Canadiense de Radiocarbono (CARD) fue desarrollada por Richard Morlan del Museo Canadiense de la Civilización (ahora Museo Canadiense de Historia) a finales de la década de 1980 como uno de los primeros sistemas públicos de datos en línea (ver Morlan 1999). La ambición de Morlan era recopilar datos de radiocarbono de contextos arqueológicos, geológicos y paleontológicos de Canadá y el norte de EE.UU. en formato digital,

siguiendo el modelo de las listas de fechas publicadas del ^{14}C establecidas en revistas como *Radiocarbon*. La plataforma digital y en línea proporcionó acceso a usuarios de todo el mundo y presentó un archivo con capacidad de búsqueda utilizando un programa de base de datos llamado *Cardbox*. La funcionalidad estaba muy avanzada, pero CARD no dejaba de ser, en realidad, un mero catálogo digital de tarjetas de archivo plano, en el que cada fecha se representaba como una sola tarjeta/archivo con campos alfanuméricos. Las funciones de búsqueda, edición y carga/descarga eran limitadas, pero al jubilarse en 2005, Morlan y sus colegas habían recopilado más de 36.000 datos de radiocarbono de Norteamérica.

CARD fue actualizado en 2015 a través de una asociación con el Laboratorio de Arqueología de la Universidad de Columbia Británica y reconstruido sobre una plataforma establecida en el proyecto de investigación *Ancient Maize Map* (en.ancientmaize.com). El formulario revisado ofrecía nuevas funciones de búsqueda, edición y carga/descarga, así como una interfaz de mapa y un muro de seguridad. Debido a que la proyección espacial se basaba en un motor de Google Maps, por defecto ya era una interfaz global. Para 2016, una comunidad global de usuarios había expandido CARD a más de 72.000 ^{14}C fechas de todo el mundo (Figura 1). Mientras que la programación y la infraestructura digital fueron actualizadas en la renovación de CARD, el diseño de la base de datos siguió siendo el que había sido desarrollado por Morlan en la década de 1980: 31 campos de datos, 9 de los cuales son obligatorios. CARD sigue siendo más apropiada para el contexto canadiense que para un uso global. Por ejemplo, exige que los sitios arqueológicos tengan códigos de identificación únicos emitidos por el gobierno, que en Canadá son emitidos por las oficinas provinciales de patrimonio y se ajustan al sistema nacional de normalización de designación de sitios. Es efectivamente una base de datos local manejada por una pequeña hoja de cálculo con una interfaz basada en la web que proporciona un conjunto de herramientas útiles de consulta de datos. Aunque se ha convertido en una empresa global, no fue diseñada para serlo y como resultado no cumple con los tres elementos de propósito, funcionalidad y escala a nivel global. Dado que CARD no fue diseñado como una herramienta global, hemos llegado a un punto muerto crítico en el que CARD actúa como una plataforma global, pero no es efectiva como tal.

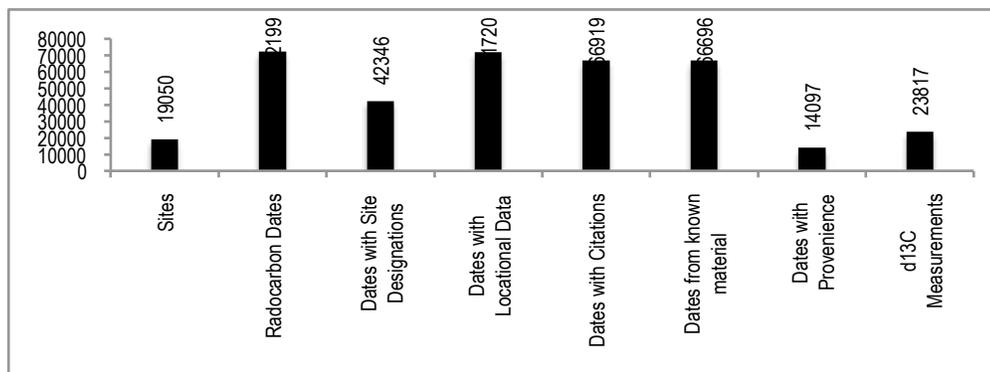


Figura 1. Perfil de los campos de datos en CARD. Obsérvese que los sitios/fechas canadienses cumplen casi al 100% con los requisitos de datos, lo que refleja la naturaleza regional de la base de datos.

4 Requisitos de una base de datos global de ^{14}C

La distancia entre lo que CARD es y lo que debiera ser una base de datos global de ^{14}C puede evaluarse con referencia a las tres preguntas iniciales sobre el diseño de una base de datos. CARD, tal como existe actualmente, es: 1) un archivo de datos de radiocarbono de interés para la arqueología en Canadá y, en menor medida, en los EE.UU; 2) funciona principalmente como depósito estático de un pequeño conjunto de campos de datos que definen las mediciones de la edad ^{14}C en un conjunto más o menos limitado de muestras arqueológicas, especificando las ubicaciones de los sitios, por lo que tiene algunos campos para el contexto de muestras y la procedencia de los datos, y 3) tiene una escala global que proyecta la herramienta de gestión de datos local desarrollada para un contexto canadiense en jurisdicciones de todo el mundo.

Por el contrario, una plataforma de datos global debería 1) proporcionar herramientas y campos de gestión de datos adecuados para todos los contextos arqueológicos del mundo, 2) ampliar su funcionalidad para facilitar tanto las herramientas de archivo como las analíticas, y 3) estandarizar los campos de datos para proporcionar valor a los usuarios locales y la revista *Nature* (Jones 2017; véase también Wood 2015 para un debate sobre la importancia de las normas de información), la ambición de equilibrar los intereses locales y mundiales puede resultar "digna pero infructuosa". Los desafíos son numerosos. Incluyen la definición de los campos de datos básicos requeridos en términos estandarizados para uso global. Por ejemplo, ¿deberían incluirse los valores de $\delta^{13}\text{C}$? Estos juegan un papel importante en la contabilidad del fraccionamiento isotópico para calcular las estimaciones normalizadas de edad de ^{14}C en las mediciones de MGA, pero se confunden fácilmente con los valores de IRMS y no son comparables con ellos para la reconstrucción de la dieta. Una serie de preguntas similares surgen a escala global que no son tan aparentes o no son tan relevantes a escala local. Sin embargo, un sistema global requiere una estandarización aceptada a nivel mundial. En segundo lugar, la funcionalidad se vuelve igualmente compleja a medida que una audiencia global de investigadores e investigadoras presenta división de intereses. Otras herramientas mundiales de investigación, como OxCal, revisan y amplían continuamente sus capacidades funcionales; para ello, una base de datos mundial sobre ^{14}C requiere una financiación considerable y una gestión colectiva que a menudo supera la capacidad de los sistemas locales. En tercer lugar, el aliento y la calidad de los campos de datos varían considerablemente para los investigadores de todo el mundo. Por ejemplo, el registro del contexto de las muestras de materiales fechados tiene un valor arqueológico, pero parece poco probable que surja de forma rápida o económica un sistema mundialmente normalizado para describir la procedencia y la estratigrafía.

Así pues, los desafíos de un sistema mundial no radican en la capacidad de construirlo, sino en el equilibrio de los intereses contrapuestos entre las necesidades particulares de los usuarios locales de datos y las necesidades generalizadoras de un archivo mundial amplio y normalizado. Las pequeñas bases de datos locales, como el sistema CARD original, tienden a sacrificar la coherencia en aras de la exhaustividad: se adaptan a las necesidades locales a expensas de las normas mundiales. Por el contrario, las comunidades de investigación pequeñas, regionales o temáticas, llegan más fácilmente a las normas por consenso o están más dispuestas a adoptar las normas impuestas por las principales figuras dentro de ellas. Sin embargo, la longevidad y la utilidad más amplia de estas bases de datos se ve obstaculizada en gran medida por su falta de coherencia disciplinaria. Una revisión de las principales bases de datos en línea del ^{14}C (Figura 2) revela su naturaleza parroquial: la mayoría se centran en pequeños conjuntos de datos regionales.

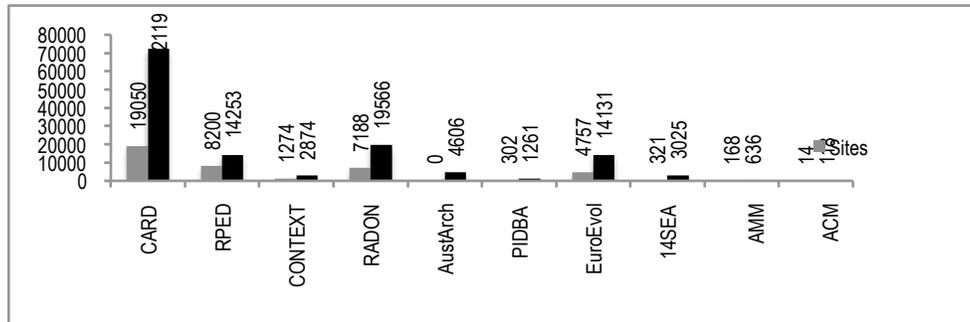


Figure 2. Comparison of the numbers of sites and radiocarbon dates collated in major databases.

Estos sistemas tienden a existir durante un corto período de tiempo, a menudo los años de la carrera activa de los investigadores que han creado ese conjunto de datos. Por ejemplo, CARD permaneció inactiva desde 2005 hasta que el actual equipo ejecutivo la resucitó en 2014. Los proyectos a pequeña escala y de corta duración también tienen dificultades para que la financiación a largo plazo se extienda más allá de su ámbito geográfico o temático original. Así, mientras que los sistemas a pequeña escala dominan las opciones existentes para los archivos ^{14}C , no tienen ni el alcance ni la estabilidad para convertirse fácilmente en proyectos globales.

La capacidad de cálculo y los requisitos informáticos necesarios para un archivo global son ahora fáciles de conseguir. Por ejemplo, todo el sistema CARD es menor que un Gigabyte y se ejecuta en un servidor de la Universidad de Columbia Británica a un costo de menos de \$200 (CDN) por año. Todo el sistema fue construido por menos de \$20,000 (CDN), y las nuevas características funcionales cuestan unos cuantos miles de dólares canadienses cada una. Como CARD ya funciona como un sistema global, demuestra que una plataforma global es factible y asequible, si se pueden conciliar los retos del diseño de datos.

5 Modelos para una Base de Datos Global de dataciones ^{14}C

Como se ha señalado anteriormente, el problema central al que se enfrenta un archivo digital ^{14}C mundial es la normalización de los campos de datos. La combinación de comunidades de datos regionales y temáticas en una sola plataforma representa un cambio de alcance y duración que reúne a investigadores con diferentes requisitos de capacidad, precisión y exactitud. Una simple combinación de campos de datos no resolverá estas diferencias. Los desafíos provienen de los propios usuarios, cuyas prioridades y objetivos en materia de datos difieren. Basándonos en nuestra experiencia con CARD y las revisiones hechas a nuestro sistema por parte de otros investigadores e investigadoras, sugerimos que existen tres opciones para el desarrollo de una plataforma global.

- **MODELO AMPLIO: LOGRAR UN MODELO SINGULAR COHERENTE A ESCALA GLOBAL.** El primer enfoque consiste en desarrollar una plataforma mundial basada en el modelo de los sistemas locales de archivo, es decir, establecer o imponer un conjunto de normas disciplinarias para los campos de datos y la taxonomía de datos. Este modelo es simplemente una expansión de los formularios actualmente disponibles a escala global. El reto, como hemos señalado anteriormente, es llegar a un acuerdo sobre las normas de datos, al tiempo que se incluye la más amplia gama posible de campos. Por ejemplo, la base de datos Radiocarbon Palaeolithic Europe (RPED: <http://ees.kuleuven.be/geography/projects/14c-palaeolithic/>), contiene más de 100

campos de datos e incluye 39 campos de procedencia. Funciona gracias al trabajo dedicado de su creador, Pierre Vermeersch, que convierte los datos publicados en sus estándares (Vermeersch 2005). Se beneficia tanto de una larga duración de varias décadas, que anima a los usuarios a ajustarse a los estándares del PRED, como de un enfoque regional y temático sobre las fechas del Paleolítico procedentes de Europa. El desarrollo de un sistema mundial análogo es posible, pero probablemente requeriría mucho tiempo y esfuerzo colectivo, por lo que se necesitarían fondos sustanciales. Un proyecto de este tipo no tendría ninguna garantía de éxito, ya que se basa en que el grupo de usuarios más heterogéneo llegue a un consenso.

- **MODELO REDUCIDO: ESTANDARIZACIÓN DE UN CONJUNTO LIMITADO DE CAMPOS DE DATOS SIN AMBIGÜIDADES.** Por el contrario, un archivo global que se esfuerza por conseguir menos es considerablemente más probable y factible. El propósito original de publicar fechas de radiocarbono en revistas y listas monográficas era poner a disposición del público un conjunto estándar de datos. Este estándar fue efectivamente el resultado de los laboratorios que hicieron las mediciones sobre las muestras arqueológicas. Aunque con el desarrollo de la tecnología AMS y la proliferación de empresas comerciales que se encargan de hacer las mediciones radiométricas estos campos se han expandido un poco, hay alrededor de 6 campos que son comunes a cada fecha ^{14}C y relativamente estandarizados en todo el mundo. Si a esto se añade un simple conjunto de campos de procedencia (por ejemplo, las coordenadas geográficas del sitio) y de procedencia (por ejemplo, las citas), emerge una estructura conceptual ampliamente compartible. Esto es efectivamente lo que CARD ha presentado y puede explicar su popularidad más allá de Norteamérica. Si bien esta posibilidad resuelve muchos de los desafíos a los que se enfrenta un archivo global, se enfrenta a un obstáculo considerable: un archivo reducido es menos útil para los investigadores locales, cuya participación es necesaria para que el archivo sea completo. Ningún sistema global puede compilar de manera asequible todos los datos del ^{14}C que se han producido, se están produciendo y se producirán. Por lo tanto, debe basarse en el crowdsourcing: la presentación de datos por parte de investigadores locales. Aunque hay un pequeño subconjunto de investigadores que compilan y modelan grandes grupos de datos, la mayoría de los arqueólogos están interesados en sus contextos locales. Para ello, se benefician de los mismos tipos de datos que un modelo reducido no puede acomodar. Como resultado, es menos probable que se recurra a un sistema de este tipo. Por lo tanto, un modelo reducido crea una base de datos global funcional, pero en una forma que es menos útil para los arqueólogos, que probablemente volverían a utilizar simplemente sus propias bases de datos localizadas, y por lo tanto es menos probable que sean exhaustivas.
- **MODELO SINTÉTICO.** Existe una tercera opción que combina elementos de las versiones amplia y reducida, equilibrando las necesidades del investigador/a local y los requisitos de estandarización de una plataforma global: una jerarquía de archivos conectados (Figura 3). Este modelo es esencialmente un sistema de dos niveles con una serie de modelos amplios (exhaustivos) enfocados geográfica o temáticamente que introducen datos básicos en un modelo delgado global. La programación podría actualizar automáticamente el archivo global a medida que se introducen los datos en las plataformas locales. CARD y RPED actualmente emplean una versión de esta relación, donde los datos de la Radiocarbon Palaeolithic Europe Database se cargan manualmente a CARD cada 6 meses. Esto permite a RPED centrarse en los ricos campos de datos en los que Vermeersch ha sido pionero, a la vez que compila un conjunto más pequeño de campos de datos para una audiencia de investigación más amplia. Lo ideal sería que este modelo mancomunara las plataformas de programación. CARD hace esto con sus sitios hermanos *Ancient Maize Map* (AMM: en.ancientmaize.com) y *Ancient Cacao Map* (ACM: en.ancientcacao.com), desarrollados por Michael Blake y alojados en el

Laboratorio de Arqueología de la Universidad de la Columbia Británica (Canadá). Estas plataformas son esencialmente clones entre sí, permitiendo que las actualizaciones de programación desarrolladas en una de ellas sean porteadas a las otras de forma económica. Prevemos que un sistema de este tipo permitiría a los investigadores evaluar rápidamente y acceder a la disponibilidad de datos dentro del área de interés a través de una serie de campos de datos limitados pero estandarizados y luego, si es necesario o deseado, recurrir a las bases de datos más específicas a nivel local que alimentan la interfaz global general.

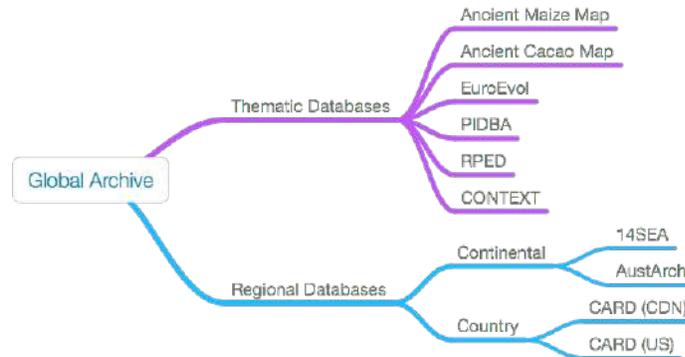


Figura 3. Una posible estructura organizativa de una base de datos mundial de 14C basada en una serie de plataformas regionales y temáticas.

6. Conclusiones

El reto fundamental de un archivo global 14C no es una limitación de la programación o de la infraestructura digital, sino la reconciliación de los intereses contrapuestos de un propósito global y de las comunidades de investigación locales. El primero se beneficia de estándares de datos consistentes, que se logran más fácilmente limitando un archivo a un pequeño número de campos de datos definidos sin ambigüedades. En cambio, esta última es útil porque proporciona datos detallados, a menudo en forma sintáctica y taxonómica, que es la más útil para las comunidades locales de investigación. La tentación es elegir uno de estos modelos como marco para un archivo global, pero argumentamos que esto resultaría impopular para los usuarios o demasiado engorroso para su implementación efectiva. En cambio, creemos que las dos formas de base de datos pueden combinarse de manera realista para hacer avanzar la ambición de un archivo global ¹⁴C completo y útil.

Esta síntesis vincularía a los archivos locales en una red más amplia, algo que se podría hacer avanzar de manera útil con una programación compartida. Como un paso hacia esto, los clones del shell de programación de CARD están disponibles en el Laboratorio de Arqueología de la UBC a precio de costo. Aunque todavía no se ha logrado un archivo global completo y funcional del ¹⁴C, es posible gracias a un equilibrio de intereses, al consenso sobre un conjunto básico de campos de datos y a cierta coordinación entre las plataformas locales. Esta plataforma sería indispensable para la comunidad arqueológica, ya que elevaría el nivel de los informes de datos, facilitaría el acceso a los mismos y ofrecería funciones útiles para los investigadores.

BIBLIOGRAFÍA

Chaput, M. A., B. Kriesche, M. Betts, A. Martindale, R. Kulik, V. Schmidt and K. Gajewski, 2015, Spatiotemporal distribution of Holocene populations in North America. *Proc Natl Acad Sci U S A* 112(39):12127-12132.

Gajewski, Konrad, Samuel E. Munoz, Mathew C. Peros, Andre Viau, Richard Morlan and M. Betts, 2011, The Canadian Archaeological Radiocarbon Database (CARD): Archaeological 14C Dates in North America and their Paleo-Environmental Context. *Radiocarbon* 53(2):371-394.

Jones, Nikola. 2017, World's largest hoard of carbon dates goes global. Nature Publishing Group, Nature News, <http://www.nature.com/news/world-s-largest-ward-of-carbon-dates-goes-global-1.22287>

Morlan, Richard, 1999, Canadian Archaeological Radiocarbon Database: Establishing Conventional Ages. *Canadian Journal of Archaeology* 23(1/2):3-10.

Peros, Matthew C., Samuel E. Munoz, Konrad Gajewski and André E. Viau, 2010, Prehistoric demography of North America inferred from radiocarbon data. *Journal of Archaeological Science* 37(3):656-664.

Vermeersch, Pierre M., 2005, European population changes during Marine Isotope Stages 2 and 3. *Quaternary International* 137(1):77-85.

Wood, Rachel, 2015, From revolution to convention: the past, present and future of radiocarbon dating. *Journal of Archaeological Science* 56:61-72.

Woodbridge, Jessie, Ralph M. Fyfe, Neil Roberts, Sean Downey, Kevan Edinborough and Stephen Shennan, 2014, The impact of the Neolithic agricultural transition in Britain: a comparison of pollen-based land-cover and archaeological 14C date-inferred population change. *Journal of Archaeological Science* 51:216-224.