

This is a repository copy of *Pensamiento Integrado: Agregación de conjuntos de datos arqueológicos a escala internacional*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/209100/>

Version: Published Version

Article:

Richards, Julian D orcid.org/0000-0003-3938-899X (2023) *Pensamiento Integrado: Agregación de conjuntos de datos arqueológicos a escala internacional*. Revista Del Museo De Antropología. pp. 451-70.

<https://doi.org/10.31048/1852.4826.v16.n2.42861>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, as long as you credit the authors and license your new creations under the identical terms. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Pensamiento Integrado: Agregación de conjuntos de datos arqueológicos a escala internacional

Joined up Thinking: Aggregating archaeological datasets at an international scale

Julian D. Richards

University of York. ORCID: <https://orcid.org/0000-0003-3938-899X>

Este trabajo se publicó originalmente como: Richards, J.D. 2023 Joined up Thinking: Aggregating archaeological datasets at an international scale, *Internet Archaeology* 64. <https://doi.org/10.11141/ia.64.3> Esta versión es obra derivada.

Traducción: Andrés D. Izeta y G. Roxana Cattaneo

Resumen

La comunidad de investigadores en arqueología fue una de las primeras en adoptar herramientas digitales para la adquisición, organización, análisis y presentación de resultados de proyectos individuales. Sin embargo, la provisión de infraestructura electrónica y servicios para compartir, descubrir, acceder y reutilizar datos ha estado rezagada. Esta situación ha sido abordada por la Infraestructura de Investigación ARIADNE. Desarrollada con financiamiento europeo, ARIADNE ha creado una infraestructura electrónica que permite a los proveedores de datos registrar y proporcionar acceso a sus recursos digitales a través del portal de datos de ARIADNE, facilitando el descubrimiento, acceso e investigación. ARIADNE ha agregado recursos de más de 45 proveedores de datos, abarcando más de 40 países y 4 continentes. El portal ahora proporciona acceso en línea a más de 3.9 millones de recursos de investigación. Se basa en tecnologías de Linked Open Data y tiene una arquitectura flexible y extensible, lo que permite múltiples combinaciones y presentaciones de los mismos datos subyacentes. Hemos tenido cuidado de no "crear un gran montón" de datos y, aprendiendo de proyectos previos de agregación de datos, hemos definido un subconjunto de estándares del CIDOC CRM para ser utilizados como una ontología estricta y prestamos atención a estándares de datos y vocabularios controlados para lograr un alto grado de interoperabilidad. Este artículo analiza algunos de los desafíos de esa integración de datos a gran escala y describe los enfoques adoptados para asegurar que la Base de Conocimientos de ARIADNE sea una herramienta efectiva para la gestión e investigación del patrimonio arqueológico a nivel nacional e internacional.

Palabras clave: Agregación de datos; Interoperabilidad; Infraestructuras de investigación; Big Data.

Abstract

The archaeological research community was an early adopter of digital tools for data acquisition, organization, analysis, and presentation of research results of individual projects. However, the provision of e-infrastructure and services for data sharing, discovery, access, and re-use has lagged behind. This situation has been addressed by the ARIADNE Research Infrastructure. Developed with European funding, ARIADNE has created an e-infrastructure that enables data providers to register and provide access to their digital resources through the ARIADNE data portal, facilitating discovery, access, and research. ARIADNE has aggregated resources from over 45 data providers, spanning over 40 countries and 4 continents. The portal now provides online access to over 3.9 million research resources. It is based upon Linked Open Data technologies and is underpinned by a flexible and extensible architecture, enabling multiple combinations and presentations of the same underpinning data. We have been keen not to "make a great heap" of all the data and, learning from previous data aggregation projects, we have defined a subset of the CIDOC CRM to be used as a strict ontology and paid close attention to data standards and controlled vocabularies to achieve a high degree of interoperability. This paper discusses some of the challenges of large-scale data integration and describes the approaches adopted to ensure that the ARIADNE Knowledge Base is an effective tool for archaeological heritage management and research at a national and international level.

Keywords: Data Aggregation; Interoperability; Research Infrastructures; Big Data.

Recibido 23-10-2023. Recibido con correcciones 24-11-2023. Aceptado 05-12-2023

Revista del Museo de Antropología 16 (3): 451-470 /2023 / ISSN 1852-060X (impreso) / ISSN 1852-4826 (electrónico)
<http://revistas.unc.edu.ar/index.php/antropologia/index>

IDACOR-CONICET / Facultad de Filosofía y Humanidades – Universidad Nacional de Córdoba - Argentina



Antecedentes: La Prehistoria de la agregación Datos Arqueológicos

En 2013, no había más de seis repositorios digitales que ofrecieran instalaciones para la preservación a largo plazo y el acceso a datos arqueológicos a nivel nacional (Richards, 2017, 233). En la última década, esta situación ha mejorado, aunque todavía es irregular. Una encuesta de 2021, realizada bajo los auspicios de ARIADNEplus y la Acción COST SEADDA, se recopiló información sobre 60 repositorios, de los cuales 43 estaban operativos y 17 se estaban configurando (Geser et al., 2022). No obstante, todavía hay solo una pequeña cantidad de repositorios de datos arqueológicos que tienen la acreditación CoreTrustSeal (ADS, DANS, SND, tDAR y ARCHE). La encuesta sobre el estado del arte realizada por SEADDA y el Consejo Arqueológico Europeo revelaron un panorama muy variado con falta de mandatos de depósito exigibles, así como de instalaciones para su almacenamiento digital en la mayoría de los países (Richards et al., 2021). Sin embargo, sigue siendo claro que, en el ámbito del patrimonio, las infraestructuras nacionales son el nivel adecuado para la curación de datos. Esto es menos directo en las ciencias físicas y químicas, donde los datos se recopilan a través de infraestructuras digitales internacionales compartidas (como el CERN) o donde los datos se relacionan con fenómenos que trascienden el planeta Tierra (como en astrofísica), entonces los repositorios nacionales tienen poco sentido.

En el sector del patrimonio, sin embargo, los datos generalmente se recopilan dentro de marcos legales y pautas nacionales, y se relacionan con características que tienen coordenadas geoespaciales fijas que existen dentro de las fronteras políticas. Existen excepciones en algunos países federales, como Alemania, donde la protección del patrimonio opera a nivel estatal (por ejemplo, Bibby, 2021), y en los Estados Unidos (Nicholson, Fernández e Irwin, 2021), aunque en este último, los principales patrocinadores del trabajo que crean información en arqueología tienden a ser agencias federales nacionales. Pero en general, la arqueología se organiza a nivel estatal nacional, y las agencias nacionales de patrimonio toman la delantera en estipular normas para el registro de los datos y dirección de marcos y prioridades de investigación, incluso si son aplicadas por organismos regionales. En el sector universitario y de investigación, la mayoría de los financiamientos están sujetos a reglas nacionales sobre depósito de datos y acceso abierto, aunque estos pueden reflejar tendencias internacionales. Además, vale la pena señalar que la sostenibilidad de los repositorios digitales capaces de manejar la amplia variedad de tipos y formatos de datos arqueológicos requiere una masa crítica de financiamiento, personal y experiencia que en muchos países no es factible a nivel regional. Los repositorios también necesitan respaldo organizativo y compromiso a largo plazo, y si sus recursos también van a ser “confiables” en el sentido más convencional,

entonces es esencial alguna huella institucional acreditada.

Ciertamente, existe el riesgo de que una red de repositorios que operan a nivel nacional cree una serie de “silos” de datos. De hecho, los repositorios digitales deben hacer dos cosas: deben reunir recursos y facilitar a los usuarios su interrogación a través de interfaces compartidas y fáciles de usar; pero también deben abrir los datos: mediante API, protocolos de cosecha como OAI-PMH y como Linked Open Data (LOD) para que puedan ser vistos y manipulados a través de múltiples rutas (May et al., 2015). También es importante reconocer las preocupaciones planteadas por Huggett (2015) de que al agregar datos sacamos los datos fuera del contexto en el que se recopilaron en primer lugar, y en ausencia de parámetros que describan ese contexto, corremos el riesgo de un mal uso. A medida que agregamos datos a un nivel cada vez más alto, los riesgos aumentan, ya que podemos separar los datos de los valores culturales y sociales en los que se registraron. Los estándares de metadatos que empleamos generalmente fueron producidos por arqueólogos educados en una tradición positivista occidental, y debemos prestar atención tanto a los principios de datos FAIR como a los de CARE para evitar imponer inadvertidamente perspectivas postcoloniales. En este sentido, el trabajo del proyecto TETRARCHs en explorar otros tipos de metadatos para fomentar diferentes formas de reutilización de datos es importante. Sin embargo, las preguntas de investigación arqueológica rara vez coinciden con las fronteras políticas modernas, que fueron irrelevantes durante la gran mayoría del lapso de tiempo del pasado humano. Grandes preguntas de investigación requieren grandes datos, y es nuestro papel catalogarlos con sensibilidad y teniendo en cuenta todos los matices de la observación humana y las diferentes perspectivas.

Tan temprano como en 1992, Henrik Jarl Hansen expresó la necesidad de unir recursos digitales en toda Europa (Figura 1) (Hansen, 1992), y la Comisión Europea ha sido una agencia de financiamiento clave que ha facilitado varios proyectos en esta área. El primero fue AQUARELLE (1996-99), que diseñó una arquitectura de descubrimiento de recursos que permite la difusión de consultas a fuentes de datos distribuidas y heterogéneas de museos y el dominio del patrimonio (Michard et al., 1998). El proyecto colaboró con el consorcio CIMI de museos y bibliotecas estadounidenses para utilizar un perfil de aplicación Z39.50 para habilitar la difusión de consultas y un sistema de software para crear y gestionar recursos terminológicos multilingües.

Entre 2002 y 2004, el Servicio de Datos de Arqueología (ADS) lideró un consorcio de socios europeos en el proyecto ARENA, financiado en el marco del programa Cultura 2000 (Kenny y Richards, 2005). Uno de los resultados del proyecto fue un portal que proporcionaba,



Figura 1: Pulpo de Hansen simbolizando la interconexión de recursos europeos (reproducido de Hansen, 1992).

Figure 1: Hansen's octopi symbolising the networking of European resources (reproduced from Hansen, 1992).

a través de una búsqueda cruzada, registros de sitios y monumentos para seis países: Dinamarca, Islandia, Noruega, Polonia, Rumania y el Reino Unido (Figura 2) (Dam y Hansen, 2005). El portal ARENA se completó en 2004 y se lanzó en la conferencia de la Asociación Europea de Arqueólogos en Lyon.

ARENA proporcionó un prototipo de prueba para un

paradigma de búsqueda que luego adoptaría ARIADNE, basado en un poderoso trío de parámetros: 'Dónde', 'Cuándo' y 'Qué', permitiendo a los usuarios identificar recursos de interés en unos pocos clics. Los socios proporcionaron asignaciones de términos de nivel superior al Tesoro de Tipos de Monumentos MIDAS del Reino Unido y utilizaron sus vocabularios controlados locales para términos de período cronológico (Figura 3). Sin embargo, no se intentó proporcionar una búsqueda temporal utilizando intervalos de tiempo absolutos como Desde y Hasta.

El portal ARENA demostró que era posible realizar búsquedas en línea en múltiples proveedores de datos europeos, pero dependía de tecnologías obsoletas como el protocolo de comunicaciones Z39.50, que se había desarrollado para la búsqueda cruzada en catálogos de bibliotecas. La misma tecnología se había implementado inicialmente para el portal del Reino Unido del catálogo del Servicio de Datos de Artes y Humanidades a fines de la década de 1990 y había sido desarrollada aún más por el Archaeological Data Service (ADS) para el portal del Reino Unido HEIRPORT para Recursos de Información del Entorno Histórico (Austin et al., 2002). En 2009-10, el ADS pudo trabajar con DANS como parte de la fase preparatoria de DARIAH, la Infraestructura de Investigación Digital para las Artes y Humanidades, para migrar el portal ARENA a una arquitectura de servicios web más flexible (Figura 4). ARENA2 cambió el enfoque técnico de ARENA a uno basado en el uso de una Arquitectura Orientada a Servicios (SOA) sobre centros de datos de socios seleccionados para demostrar la viabilidad de utilizar un enfoque SOA en lugar del enfoque original de Z39.50 y la recolección de metadatos de OAI. Un enfoque SOA difiere de los enfoques de recolección de metadatos en que permite el acceso directo y en línea a bases de datos remotas, poniendo los datos más

Figura 2: La página de inicio del portal de búsqueda ARENA, desarrollado durante el proyecto ARENA 2002-2004, proporcionaba una búsqueda cruzada distribuida de catálogos de datos de sitios y monumentos para Dinamarca, Islandia, Noruega, Polonia, Rumania y el Reino Unido.

Figure 2: The landing page of the ARENA search portal, developed during the ARENA project 2002-2004, provided a distributed cross-search of catalogues of sites and monuments data for Denmark, Iceland, Norway, Poland, Romania and the United Kingdom.



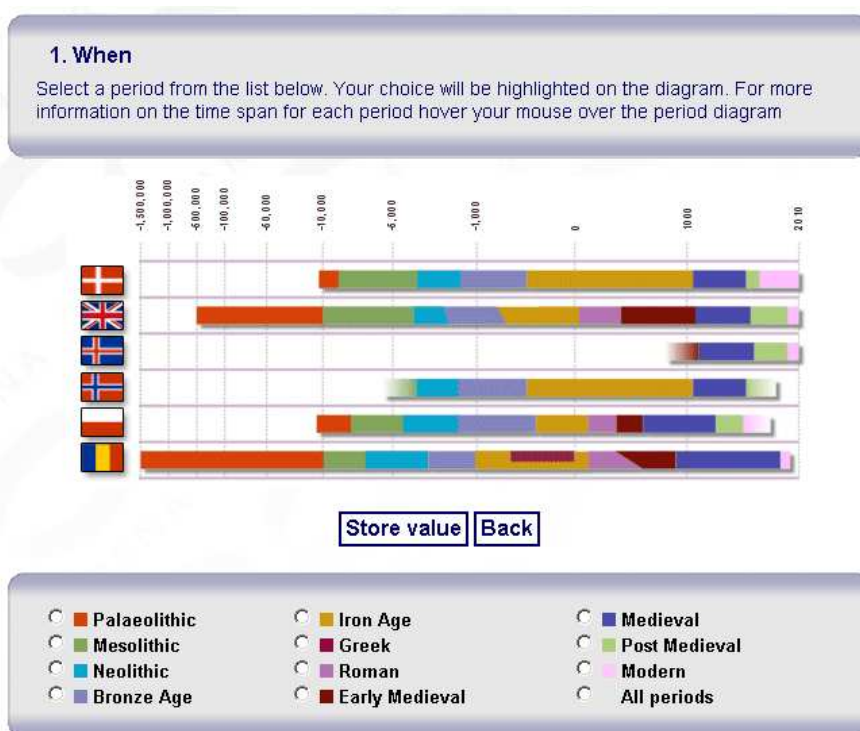


Figura 3: La búsqueda temporal de ARENA proporcionó una visualización de las diferentes fechas de inicio y finalización de los términos de período cultural aplicados por cada uno de los socios, y los usuarios podían seleccionar todos los recursos para los cuales se utilizaba el mismo término de período, pero no eran posibles búsquedas utilizando rangos de fechas absolutas.

Figure 3: The ARENA temporal search provided a visualisation of the different start and end dates for cultural period terms applied by each of the partners and users could select all resources for which the same period term was used, but searches using absolute date ranges were not possible.

actuales disponibles para ser consultados y minimizando la recolección y gestión de datos por parte del agregador.

La misma infraestructura tecnológica también se utilizó en un proyecto colaborativo entre el ADS y tDAR para construir un Portal de Arqueología Transatlántica (TAG) como buscador de archivos de datos del Reino Unido y Estados Unidos (Jeffrey et al., 2012). La primera etapa fue crear una infraestructura para permitir la búsqueda básica de registros de metadatos compatibles con Dublin Core para recursos digitales que abarcan la arqueología de Estados Unidos y el Reino Unido, con los desafíos adicionales en relación con la periodización y el tipo de tema (Figura 5). La segunda etapa de TAG fue un intento de desarrollar un nivel mucho más profundo y rico de búsqueda cruzada para un subconjunto específico de datos archivados: datos arqueofaunísticos de América del Norte y Europa. Esta subdisciplina se eligió porque hay un nivel relativamente alto de acuerdo sobre las clasificaciones básicas. El proyecto proporciona un ejemplo temprano del enfoque de perfiles de aplicación que se adoptó posteriormente en ARIADNEplus.

Estos primeros demos proporcionaron ejemplos de la escala de las preguntas de investigación que la tecnología podría respaldar. Sin embargo, todos fueron financiados como proyectos de investigación y desarrollo, y aunque hubo apoyo continuo para los repositorios subyacentes a nivel nacional, no hubo una fuente de financiamiento obvia para la infraestructura internacional, y ninguno se encontraba dentro de un marco sostenible. La excepción potencial era Europeana, una iniciativa para reunir de manera virtual las ricas colecciones de museos, galerías

y archivos de Europa. El modelo de datos inicial, los Elementos Semánticos de Europeana (ESE), se basaba en el esquema de metadatos Dublin Core y su tratamiento de la cobertura temporal y espacial era inadecuado para el dominio arqueológico (Hansen y Fernie, 2010). En 2010, esto se actualizó al Modelo de Datos de Europeana (EDM), un esquema semánticamente más rico que tiene en cuenta los principios basados en eventos que respaldan el Modelo de Referencia Conceptual CIDOC (CIDOC CRM).

Se financiaron varios proyectos de agregadores y algunos de ellos son discutidos por Vassallo et al. (2023), pero aquellos con un componente arqueológico importante fueron CARARE (2010-13) y LoCloud (2013-16). CARARE se centró en arqueología y arquitectura, con un énfasis adicional en datos 3D (Hansen y Fernie, 2010), mientras que LoCloud enfatizó el contenido local de museos pequeños y medianos. Ambos proyectos adoptaron un subconjunto de la ontología CIDOC CRM como su modelo de metadatos central, aunque solo un subconjunto de esto podía exponerse a través de Europeana. Aunque el portal de Europeana ahora proporciona acceso a más de 50 millones de elementos, su cobertura es tan amplia y su interfaz de búsqueda tan generalizada que es inadecuado para la mayoría de la investigación arqueológica, y no es muy conocido entre los arqueólogos. Como destacó Kilbride (2004): 'la tecnología no es, por sí sola, la respuesta'.

También hay un creciente interés en los Datos Enlazados Abiertos en arqueología, proporcionando mecanismos para la agregación de datos, incluyendo servicios como Pleiades (un gazetteer comunitario de lugares antiguos)

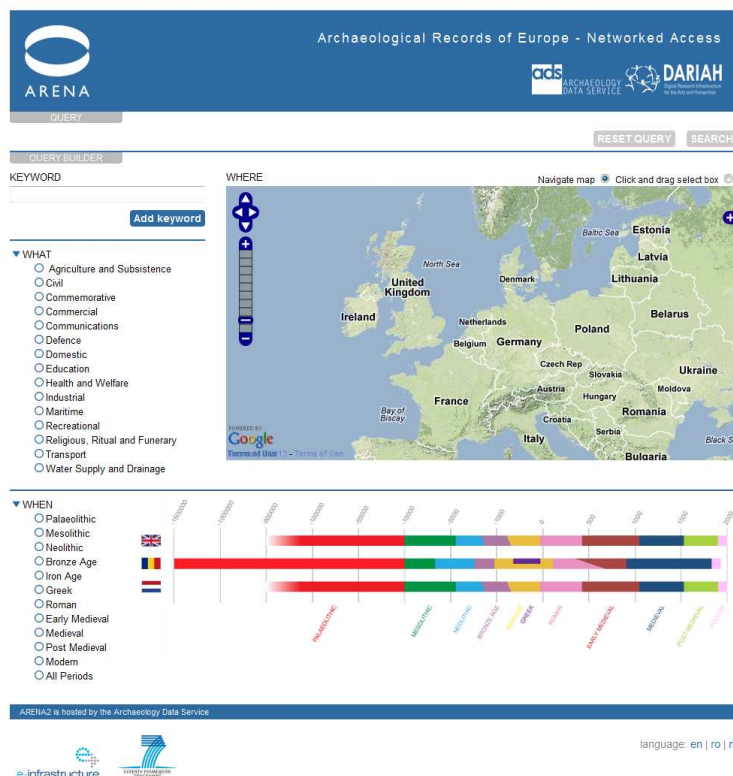


Figura 4: El portal ARENA2, desarrollado como un prototipo técnico para DARIAH.

Figure 4: The ARENA2 portal: developed as a technical demonstrator for DARIAH.

y Pelagios (también conectando lugares en el mundo clásico), que se centran en enfoques geoespaciales para la interoperabilidad de datos (Isaksen et al., 2014). Asimismo, debemos tener en cuenta la implementación de enfoques de Datos Enlazados Abiertos por parte de Open Context (Kansa, 2014). Se presenta como una plataforma de publicación de datos basada en la web en lugar de un agregador de datos, aunque en 2012 se utilizó la misma arquitectura técnica para desarrollar la plataforma Digital Index of North American Archaeology (DINAA), que agrega conjuntos de datos arqueológicos e históricos en su mayoría recopilados de agencias gubernamentales estatales y federales, y proporciona acceso a más de 880,000 registros a través de una interfaz basada en mapas. La naturaleza distribuida de los Datos Enlazados Abiertos también lo convierte en un mecanismo adecuado para la agregación de datos que abarcan varios dominios, con datos gestionados y curados por especialistas en el dominio y la creación de interfaces adaptadas que reflejen conjuntos específicos de necesidades de los usuarios. Varios proyectos financiados dentro del programa Towards a National Collection del Reino Unido han explorado el uso de Datos Enlazados Abiertos (LOD) como la única solución viable para proporcionar vistas de datos nacionales.

ARIADNE proporciona un nuevo capítulo en la historia de la agregación de datos arqueológicos y es el enfoque principal de este artículo. Financiado por el programa de infraestructura de investigación de la Comisión Europea, tuvo como objetivo proporcionar acceso unificado

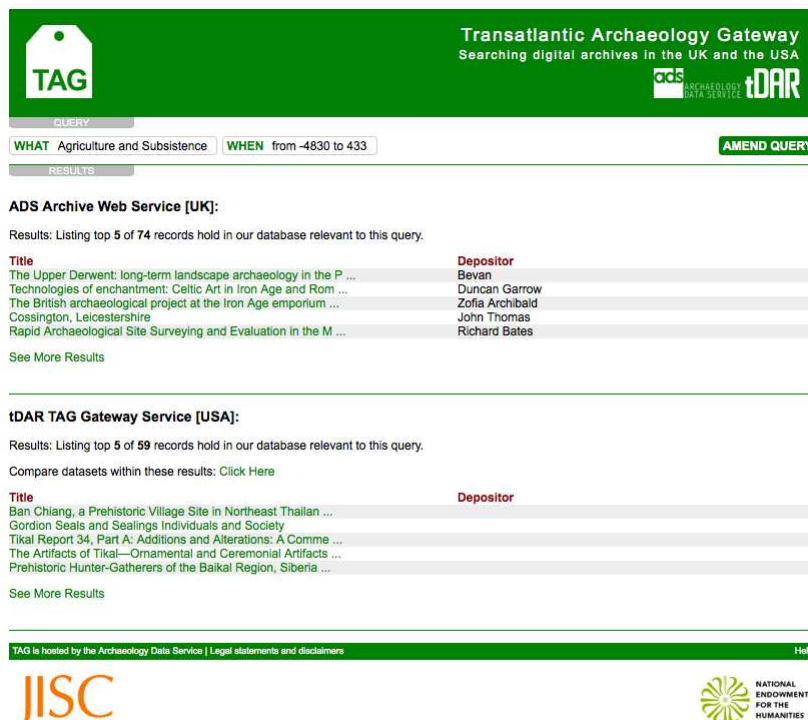
a datos arqueológicos procedentes de varios países de una manera que permitiera a los investigadores abordar nuevas preguntas. ARIADNE buscó aprender de lecciones pasadas al regresar al enfoque arqueológico de ARENA y enfatizar un proceso de mapeo impulsado por las necesidades de los usuarios de investigación arqueológica. Ha desarrollado un potente portal de búsqueda cruzada, basado en Datos Enlazados Abiertos (Aloia et al., 2017; Meghini et al., 2017; Richards y Niccolucci, 2019). Lograr la interoperabilidad entre diferentes idiomas y culturas europeos es un desafío, y la adhesión a estándares de datos es esencial para cualquier nivel de interoperabilidad semántica y búsqueda cruzada. En ARIADNE, los términos nacionales del tema se han mapeado a un estándar central común (en este caso, el Tesoro de Arte y Arquitectura Getty) y los términos arqueológicos de período se han definido según criterios explícitos, trabajando con la iniciativa norteamericana PeriodO (Shaw et al., 2016).

Agregación de Datos en ARIADNE

ARIADNE ha agregado recursos de más de 45 proveedores de datos, abarcando más de 40 países y cuatro continentes. A partir de mayo de 2023, el portal de ARIADNE proporciona acceso en línea a más de 3.9 millones de recursos de investigación, y este número sigue aumentando. El portal se basa en un almacén triple gestionado en Graph DB. En el momento de la redacción, el triple store (almacén de RDF) utilizado por el portal público consta de 322,218,928 triples.

Figura 5: El Portal de Arqueología Transatlántica (2009-11), mostrando una búsqueda cruzada en los archivos del ADS y tDAR.

Figure 5: *The Transatlantic Archaeology Gateway (2009-11), showing a cross search across the archives of the ADS and tDAR.*



La Infraestructura de Investigación ARIADNE se desarrolló durante el primer proyecto ARIADNE, con financiamiento de la Comisión Europea, en el marco del Programa Marco 7 para el período 2013-2017. El objetivo de ARIADNE era proporcionar acceso abierto al patrimonio arqueológico de Europa y superar la fragmentación de los repositorios digitales, ubicados en diferentes países y compilados en diferentes idiomas (Niccolucci y Richards, 2019). Técnicas innovadoras de mapeo de vocabulario aportaron interoperabilidad a una colección enorme y heterogénea de textos, dibujos, imágenes, videos, modelos 3D y mapas. Al final del proyecto ARIADNE original, habíamos logrado catalogar aproximadamente 1.9 millones de

conjuntos de datos arqueológicos, utilizando el Modelo de Datos Común de ARIADNE (ACDM) (Figura 6) (Aloia et al., 2017).

En 2019, comenzamos el trabajo en la segunda iteración de ARIADNE, con un consorcio ampliado y financiamiento de Horizon Europe. El modelo de datos fue revisado y simplificado, adoptando también algunas lecciones del proyecto PARTHENOS, y se lanzó como el AO-Cat (Felicetti et al., 2023). Nos interesaba no “hacer un gran montón+” de todos los datos y, aprendiendo de proyectos anteriores de agregación de datos, definimos una ontología revisada, el AO-Cat, como un subconjunto

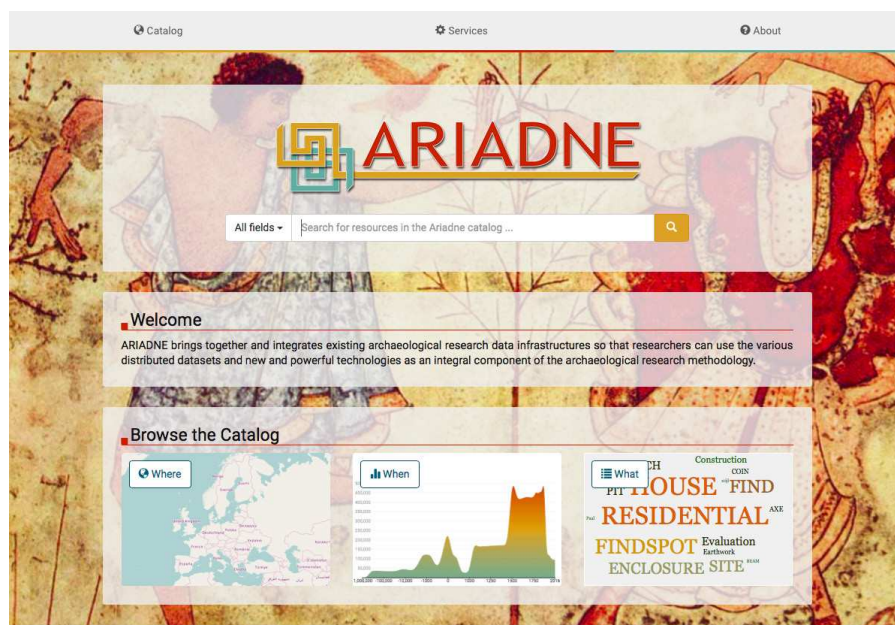


Figura 6: Captura de pantalla del portal original de ARIADNE. La página de inicio proporcionaba un cuadro de búsqueda de texto libre y la oportunidad de explorar el catálogo mediante los tres parámetros clave de Dónde, Cuándo y Qué.

Figure 6: *Screen shot of the original ARIADNE portal. The landing page provided a free text search box and the opportunity to browse the catalogue by the three key parameters of Where, When and What.*

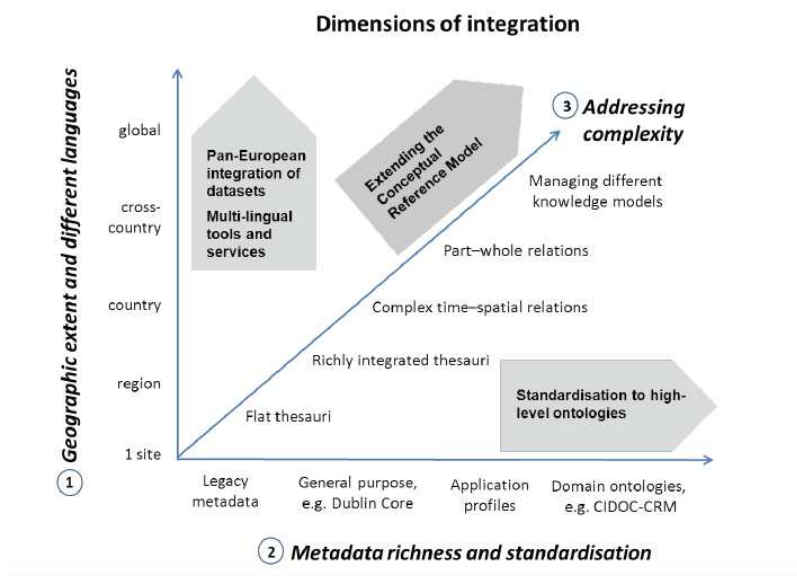


Figura 7: El enfoque de ARIADNE para abordar la complejidad del desarrollo de servicios integrados avanzados para las comunidades de investigación arqueológica.

Figure 7: *The ARIADNE approach for addressing the complexity of developing advanced integrated services for archaeological research communities.*

del CIDOC CRM como una ontología estricta, y también prestamos mucha atención a los estándares de datos y vocabularios controlados para lograr un alto nivel de interoperabilidad. El catálogo se migró de una estructura de base de datos tradicional a un triple store de Linked Open Data, evitando la creación de otro silo de datos y abriendo la Base de Conocimientos para la vinculación desde y hacia otros recursos en línea.

ARIADNEplus ha reunido un conjunto de registros mucho más amplio que el proyecto predecesor ARIADNE, sobre cuyo éxito ha podido construir (Figura 7). La intención fue integrar conjuntos de datos que abarcaran ampliamente en espacio y tiempo, y a través de un rango más amplio de subdominios arqueológicos. Al final del proyecto financiado, tenemos 7,710 recursos que datan de antes de 1,000,000 aC, y 517,535 recursos que datan después del año 1900 dC, demostrando el alcance completo de la investigación arqueológica del siglo XXI. Tenemos recursos de América del Norte y del Sur, Europa y Escandinavia, y Asia. Los datos integrados cubren todos los tipos de recursos de ARIADNE, desde inscripciones y arte rupestre hasta análisis científicos y recursos de datación. Hemos integrado más de un millón de registros para sitios y monumentos individuales, más de 855,250 artefactos y 470,000 monedas. Nuestro catálogo incluye más de 290,000 informes de trabajo de campo inéditos y más de 110,000 archivos completos de trabajo de campo. Estos representan una enorme infraestructura de investigación digital ahora disponible de manera gratuita y fácil como datos de acceso abierto para científicos e investigadores europeos, profesores y estudiantes, así como para miembros interesados del público.

La tarea de agregación será ahora sostenida a través del trabajo de la ARIADNE Research Infrastructure AISBL, una asociación sin fines de lucro registrada bajo la ley belga, pero que opera a nivel internacional. Actualmente, opera bajo un modelo de suscripción mediante el cual los

miembros pueden pagar una tarifa anual para actualizar y agregar sus recursos en línea. En este sentido, es importante señalar que ARIADNE es tanto una comunidad como una infraestructura técnica. Las infraestructuras de investigación en línea pueden actuar como catalizadores para reunir a diferentes comunidades de experiencia e interés, y ARIADNE ha creado un ecosistema académico, así como un portal (ver Benardou et al. 2017; Wright y Richards 2018).

La ontología ARIADNE

La Ontología ARIADNE, AO por sus siglas, fue desarrollada al inicio del proyecto ARIADNEplus con el propósito de integrar los datos arqueológicos de los socios de ARIADNE en un espacio de información común. Basándose en proyectos anteriores, el CIDOC CRM, la ontología estándar en el dominio del patrimonio cultural, fue adoptado como la columna vertebral conceptual de AO, proporcionando un marco lingüístico y axiomático unificado y coherente. Posteriormente, el CRM se especializó, bajo un espacio de nombres especialmente diseñado para satisfacer las necesidades de los diferentes subdominios abordados por el proyecto ARIADNEplus. La primera de esas especializaciones fue el AO-Cat, la parte de AO que se ocupa de la representación de los recursos en el Catálogo ARIADNE. El AO-Cat se utilizó para construir el Catálogo ARIADNE. La segunda ronda de especializaciones se conoce como perfiles de aplicación, que se han desarrollado para respaldar la integración de datos más especializados a nivel de ítem de los socios de ARIADNE en la Base de Conocimientos ARIADNE.

El Catálogo ARIADNE proporciona una representación a nivel de colección de los recursos compartidos por los socios del proyecto. Sin embargo, aunque la distinción entre nivel de colección y nivel de ítem puede parecer obvia, se puede aplicar de muchas maneras, según cómo definamos lo que consideramos que son ítems. Esto, a

su vez, depende del contexto de investigación específico y de la pregunta de investigación. Por ejemplo, a nivel de investigación de paisajes, cada sitio arqueológico puede considerarse un ítem de observación, y el registro a nivel de colección se refiere a la base de datos de sitios y monumentos nacionales, mientras que para un estudio basado en artefactos, los objetos individuales pueden ser los ítems de interés, y el registro a nivel de colección se refiere ahora a la base de datos de artefactos, potencialmente de varios sitios.

En la primera fase de ARIADNE, produjimos un catálogo que ya incluía registros de metadatos resumidos para sitios y monumentos arqueológicos a nivel de ítem y este enfoque se mantuvo en ARIADNEplus. También se realizaron experimentos en una integración más granular a nivel de ítem, investigando especialmente la interoperabilidad entre varias bases de datos de monedas mantenidas por diferentes socios (Felicetti et al., 2015). En ARIADNEplus, estos desarrollos se extendieron mediante la adopción de perfiles de aplicación.

Los Perfiles de Aplicación son modelos ontológicos particulares cuyo propósito es describir la información relacionada con un dominio de investigación específico de manera coherente y completa. La filosofía detrás de la definición de un Perfil de Aplicación es la de reutilizar los elementos de ontologías y modelos existentes para la descripción de entidades similares en el nuevo contexto de investigación, limitando el desarrollo de nuevos elementos solo para la descripción de aspectos peculiares y típicos

de la disciplina específica. En ARIADNEplus, por ejemplo, seleccionamos clases de la ontología CIDOC CRM y del estándar Dublin Core para definir perfiles de aplicación para la descripción de datos científicos y datos mortuorios a nivel de ítem. Sin embargo, vale la pena señalar que el AO-Cat es en sí mismo un perfil de aplicación para lo que podríamos considerar observaciones arqueológicas de un nivel superior. Captura la información básica de 'Qué', 'Cuándo' y 'Dónde' que impulsa la interfaz de búsqueda del portal, y proporciona una representación adecuada para el descubrimiento de recursos relacionados con sitios y monumentos arqueológicos. También ha quedado claro que proporciona información básica adecuada para otros subdominios donde 'Qué', 'Cuándo' y 'Dónde' capturan los metadatos y datos principales, como en el caso de artefactos arqueológicos, por ejemplo. En otros casos, puede proporcionar un registro a nivel de colección para clases relacionadas de datos, como análisis científicos, aunque pueda necesitarse un perfil de aplicación para abordar las preguntas de investigación específicas del subdominio. La tarea de crear perfiles de aplicación se delegó a grupos de interés especiales de especialistas en subdominios. Las actividades y las discusiones de los grupos de trabajo han demostrado que, en muchos casos, el AO-Cat es suficiente, sin necesidad de modificaciones adicionales, para cubrir las necesidades de sus subdominios, incluso para la agregación de datos a nivel de ítem. También es notable que la adición de una gran colección de recursos al Catálogo durante la vida útil de ARIADNEplus no ha requerido ningún cambio en la ontología. Esto indica que el AO-Cat ha alcanzado

Vocabulary Matching Tool

English

Source Concept	Label	Match	Target Concept	Suggest	Delete Row
Identifier	Filter column...		Filter column...		
http://purl.org/heritagedata/schemes...	ABATTOIR	en Exact Match	slaughterhouses	Q	
http://purl.org/heritagedata/schemes...	ABBEY	en Exact Match	abbeys (monasteries)	Q	
http://purl.org/heritagedata/schemes...	ABBOTS SUMMER PALACE	en Close Match	summer palaces	Q	
http://purl.org/heritagedata/schemes...	ABLUTIONS BLOCK	en Close Match	rest rooms	Q	
http://purl.org/heritagedata/schemes...	ACADEMY SCHOOL	en Close Match	academies (buildings)	Q	
http://purl.org/heritagedata/schemes...	ACCIDENT HOSPITAL	en Close Match	hospitals (buildings for health facility)	Q	
http://purl.org/heritagedata/schemes...	ACCOMMODATION BRIDGE	en Close Match	bridges (built works)	Q	
http://purl.org/heritagedata/schemes...	ACCOMMODATION HUT	en Close Match	huts (houses)	Q	
http://purl.org/heritagedata/schemes...	ACCUMULATOR HOUSE	en Close Match	power plant buildings (utilities buildi...	Q	
http://purl.org/heritagedata/schemes...	ACETONE FACTORY	en Close Match	factories (structures)	Q	
http://purl.org/heritagedata/schemes...	ACID WORKS	en Close Match	chemical plants	Q	
http://purl.org/heritagedata/schemes...	ACTIVITY CENTRE	en Close Match	athletic clubs	Q	
http://purl.org/heritagedata/schemes...	ADIT	en Close Match	mine shafts	Q	
http://purl.org/heritagedata/schemes...	ADMIRALTY SIGNAL ESTABLISHM...	en Close Match	signal towers (elevated structures)	Q	
http://purl.org/heritagedata/schemes...	ADMIRALTY SIGNAL STATION	en Close Match	signal houses	Q	
http://purl.org/heritagedata/schemes...	ADMISSION HOSPITAL	en Close Match	hospitals (buildings for health facility)	Q	
http://purl.org/heritagedata/schemes...	ADULTERINE CASTLE	en Broad Match	castles (fortifications)	Q	

3979 rows

FIRST PREV 1 2 3 4 5 NEXT LAST

IMPORT JSON EXPORT JSON EXPORT CSV ADD NEW ROW CLEAR ROWS SHOW HELP

ARIADNEplus
Created by University of South Wales

ARIADNEplus is a Horizon 2020 project funded by the European Commission (Grant Agreement No 823914). This application retrieves some information originating from Getty Art & Architecture Thesaurus (AAT)® which is made available under the ODC Attribution License. See <http://vocab.getty.edu/> for further details.

Figura 8: La Herramienta de Coincidencia de Vocabulario, ilustrando la capacidad de definir coincidencias Exactas, Cercanas o Amplias entre el AAT de Getty y los conceptos principales de los vocabularios nativos utilizados por los socios.

Figure 8: The Vocabulary Matching Tool, illustrating the ability to define Exact, Close, or Broad matches between the Getty AAT and target concepts in native vocabularies deployed by partners.

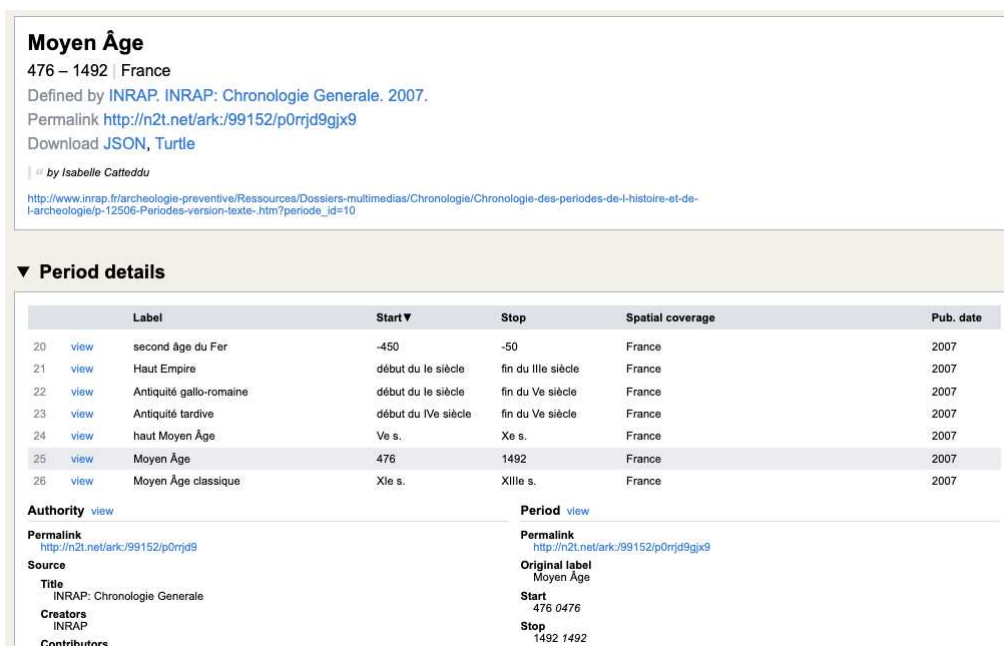


Figura 9: PeriodO: definiciones de periodos para Francia proporcionadas por Inrap.

Figure 9: PeriodO: period definitions for France provided by Inrap.

un nivel maduro de estabilidad y, por lo tanto, está listo para un uso más amplio fuera de los límites del proyecto ARIADNEplus. De hecho, en el Reino Unido, el AO-Cat ha sido adoptado con pocas modificaciones por Unpath'd Waters, un proyecto importante para caracterizar y agregar datos digitales sobre el rico patrimonio marítimo del Reino Unido.

Mapeo de Datos

Habiendo diseñado el AO-Cat como una ontología compartida, la interoperabilidad dependía del mapeo de los modelos de datos de conjuntos de datos heterogéneos de los socios al AO-Cat. Para esto se utilizó la X3ML Toolkit desarrollada por FORTH. La X3ML Toolkit es un conjunto de microservicios pequeños, de código abierto, que siguen el Modelo de Referencia SYNERGY para la provisión y agregación de datos. Nos permitió definir equivalencias entre los diversos atributos en los datos proporcionados por los socios con propiedades en el AO-Cat. Sin embargo, para facilitar una búsqueda cruzada útil, también era esencial lograr la interoperabilidad de los metadatos utilizados para describir las tres facetas principales de Dónde, Cuándo y Qué. Para respaldar una búsqueda cruzada efectiva de metadatos originarios de muchos países diferentes, se requiere cierto entendimiento común compartido del significado de los metadatos. Las facetas 'Dónde' y 'Cuándo' pueden comunicarse utilizando tipos de datos comunes y valores comparables (por ejemplo, coordenadas espaciales relativas a una ubicación conocida en la Tierra, rangos de fechas relativos a una época conocida). La faceta 'Qué' (sujeto) puede ser más difícil de definir en un formato comúnmente entendido y comparable.

Tesaurus de Arte y Arquitectura de Getty

En lugar de intentar mapear cada vocabulario de temas con todos los demás vocabularios de temas, se acordó mapear cada vocabulario a una columna vertebral común (Binding et al., 2015). El Tesaurus de Arte y Arquitectura de Getty (AAT) proporciona conceptos y términos para describir conceptos del patrimonio cultural. El AAT está disponible como Datos Enlazados Abiertos, por lo que cada concepto en el tesaurus tiene un identificador único en forma de un URI (por ejemplo, <http://vocab.getty.edu/aat/300211545> es el identificador del concepto "broches penannulares"). ARIADNE adoptó el AAT como un centro central para la interconexión escalable de vocabularios locales de temas. Se requirió que los proveedores de datos proporcionaran un conjunto de mapeos, desde su propio vocabulario local de temas (todos los términos utilizados para describir temas en sus propios metadatos) hasta el AAT, y se desarrolló una herramienta de mapeo de vocabulario para facilitar este proceso (Figura 8). El objetivo del ejercicio de mapeo de temas era mejorar la recuperación y precisión al navegar y buscar posteriormente los datos por tema. Sin embargo, uno de los problemas a los que se enfrentaron fue que para varios países no existía un vocabulario de temas acordado. En estos casos, los socios a menudo llevaron a cabo un extenso ejercicio de mejora de datos, adoptando el AAT desde el principio. Para otros, esto no fue posible, pero se decidió incluir sus datos en la agregación, siempre y cuando aún fueran encontrables por las facetas de Cuándo y Dónde.

PeriodO

PeriodO es un gacetero multilingüe de definiciones de periodos históricos y arqueológicos para describir

periodos con nombres. Incluye una serie de 'autoridades' (listas de periodos con nombres) donde cada definición de periodo individual está asociada con una área geográfica específica, una fecha de inicio y una fecha de finalización. Los periodos y autoridades tienen identificadores de 'permalink' (URIs) para que puedan ser referenciados de manera clara y unívoca, por ejemplo, el identificador <http://n2t.net/ark:/99152/p0rrjd9gix9> define el periodo 'Moyen Âge' (Edad Media) dentro de la autoridad <http://n2t.net/ark:/99152/p0rrjd9> 'INRAP: Chronologie Generale. 2007' (Figura 9). Todos los periodos con nombres en ARIADNE se han definido con referencia a periodos dentro de PeriodO.

****PeriodO**** en ARIADNE se utiliza en una etapa posterior de enriquecimiento de datos con un propósito diferente al uso del enriquecimiento de datos con AAT; en lugar de alinear datos con identificadores comunes, PeriodO se utiliza en una etapa posterior de enriquecimiento de datos para mejorar los registros que ya se han indexado con periodos con nombres (como "Edad de Bronce", "Edad de Hierro", etc.), donde las fechas asociadas con estos periodos no se hacen explícitas en los datos de entrada, proporcionando fechas de inicio/fin para que estos registros sean comparables y encontrables de la misma manera que otra información de fecha.

Nuevamente, hubo casos en los que los nombres de los periodos no estaban disponibles, aunque a veces los proveedores ya tenían fechas de inicio y fin para sus recursos, por lo que se podía pasar por alto la mejora de PeriodO. En algunos casos, sin embargo, no se disponía de datos temporales, pero los recursos se incluyeron si eran descubribles a través de los otros aspectos fundamentales.

Datos Geográficos

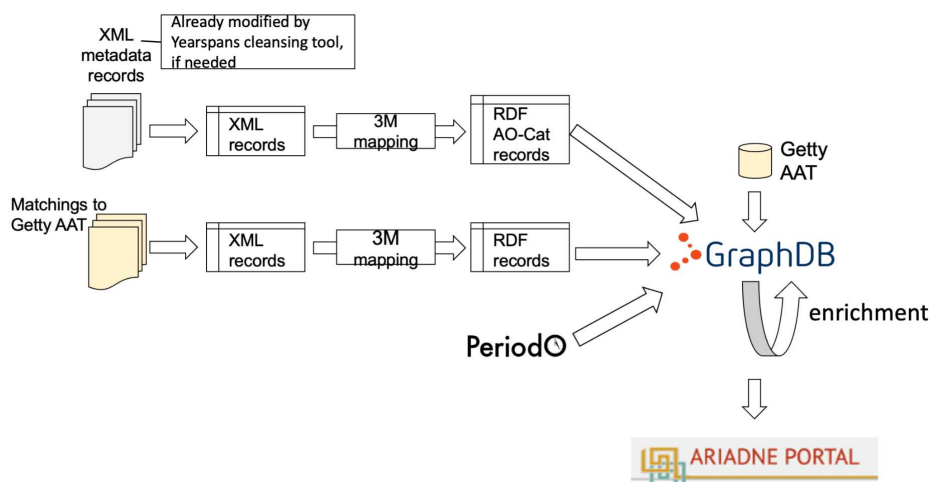
Para abordar la dimensión espacial, todas las coordenadas espaciales en ARIADNE se expresan utilizando el Sistema Geodésico Mundial 1984 (WGS84). Este es un sistema de coordenadas geográficas estándar que comprende un datum horizontal y vertical global, así como un sistema de coordenadas utilizado (especialmente por sistemas de GPS) para expresar la posición global en la superficie de la Tierra. Para los propósitos de ARIADNE, todas las coordenadas espaciales se expresan utilizando WGS84. Las coordenadas en conjuntos de datos locales a veces requerían normalización/transformación antes de la agregación para mejorar las oportunidades de búsqueda cruzada de los conjuntos de datos integrados. Las coordenadas WGS84 se expresan en grados: la posición horizontal es la longitud (con valores entre -180° y +180°, en relación con un datum fijo de 0° en Greenwich, Reino Unido) y la posición vertical es la latitud (con valores entre -90° y +90°, en relación con un datum fijo de 0° en el ecuador). Como ejemplo, la ubicación global de la Torre Inclinada de Pisa (Torre di Pisa) se da mediante las coordenadas WGS84 de latitud 43.723056 (43° 43' 23" N) y longitud 10.396389 (10° 23' 47" E).

El flujo de trabajo de agregación

A los socios se les ofrecieron dos opciones para la agregación: el enfoque estándar que utiliza un conjunto de herramientas para la agregación semiautomática de conjuntos de datos grandes, y un enfoque básico para la carga manual de pequeños números de registros. La mayoría de los socios utilizó el enfoque estándar, pero un pequeño número utilizó FastCat, una herramienta proporcionada para la carga de unos pocos registros.

Figura 10: El flujo de trabajo durante la agregación.

Figure 10: The aggregation workflow



Esto también resultó invaluable para la adición de registros de colecciones personalizadas para recursos recolectados.

Los socios que siguieron el enfoque estándar tuvieron que:

1. Describir sus datos según el AO-Cat utilizando la herramienta X3ML, generalmente con un mapeo, por socio.
2. Mapear los términos de sujetos al Getty AAT utilizando la Herramienta de Coincidencia de Vocabulario.
3. Definir cualquier término temporal utilizado para que se cargue en PeriodO.

Cuando los datos temporales necesitaban limpieza para crear un uso consistente de rangos de fechas y períodos, los socios utilizaron una herramienta adicional, Yearspans, para normalizar los rangos de fechas (Binding y Tudhope, 2023). También tuvieron que asegurarse de que sus datos espaciales fueran compatibles con WGS84.

Los socios que utilizaron FastCat en su lugar introdujeron manualmente sus registros de datos en la herramienta tipo hoja de cálculo, donde los encabezados de las columnas ya correspondían a los campos obligatorios básicos de AO-Cat, de modo que podría haber un solo mapeo X3ML que cubriera a varios socios.

Los proveedores de datos suministraron sus datos a ARIADNE ya sea como volcados de datos XML o JSON o, cuando se requerían actualizaciones regulares, se prefería la alimentación OAI-PMH. Luego, los datos se convirtieron en la ontología AO Cat como archivos

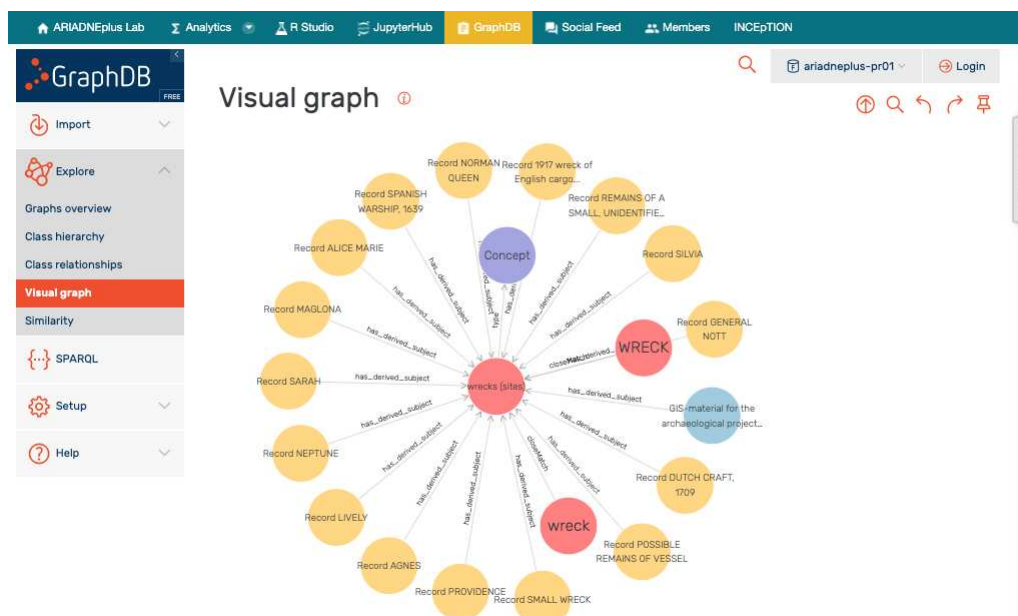
RDF y se almacenaron en la base de conocimientos ARIADNEplus (basada en la tecnología GraphDB). Los datos suministrados podrían incluir mapeos al Getty AAT o PeriodO, o estos podrían agregarse a la base de conocimientos en esta etapa de agregación de datos (Figura 10). Desde aquí, los recursos de datos de ARIADNE se pueden reutilizar para múltiples propósitos, por ejemplo, en portales o en VRE, o por consumidores externos, evitando la creación de un silo de datos. La base de conocimientos también facilita la vinculación con fuentes externas de Datos Abiertos Vinculados, como Wikidata (Figura 11). Los datos se publicaron primero en un portal de ensayo, para permitir una verificación de calidad, antes de su liberación en el portal público.

Uso del Portal ARIADNE para investigación arqueológica y gestión del patrimonio

Parte de la fortaleza del portal ARIADNE se encuentra en índices que se localizan en ese triple almacenamiento y ofrecen un medio rápido y efectivo de filtrar la Base de Conocimientos. Inicialmente, se implementaron en Elastic Search, pero se migraron a Open Search para mantener el énfasis en el software de código abierto; ofrecen una potente instalación de navegación por partes. Reflejando su origen y trayectoria en Dublin Core, el AO-Cat permite la definición de varios agentes asociados con un recurso de datos: Creador, Colaborador, Propietario, Editor, Responsable Científico. Estos pueden ser individuos u organizaciones. En muchos casos, la mayoría de los agentes declarados por los proveedores de datos eran la misma entidad, pero la variedad de agentes posibles permitía la atribución de una amplia gama de roles. Dado

Figura 11: Gráfico visual que muestra parte del triple almacenamiento ARIADNE en GraphDB.

Figure 11: Visual graph showing part of the ARIADNE triple store in GraphDB.



que Editor y Colaborador tenían filtros en la interfaz del portal, estos eran los más importantes, y uno de los desafíos del proceso de agregación fue asegurar que las categorías se aplicaran de manera consistente. Así, AO-Cat definía al Editor como el agente (generalmente el proveedor de datos o socio) responsable de publicar los metadatos en línea, mientras que el Colaborador era generalmente el mismo que el Propietario o Creador (Figura 12).

De igual importancia es el filtro “Tipo de Recurso”. Esto es efectivamente una especialización de la categoría Qué o Sujeto, extraída de una lista corta de 13 categorías

Figura 12: El filtro “Publicador” en el portal ARIADNE. Esto proporciona una identificación rápida del número de registros de cada proveedor y, en muchos casos, también actúa como una búsqueda de países.

Figure 12: *The Publisher filter in the ARIADNE portal. This provides a rapid identification of the number of records from each provider, and in many cases also acts as a country search.*

Name	Hits
Archaeology Data Service	1080189
British Museum	945228
Historic Environment Scotland	334636
Museum of Cultural History, University of Oslo	295674
Archaeological Information System of the Czech Republic (AIS CR)	229999
National Monuments Service, Department of Housing, Local Government and Heritage, Republic of Ireland	161383
Nara National Research Institute for Cultural Properties	141159
Data Archiving and Networked Services (DANS)	99291
Hungarian National Museum (HNM)	63019
CEIPAC - Universitat de Barcelona	57089
Historic England	50680
Institut national de recherches archéologiques préventives (Inrap)	37616
Swedish Rock Art Research Archives	25223
National Heritage Institute - Romania	23888
THANADOS	23580
Israel Antiquities Authority	23240
Aarhus University	16826
International Association for Classical Archaeology	14279
The Institute of Archaeology, Iceland	10344

Get 20 more results..

predefinidas. Aunque los términos pueden asignarse al AAT más amplio, proporcionan un conjunto pragmático de tipos de recursos que se definieron “de abajo hacia arriba” según los tipos más comunes de datos proporcionados por los socios. Nuevamente, la coherencia en la aplicación fue importante, pero el filtro proporciona un medio efectivo para definir el nivel de detalle requerido en una búsqueda (Figura 13). Los registros dentro de un “Tipo de Recurso” tienden a ser más homogéneos que entre ellos, asegurando un mayor grado de interoperabilidad dentro de las categorías. Si más proveedores se unieran a ARIADNE RI, las categorías son capaces de extenderse, aunque esto podría tener implicaciones para la necesidad de volver a categorizar colecciones existentes. También dependen del nivel de información proporcionado. Por lo tanto, fue relativamente fácil distinguir entre las 470,677 monedas y los 474,552 artefactos proporcionados por el Portable Antiquities Scheme (PAS) del Museo Británico, y algunos proveedores de datos, como el RGK, solo proporcionaron información sobre monedas, pero para la base de datos DIME de Dinamarca, los registros de monedas estaban integrados en el conjunto de datos más amplio.

La colaboración con THANADOS, una base de datos antropológica y arqueológica austriaca de cementerios y tumbas, proporciona un ejemplo especialmente bueno del valor de los tipos de recursos. El modelo de datos de THANADOS proporciona información en cuatro niveles: cementerio, tumba, entierro y artefacto, y esto se refleja en el perfil de aplicación de ARIADNE para datos mortuorios (Aspöck et al., 2023). El mapeo de AO-Cat permite el uso de los tipos de recursos de ARIADNE ‘Sitio/monumento’, ‘Entierro’ y ‘Artefacto’ para reflejar la estructura relacional de la base de datos de THANADOS, lo que permite a los usuarios filtrar según el nivel de detalle que les interese. A su vez, el perfil de aplicación mortuoria se puede utilizar para codificar los detalles de cada uno de los aspectos únicos de esta actividad arqueológica.

Para búsquedas más detalladas basadas en el tema, el uso del AAT como vocabulario central ha sido muy efectivo, y la Herramienta de Mapeo de Vocabulario ha permitido a los socios realizar el mapeo de manera relativamente fácil. Aunque el tesoro de Getty no define términos equivalentes en todos los idiomas, permite un nivel de búsqueda multilingüe. Habiendo especificado una búsqueda de términos de tema AAT, por ejemplo, los usuarios pueden ingresar ‘Swords’ (inglés), ‘Zwaarden’ (neerlandés), ‘Épée’ (francés), ‘Spada’ (italiano) o ‘Espadas’ (español) y obtener los mismos resultados (Figura 14). Además, siempre que el proveedor haya mapeado su término en su idioma nativo para el equivalente al concepto ‘Swords’ en el AAT de Getty, también se recuperarán sus registros. La búsqueda también utiliza su diccionario para proporcionar una función de autocompletar como una ayuda adicional

para el investigador.

La potencia del jerárquico AAT de Getty también se demuestra por la capacidad del usuario para encontrar objetos definidos a través de términos más estrechos, incluso si el término más amplio no se ha incluido en los metadatos (Figura 15). Por lo tanto, una búsqueda de ‘Swords’ también recupera ‘Cutlasses’, ‘Foil’, ‘Rapier’

▼ Resource type	
Enter text to filter on Resource types.	Name Hits ▼
Site/monument	1389759
Artefact	855253
Coin	471403
Fieldwork	421697
Fieldwork report	293825
Fieldwork archive	113941
Maritime	86509
Inscription	83365
Rock art	26805
Burial	11573
Dating	9015
Scientific analysis	7766
Building survey	1407

Figura 13: El filtro de “Tipo de Recurso” en el portal ARIADNE permite a los usuarios extraer un subconjunto de datos interoperables.

Figure 13: The Resource Type Filter in the ARIADNE portal allows users to extract a subset of interoperable data.

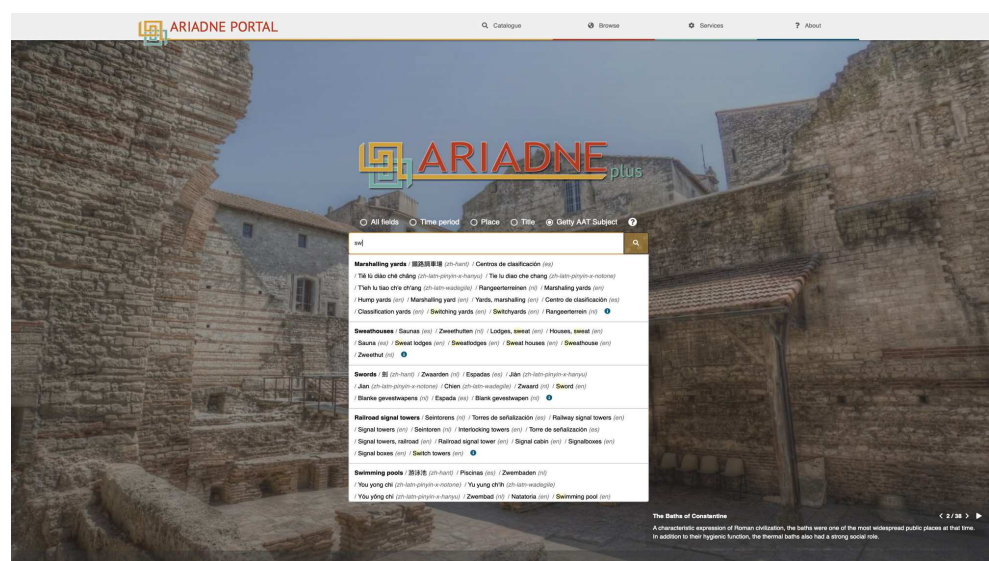
y cualquiera de las 22 categorías más estrechas. En algunos casos, el AAT de Getty no llegaba al nivel de detalle que los investigadores podrían desear, por ejemplo, en tipos detallados de broches o monedas, aunque donde hay respaldo y autoridad de dominio, se podría hacer un caso futuro para extender el tesoro para uso especializado.

Una de las características más populares y utilizadas de la interfaz del portal ha resultado ser la búsqueda basada en mapas (Figura 16). El mapa de calor proporciona una visualización rápida de la densidad de registros, pero se resuelve en puntos y polígonos individuales que se pueden hacer clic a medida que el investigador utiliza la función de zoom. Las herramientas adicionales incluyen la opción de utilizar diferentes bases de mapas y buscar a través de un cuadro delimitador o un polígono definido por el usuario. El menú proporciona los mismos filtros disponibles en la página principal de resultados, lo que permite a los usuarios seleccionar registros específicos de interés.

El soporte para polígonos se incorporó en el AO-Cat, ya que estos pueden ser importantes para mostrar la extensión completa de sitios arqueológicos o áreas de trabajo de campo. Los polígonos se ven mejor en los mapas incluidos en las páginas de inicio de registros individuales (Figura 17), que también ofrecen la opción de ver recursos cercanos. Los metadatos mostrados en las páginas de inicio de recursos individuales generalmente también proporcionan enlaces a la página de inicio del recurso principal. Esto es de considerable importancia, ya que los enlaces pueden dar acceso a información mucho más rica y detallada, incluidos informes de texto descargables y, en algunos casos, archivos digitales completos en línea. Idealmente, el enlace debería ser a través de identificadores permanentes, como Identificadores de Objetos Digitales (DOIs), contribuyendo a que el

Figura 14: La página de inicio del portal ARIADNEplus, que muestra las funciones multilingües y de autocompletar proporcionadas por la aplicación de Getty AAT.

Figure 14: The ARIADNEplus portal landing page, showing the autocomplete and multilingual features provided by the application of the Getty AAT.



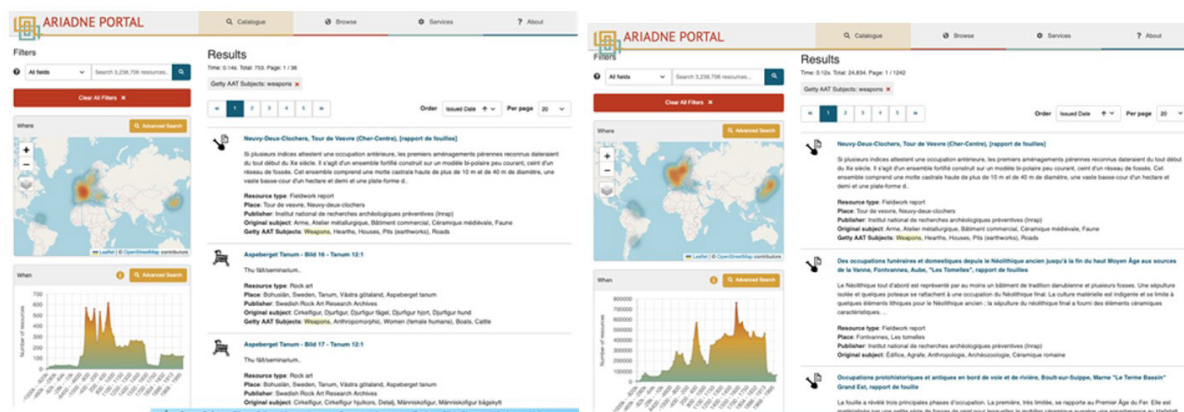


Figura 15: Una búsqueda de “armas” sin términos más específicos (izquierda) arroja solo 753 recursos, mientras que cuando se incluyen términos más específicos (derecha) se recuperan 24,834 recursos.

Figure 15: A searching on “weapons” without narrower terms (left) yields only 753 resources, whereas when narrower terms are included (right) 24,834 resources are retrieved.

portal proporcione acceso a datos justos y accesibles. Esto también proporciona una importante función tipo “vitrina” para los proveedores de datos, atrayendo a los investigadores a sus propios sitios web, de los cuales podrían no haber tenido conocimiento previamente. En algunos casos, sin embargo, los datos se han publicado en línea por primera vez a través del portal ARIADNE, y no hay información adicional para enlazar.

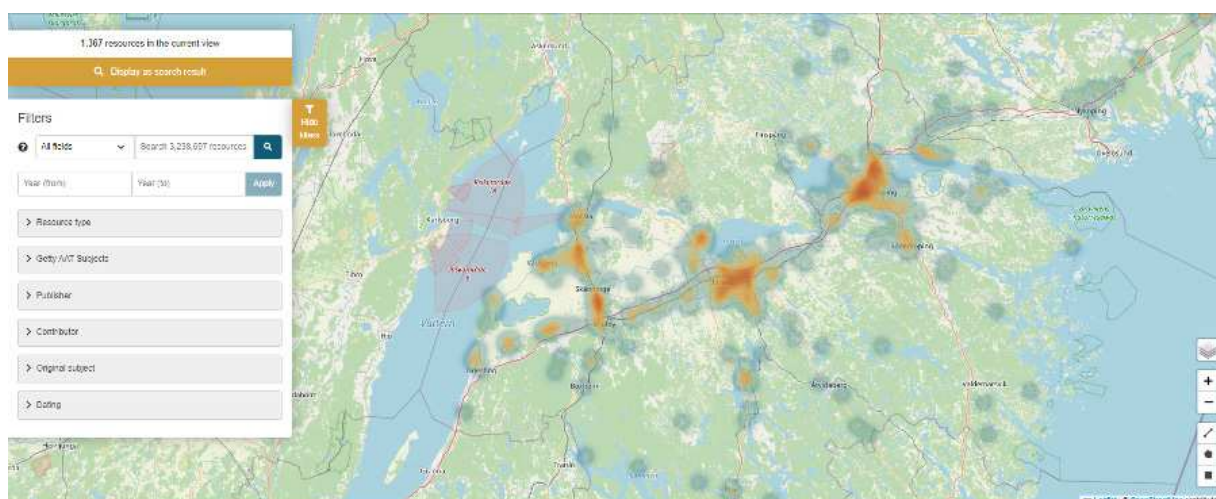
La función de exploración temporal proporciona una funcionalidad equivalente para la faceta “Cuándo”. Permite una visualización rápida del rango temporal de los recursos seleccionados, lo que proporciona una valiosa herramienta de investigación por derecho propio, pero el uso de PeriodO para definir rangos en años absolutos también permite al usuario seleccionar un lapso de tiempo específico de interés y ver los resultados. (Figura 18). Una segunda opción, agregada

durante el proyecto, permite filtrar por períodos de tiempo, según las regiones temporales seleccionadas. Por ejemplo, esto permite al usuario seleccionar sitios clasificados según la etiqueta cultural Edad del Hierro en diferentes regiones temporales, independientemente de las fechas Desde y Hasta de la Edad del Hierro en esas regiones.

La ontología AO-Cat demostró ser adecuada para las necesidades de la mayoría de los proveedores de datos durante la vida del proyecto, y solo se requirieron extensiones en dos áreas. La primera surgió del deseo de incluir imágenes entre los metadatos mostrados en las páginas de destino de los recursos, y miniaturas dentro de la lista de resultados. Esto surgió especialmente de los investigadores de artefactos que querían poder comparar objetos individuales (Figura 19). Claramente, esto también podría mejorarse

Figura 16: La función de exploración de mapas utiliza un mapa de calor para visualizar la densidad de registros.

Figure 16: The map browsing function uses a heat map to visualise the density of records.



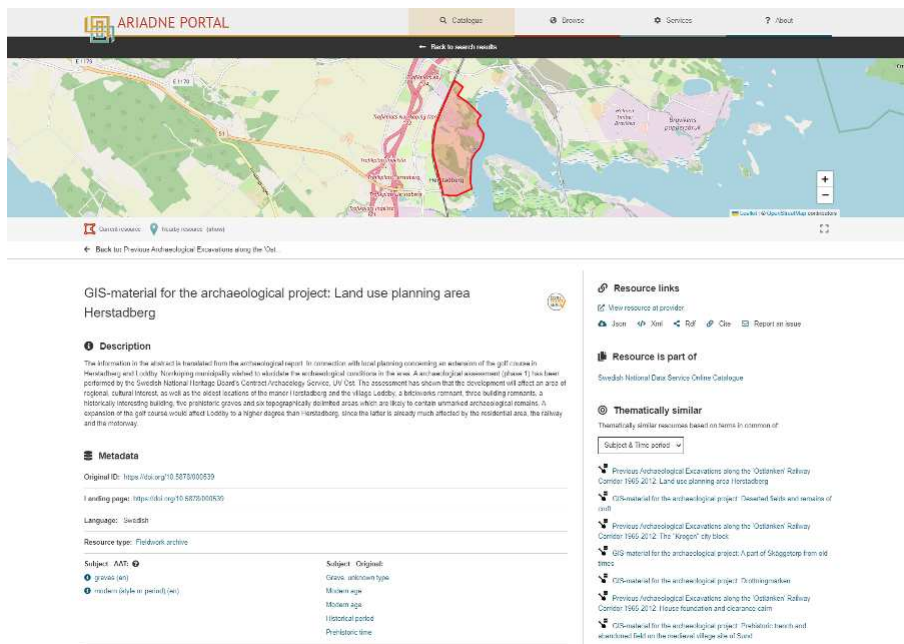


Figura 17: Página de inicio para un recurso de datos individual, que ilustra la capacidad de mostrar la extensión geoespacial del recurso como un polígono.

Figure 17: Landing page for an individual data resource, illustrating the ability to display the geospatial extent of the resource as a polygon.

utilizando el Marco Internacional de Interoperabilidad de Imágenes (IIIF) y visores apropiados en un proyecto futuro.

La segunda necesidad de una extensión al AO-Cat surgió porque algunos proveedores de datos pidieron incluir información sobre la región espacial cubierta por un recurso de datos, con una especificación sobre la precisión de la información. Esto se debió a dos razones principales. En primer lugar, en algunos casos, el proveedor de datos solo tenía datos de la ubicación del sitio a un nivel impreciso, por lo que era necesario dejar claro, al representar ubicaciones en un mapa, que el punto podría representar la ubicación del sitio hasta

un nivel cercano a 10 km, por ejemplo. Una vez que esto se codificó dentro del AO-Cat y los datos se almacenaron en el triple sistema, el portal lo manejó utilizando una simbología diferente. En segundo lugar, había casos en los que el proveedor de datos tenía datos espaciales sensibles que no deseaban hacer públicos, tal vez para proteger la ubicación precisa de un sitio. En estos casos, se suministraron datos espaciales a la resolución que estaban dispuestos a hacer disponibles, y la ubicación se representó como un cuadro delimitado, que incluía el punto de hallazgo real en algún punto aleatorio. Por ejemplo, los datos espaciales para la base de datos del PAS del Reino Unido se suministraron con un cuadro delimitador de 1 km o 10 km, según la sensibilidad del

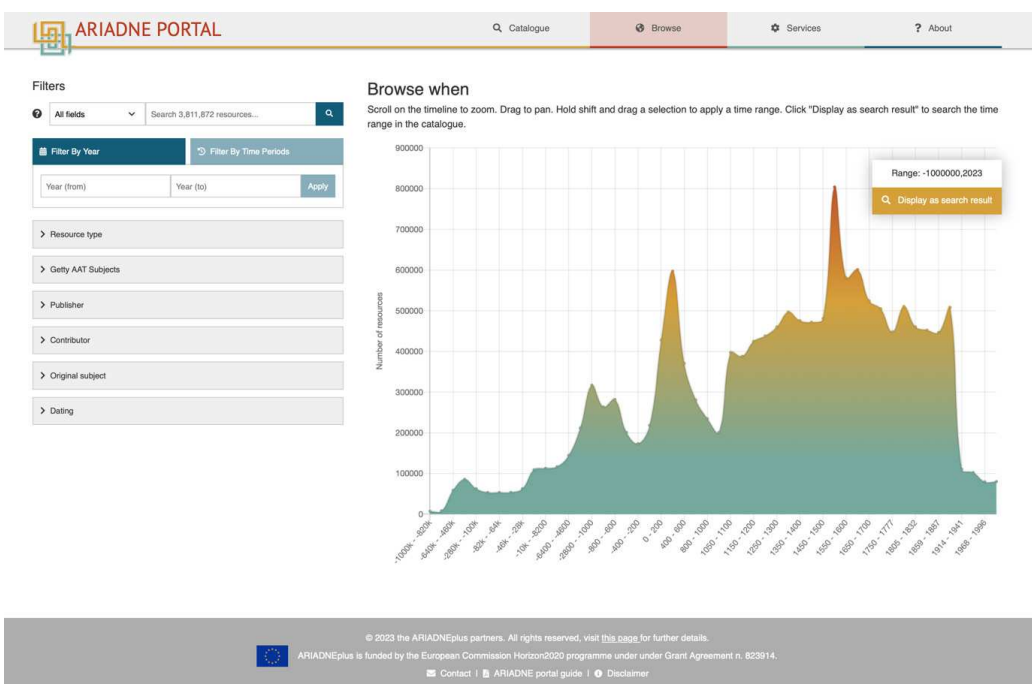


Figura 18: La función de navegación temporal en el portal.

Figure 18: *The temporal browse facility in the portal.*

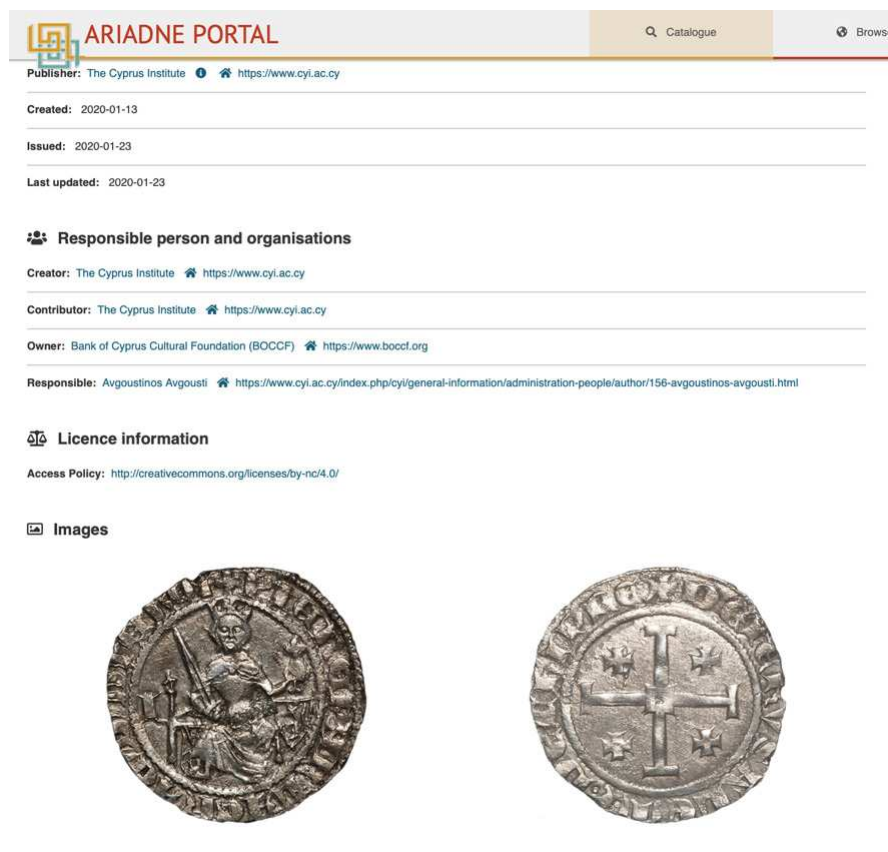


Figura 19: La página de inicio de una moneda de la Fundación Cultural del Banco de Chipre que muestra el valor de incluir imágenes entre los metadatos recopilados por el portal.

Figure 19: The landing page for a coin held by the Bank of Cyprus Cultural Foundation showing the value of including images amongst the metadata harvested by the portal.

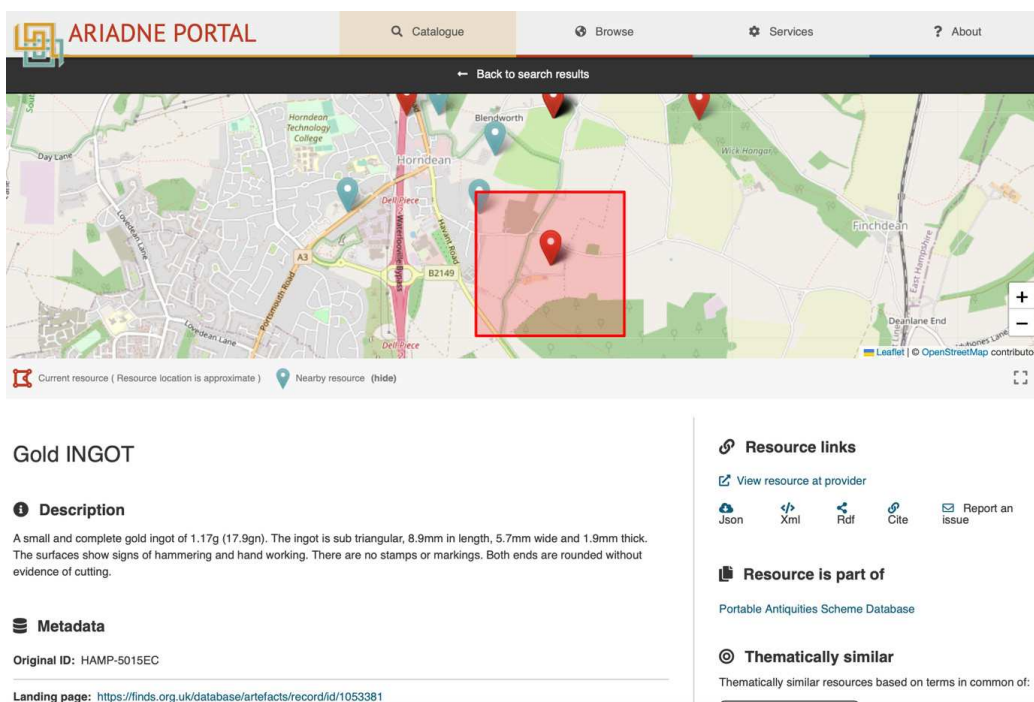
lugar de hallazgo (Figura 20).

Finalmente, la base de conocimientos ARIADNE y la arquitectura del portal permiten adaptarse fácilmente para proporcionar una amplia gama de interfaces más especializadas. Por ejemplo, como parte del proyecto "Towards a National Collection Unpath'd Waters",

el ADS pudo desarrollar un portal que ofrece una búsqueda cruzada distribuida de las muchas agencias que tienen datos sobre el patrimonio marítimo de las aguas costeras del Reino Unido mediante la implementación de una instancia del portal ARIADNE (Figura 21). En este caso, la búsqueda estaba "cableada" para limitarla a aquellos recursos para

Figura 20: La ubicación del lugar de búsqueda de un lingote de oro mostrada con una resolución de 1 km, de modo que el lugar de búsqueda real puede estar en cualquier lugar dentro del cuadro delimitador rojo.

Figure 20: The location of the find spot of a gold ingot displayed to a 1km resolution, such that the actual find spot may be anywhere within the red bounding box.



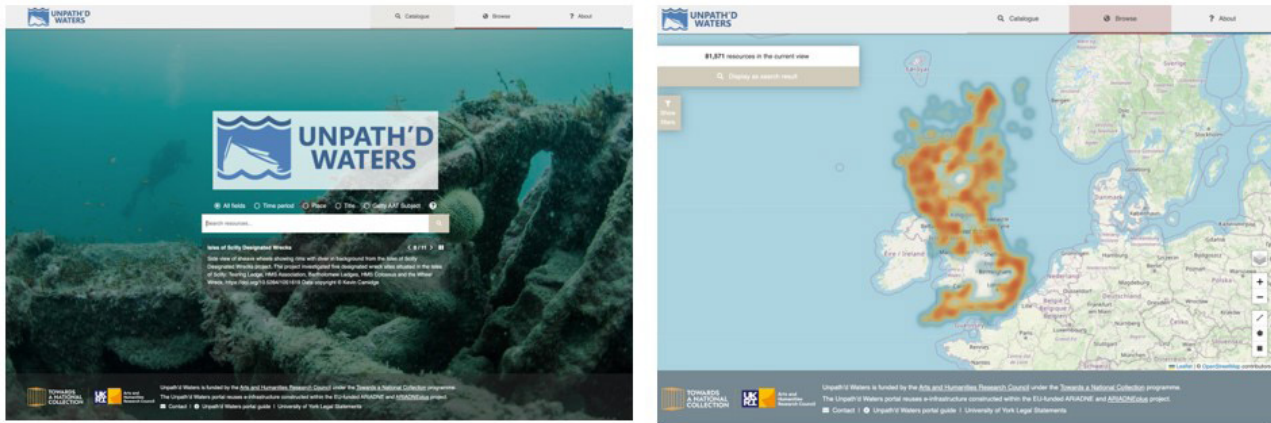


Figura 21: El portal *Unpath'd Waters* utiliza la base de conocimientos y la interfaz del portal ARIADNE para proporcionar una búsqueda integrada del patrimonio marítimo del Reino Unido.

Figure 21: The *Unpath'd Waters* Portal uses the ARIADNE Knowledge Base and Portal interface to provide a hard-wired UK maritime heritage search.

los cuales el Tipo de Recurso era 'Marítimo', y para que solo se recuperaran registros de proveedores de datos del Reino Unido. Dado que la fuente de metadatos es el almacenamiento triple de ARIADNE RI, los metadatos se almacenan solo una vez, y los recursos agregados para el proyecto *Unpath'd Waters* aparecen simultáneamente en el portal de ARIADNE, proporcionando así una vista europea del conjunto de datos del Reino Unido. El enfoque podría ampliarse en cualquier dirección, por ejemplo, proporcionando vistas nacionales para agencias que no desean (o no tienen la capacidad) de desarrollar su propia interfaz de búsqueda, evitando la duplicación, o para vistas temáticas de datos, a nivel internacional.

Conclusión

En conclusión, el portal ARIADNE proporciona acceso al conjunto de datos arqueológicos en línea más grande a nivel internacional, y no hay nada comparable en otras disciplinas de Artes y Humanidades. Sin embargo, es esencial tener en cuenta que el papel del servicio de agregación de ARIADNE no es duplicar los servicios ya disponibles en los sitios web de los proveedores de datos, sino proporcionar interoperabilidad y descubrimiento de recursos a través de múltiples fuentes. No obstante, en algunos casos, el portal proporciona una herramienta de investigación poderosa por sí misma (Richards, 2023). El servicio se construyó sobre las bases establecidas por varios proyectos anteriores, pero se han aprendido lecciones adicionales. En primer lugar, para desarrollar una búsqueda cruzada útil, es evidente que el uso consistente de una ontología es esencial. El CIDOC CRM era lo suficientemente amplio como para abarcar todos los subdominios involucrados, pero era importante realizar un mapeo consistente, y el desarrollo de perfiles de aplicación liderados por la comunidad fue

esencial en este sentido. En segundo lugar, para lograr la interoperabilidad, el control del vocabulario también es esencial, y tuvimos la suerte de poder basarnos en décadas de inversión en estándares de datos y tesauros en arqueología. El Getty AAT proporcionó una columna vertebral común potente, pero los mapeos requieren supervisión nuevamente para garantizar la consistencia en los mapeos de los socios. En tercer lugar, es esencial darse cuenta de que, al tratar con datos tan heterogéneos, es imposible hacer que todo sea interoperable, y es necesario considerar el papel de la agregación y cómo los investigadores utilizarán los datos. En ARIADNE, basándonos en proyectos anteriores, Que, Cuando, Donde resultó ser extremadamente efectivo, aunque a posteriori Quién también podría haber sido útil, especialmente si el portal se extendiera para incluir más informes y libros escritos por individuos. Sin embargo, actualmente carecemos de listas de autoridades que nos permitan definir a las personas. Una mayor adopción de identificadores ORCID podría ofrecer una solución para personas en la actualidad, pero necesitamos examinar las soluciones e identificadores ofrecidos por Wikidata para tratar con personas del pasado.

Por sobre todo, nuestro trabajo en la agregación de datos ha destacado la importancia de equipos interdisciplinarios en este esfuerzo y la necesidad de que los arqueólogos trabajen en estrecha colaboración con científicos de la información. Es necesario tomar decisiones clave sobre los niveles de detalle o granularidad y las escalas de la agregación de datos, y qué es útil agregar a nivel de ítem, o dónde tiene sentido agregar solo a nivel de colección. En este sentido, nuestro enfoque debe ser impulsado por la investigación y las preguntas de los investigadores y especialistas de este campo.

Agradecimientos

El flujo de trabajo de agregación aplicado en ARIADNE se construyó sobre una variedad de soluciones de software, pero su éxito dependió de las habilidades y el compromiso humanos. Durante el proyecto ARIADNEplus, el grupo de trabajo de agregación se reunió en casi 50 ocasiones para gestionar la canalización y me gustaría aprovechar esta oportunidad para agradecer a sus miembros: Alessia Bardi, Ceri Binding, Achille Felicetti, Carlo Meghini, Enrico Ottonello y Maria Theodoridou, así como a nuestros colegas en SND, en particular a Pablo Millett, por la entrega de la interfaz del portal. Finalmente, muchas de las ilustraciones utilizadas en este documento se tomaron de varios productos del proyecto ARIADNE y se reproducen cortesía de los socios del proyecto.

La investigación que llevó a estos resultados recibió financiamiento del Séptimo Programa Marco de la Unión Europea bajo el acuerdo de subvención No 313193, y del programa de investigación e innovación Horizonte 2020 de la Unión Europea bajo el acuerdo de subvención No 823914.

Bibliografía

- Aloia, N., Binding, C., Cuy, S., Doerr, M., Fanini, B., Felicetti, A., Fihn, J., Gavrillis, D., Geser, G., Hollander, H., Meghini, C., Niccolucci, F., Nurra, F., Papatheodorou, C., Richards, J., Ronzino, P., Scopigno, R., Theodoridou, M., Tudhope, D., Vlachidis, A. and Wright, H. (2017) Enabling European Archaeological Research: The ARIADNE E-Infrastructure, *Internet Archaeology* 43. <https://doi.org/10.11141/ia.43.11>
- Aspöck, E., Eichert, S., Theodoridou, M., Richards, N. and Felicetti, A. (2023) *Integrating data on early medieval graves: Mapping the THANADOS database to the ARIADNE infrastructure with the ARIADNE Mortuary Data Application Profile*, *Internet Archaeology* (this volume).
- Austin, A., Richards, J.D., Pinto, F. and Ryan, N. (2002) Joined up writing: an Internet portal for research into the historic environment, in G. Burenhult (ed) *Archaeological Informatics: Pushing the Envelope CAA2001*, BAR International Series 1016 (pp. 243-51).
- Benardou, A., Champion, E., Dallas, C. and Hughes, L. (2017) Introduction: A Critique of Digital Practices and Research Infrastructures, In A. Benardou, E. Champion, C. Dallas, and L. Hughes (eds) *Cultural Heritage Infrastructures in Digital Humanities*. (pp. 1-14) London: Routledge.
- Bibby, D. (2021) 'Digital Archaeological Archiving in Baden-Württemberg, Germany: an evolving system', *Internet Archaeology* 58. <https://doi.org/10.11141/ia.58.3>
- Binding, C., Charno, M., Jeffrey, S., May, K. and Tudhope, D. (2015) Template based semantic integration: from legacy archaeological datasets to linked data, *International Journal on Semantic Web and Information Systems* 11(1), 1-29. <https://doi.org/10.4018/IJSWIS.2015010101>
- Binding, C. and Tudhope, D. (2023) Automatic Normalization of Temporal Expressions, *Journal of Computer Applications in Archaeology* 6, 24-39. <https://doi.org/10.5334/jcaa.105>
- Dam, C. and Hansen, H.J. (2005) The European Digital Resource in Archaeology: Sites and Monuments Data as a common European Web Resource, *Internet Archaeology* 18. <https://doi.org/10.11141/ia.18.4>
- Felicetti, A., Gerth, P., Meghini, C. and Theodoridou, M. (2015) *Integrating Heterogeneous Coin Datasets in the Context of Archaeological Research*, *EMF-CRM@TPDL*. <https://ceur-ws.org/Vol-1656/paper2.pdf>
- Felicetti, A., Meghini, C., Richards, J.D. and Theodoridou, M. (2023) *The AO-Cat Ontology*. <https://doi.org/10.5281/zenodo.7818375>
- Geser, G., Richards, J.D., Massara, F. and Wright, H. (2022) Data Management Policies and Practices of Digital Archaeological Repositories', *Internet Archaeology* 59. <https://doi.org/10.11141/ia.59.2>
- Hansen, H.J. (1992) European archaeological databases: problems and prospects, in J. Andresen, T. Madsen and I. Scollar (eds) *Computing the Past. Computer Applications and Quantitative Methods in Archaeology*. (pp. 229-37) Aarhus: Aarhus University Press.
- Hansen, H.J. and Fernie, K. (2010) *CARARE: Connecting Archaeology and Architecture in Europeana*, International Euro-Mediterranean Conference. <https://doi.org/10.1007/978-3-642-16873->

4_36

- Huggett, J. (2015) Digital Haystacks: Open Data and the Transformation of Archaeological Knowledge', in A.T. Wilson and B. Edwards (eds) *Open Source Archaeology: Ethics and Practice*. (pp. 6-29) Warsaw, Poland: De Gruyter Open. <https://doi.org/10.1515/9783110440171-003>.
- Isaksen, L., Simon, R., Barker, E. and de Soto, C. (2014) *Pelagios and the emerging graph of ancient world data*, *Proc. ACM Conference on Web Science (WebSci'14)* (pp.197-201). ACM. <https://doi.org/10.1145/2615569.2615693>
- Jeffrey, S., Xia, L., Richards, J., Bateman, J., Kintigh, K., Pierce-McManamon, F. and Brin, A. (2012) The Transatlantic Archaeology Gateway: Bridging the Digital Ocean, in M. Zhou, I. Romanowska, Z. Wu, P. Xu, and P. Verhagen (eds) *CAA2011-Revive the Past: Proceedings of the 39th Conference in Computer Applications and Quantitative Methods in Archaeology*, (pp. 198-208) Beijing, China, 12-16 April 2011. Pallas Publications.
- Kansa, E. (2014) Open Context and Linked Data, in T. Elliott, S. Heath and J. Muccigrosso (eds) *Current Practice in Linked Open Data for the Ancient World*, ISAW Papers 7(10) <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/kansa/>
- Kenny, J. and Richards, J.D. (2005) Pathways to a Shared European Information Infrastructure for Cultural Heritage, *Internet Archaeology* 18 <https://doi.org/10.11141/ia.18.6>
- Kilbride, W.G. (2004) 'The Danube in prehistory in the digital age: Towards a common information environment for European archaeology', *Archeologia e Calcolatori* 15, 129-44.
- May, K., Binding, C. and Tudhope, D. (2015) Barriers and Opportunities for Linked Open Data use in Archaeology and Cultural Heritage, *Archäologische Informationen* 38, 173-84.
- Meghini, C., Scopigno, R., Richards, J.D., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F., Felicetti, A., Ronzino, P., Nurra, F., Papatheodorou, Ch., Gavrili, D., Theodoridou, M., Doerr, M., Tudhope, D., Binding, C. and Vlachidis, A. (2017) ARIADNE: A Research Infrastructure for Archaeology, *Journal of Computing and Cultural Heritage* 10.3, 1-27. <https://doi.org/10.1145/3064527>
- Michard A, Christophides V, Scholl M, Stapleton M, Sutcliffe D, Vercoustre A-M. (1998) The Aquarelle resource discovery system, *Computer Networks and ISDN Systems* 30(13), 1185-1200.
- Niccolucci, F. and Richards, J.D. (2019) ARIADNE and ARIADNEplus in J.D. Richards and F. Niccolucci (eds) *The ARIADNE Impact*. (pp. 7-25) Budapest: Archaeolingua.
- Nicholson, C., Fernandez, R. and Irwin, J. (2021) Digital Archaeological Data in the Wild West: the challenge of practising responsible digital data archiving and access in the United States, *Internet Archaeology* 58 <https://doi.org/10.11141/ia.58.22>
- Richards, J.D. (2017) Twenty Years Preserving Data: A View From the UK, *Advances in Archaeological Practice* 5.3, 227-37. <https://doi.org/10.1017/aap.2017.11>
- Richards, J.D. (2023) Getting it Together: Combining information about archaeological sites and artefacts in ARIADNE, *Internet Archaeology* 64. <https://doi.org/10.11141/ia.64.14>
- Richards, J.D. and Niccolucci, F. (eds) (2019) *The ARIADNE Impact*. Budapest: Archaeolingua. <https://doi.org/10.5281/zenodo.4319058>
- Richards, J.D., Jakobsson, U., Novák, D., Štular, B. and Wright, H. (2021) Digital Archiving in Archaeology: The State of the Art. Introduction, *Internet Archaeology* 58. <https://doi.org/10.11141/ia.58.23>
- Shaw, R., Rabinowitz, A., Golden P. and Kansa, E. (2016) A Sharing-oriented Design Strategy for Networked Knowledge Organization Systems, *International Journal on Digital Libraries* 17, 49-61. <https://doi.org/10.1007/s00799-015-0164-0>
- Vassallo, V., Theodoridou, M., Felicetti, A. and Avgousti, A. (2023) Tools and Ontologies for the Aggregation and Management of Cypriot Archaeological Datasets, *Internet Archaeology* 64. <https://doi.org/10.11141/ia.64.14>

[org/10.11141/ia.64.10](http://doi.org/10.11141/ia.64.10)

Wright, H. and Richards, J.D. (2018) Reflections on
Collaborative Archaeology and Large-Scale

Online Research Infrastructures, *Journal of Field
Archaeology* 43: S60-S67, <https://doi.org/10.1080/00934690.2018.1511960>