Article:
Richards, Julian D, Brunello, Marco, Carrillo-Garza, Valeria et al. (2026) Dealing with Unprecedented Scale and Complexity: Lessons from Archiving HS2 Digital Archaeological Data. International Journal of Digital Curation. 1138. ISSN: 1746-8256

https://doi.org/10.2218/ijdc.v20i1.1138

IJDC / *Conference Paper*

Dealing with Unprecedented Scale and Complexity: Lessons from Archiving HS2 Digital Archaeological Data

Marco Brunello
Archaeology Data Service

Valeria Carrillo-Garza
Archaeology Data Service

Olivia Foster
Archaeology Data Service

Katie Green
Archaeology Data Service

Julian D. Richards
Archaeology Data Service

Abstract

As the UK's largest linear infrastructure project, High Speed 2 (HS2) necessitated the UK's most extensive archaeological programme. The project's monumental scale and complexity posed risks to data consistency, preservation, and reuse. The Archaeology Data Service (ADS) was mandated to manage this challenge, overseeing over 34 terabytes of heterogeneous data. This volume exposed the limitations of existing frameworks while providing a unique opportunity for innovation in archiving infrastructure and rationale.

This paper reflects on how the ADS was transformed by HS2, evolving its technical capacity alongside its understanding of the organisational and ethical dimensions of large-scale curation. These challenges were a proving ground for testing future-proofed data management approaches. The resulting tools, workflows, and perspectives have expanded the ADS's capacity to ensure data integrity in perpetuity, providing a long-term infrastructural benefit to the wider archaeological sector.

Submitted 18 March 2026 ~ *Accepted* 15 April 2026

Correspondence should be addressed to Marco Brunello, email: marco.brunello@york.ac.uk.

This paper was presented at the International Digital Curation Conference IDCC26, Zagreb, Croatia 16-18 February 2026.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/.

From Field to Archive: Standardising Distributed Data in National Infrastructure

As the largest linear infrastructure project in the UK, High Speed Two (HS2) was designed to modernise national transport via a new rail line between London and Birmingham. Aimed at providing faster transport across the South of England and the West Midlands, the project necessitated the largest archaeological programme ever seen in the country (Aryankhesal, 2023).

Works on the new network required archaeological interventions on numerous sites along the approximately 200 km line, which resulted in a large amount of archaeological evidence to be unearthed, analysed, and recorded. The monumental scale of this project, defined by its geographic spread and extended timeline, marks a significant shift in the practice of rescue archaeology. The project has influenced how heritage is investigated and managed across the UK, creating new challenges for the recording and preservation of archaeological data. Central to this effort is ensuring that these findings remain publicly accessible to support future research and public understanding (HS2 Ltd., 2017).

As a government-mandated initiative, HS2's archaeological framework was defined by specific, top-down requirements set by HS2 Ltd. This company, specifically conceived to develop and promote the new rail network, endeavoured to ensure consistency across the vast array of contractors involved (*High Speed Rail (Preparation) Act 2013*). However, implementing this centralised mandate presented significant logistical complexities. The scale of the project necessitated the involvement of numerous archaeological organisations, both concurrently and sequentially, each utilising established internal methodologies. Navigating this inherent variability was essential to ensuring the long-term coherence of digital outputs and safeguarding the future accessibility and reuse of the archaeological record.

These concerns are not new; the discipline has long grappled with the tension between raw data production and its interpretative synthesis. Decades ago, Hills (1993) famously observed that the discipline was "quietly drowning" in its own data, even as practitioners lamented the nonappearance of the work of others. This critique highlighted a fundamental crisis in dissemination. The volume of excavated material was growing, but coherent, synthesised outputs remained elusive. The transition to the digital age offered a potential lifeline; it also introduced a shift from "order to chaos," requiring new modes of dissemination to prevent data from becoming siloed or inaccessible (Gaffney & Exon, 1999). Richards (2004), over 20 years ago at the time of writing, observed that the mere existence of online archives does not guarantee utility; true value lies in the ability to move beyond simple storage towards integrated, reusable digital resources. There have been calls for more dynamic archives. For example, Opgenhaffen (2022) argues that "a digital archaeological project archive is not a passive repository of data, but a dynamic and participatory environment of data and knowledge exchange and learning."

To mitigate these risks within the HS2 project, the Archaeology Data Service (ADS) was tasked with setting standards for data deposition and providing long-term digital preservation and access to the archaeological data created by the HS2 Historic Environment Programme. To bridge the gap between rigid government requirements and diverse contractor outputs, the ADS had to move beyond standard deposition and archiving procedures, adapting to accommodate the unprecedented scale and complexity of the data received.

Therefore, this paper aims to analyse and reflect on the challenges faced by the ADS team in their pioneering effort when dealing with archaeological data at an unparalleled scale, within the constraints due to the complexities and multi-stranded nature of a large-scale government project. The broad challenges ranged from coordinating communication with diverse contractor teams to enforcing rigorous metadata standards; this paper focuses on the specific innovations compelled by the project's scale and complexity. In particular, how the automation of preservation workflows, the development of bespoke Graphical User Interface (GUI)-based tools, and the implementation of a scalable reuse framework allowed a coherent approach to be defined that accommodates diverse contractor outputs is examined. By reflecting on these focused interventions, how the pressures of a "mega-project" served as a catalyst for broader infrastructural

advancement within the organisation is demonstrated. As tensions between data volume and archival consistency are widely shared, focused interventions in these areas have the potential to benefit both the archaeology sector and the wider digital curation community.

Managing the Mega-Project: Technical Innovation and Curatorial Care

Institutional Capacity and Policy Shifts

The HS2 Historic Environment Programme presented an unprecedented challenge to the digital curation landscape in the UK, simultaneously providing a significant opportunity for innovation in archiving practice, infrastructure, and institutional rationale. Some obstacles were anticipated, as discussed in the following section (Navigating Expected and Unexpected Challenges), and the magnitude of the project required a fundamental shift in ADS operations. Eventually consisting of over 360 distinct datasets and 30+ terabytes, HS2 data rapidly increased the ADS's holdings over 5 years and now represents a third of our total preservation footprint (Figure 1). Receiving this high volume of data in a short timeframe highlighted the limitations of existing preservation and archiving infrastructure. Initially, ADS tools and systems were not designed to accommodate this volume and complexity; however, these bottlenecks proved to be fertile ground for developing structural improvements.

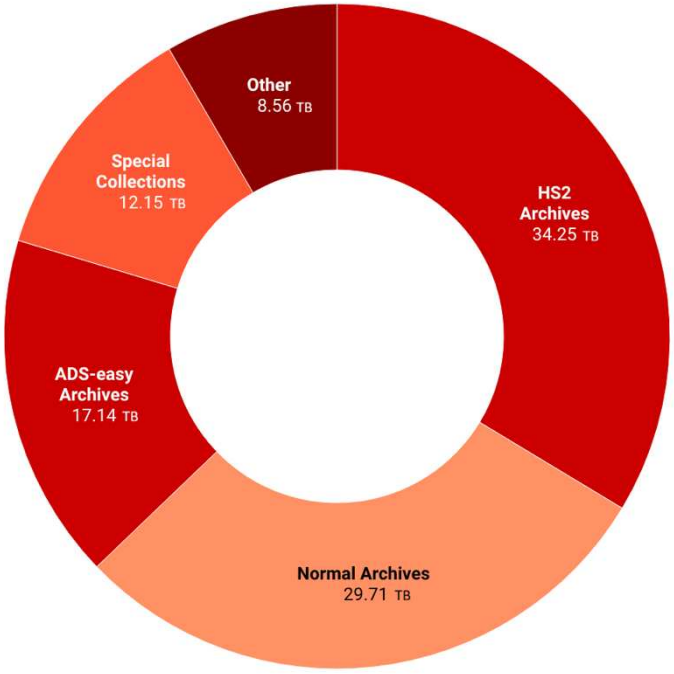

Figure 1. The Archaeology Data Service's (ADS) total preservation footprint by collection type as of March 2026, including High Speed Two (HS2) archives, datasets deposited using the ADS-easy deposit system, "Special Collection" archives and so-called "Normal archives," which are typically projects that exceed the deposit limit of ADS-easy. TB = Terabytes.

The HS2 Historic Environment Digital Archive required the appraisal, accession, and dissemination of data in addition to the pre-existing workload of ADS staff. The sheer volume of HS2 data necessitated a significant expansion of institutional capacity, with new appointments

dedicated primarily to the HS2 project. Similarly, pre-existing policies and procedures were expanded to ensure the workflow remained efficient when handling specific needs and complex archives. Existing ADS guidance provided a foundation for data structuring; the unparalleled throughput necessitated a transition from bespoke manual processing towards a more automated, systematic approach.

Navigating Expected and Unexpected Challenges

The first step in managing this wealth of data was directing archaeological units towards established depositor guidance (ADS, 2022a). These resources were vital, providing a standardised framework for dataset structuring, file naming, and metadata compilation. Establishing these expectations early on ensured that documentation, including archive decision records and HS2-approved reports, remained consistent across deposits, offering a "common language" to a heterogeneous landscape of contractors.

However, a notable "unexpected" challenge was the disparity in scale between individual sites. Initially, the assumption that all HS2 archives would be large and high-complexity led to a decision not to use ADS-easy, the automated portal designed for small-to-medium deposits (limited to 100 MB per file and 1,000 files per archive). In retrospect, many HS2 archives met these criteria and using ADS-easy for smaller datasets could have accelerated archiving by reducing some manual pre-accessioning checks. However, using a multi-strand approach to data deposition may have introduced complexities that the unified bespoke workflow sought to avoid.

Ultimately, the HS2 Historic Environment Digital Archive called for a dual transformation: capacity could not simply be expanded via human resources alone, nor could automation be relied on to fix inconsistent data. Instead, the project required an integrated model where technical innovation and proactive curation functioned in tandem.

Results: Technical Evolution and Workflow Automation

Reducing Technical Barriers: The OMS Viewer

To manage this massive throughput, the ADS transitioned from a code-heavy, manual infrastructure to an accessible, tool-based environment. The primary objective was to reduce technical barriers or bottlenecks by replacing manual Structured Query Language (SQL) tasks with robust GUI operations. By developing bespoke interfaces, augmenting existing tools, and creating batch processing utilities, the ADS enabled a more diverse team to manage archival tasks that previously required specialised database expertise.

A central requirement for all digital objects archived by the ADS is the inclusion of core object metadata (ADS, 2022b) and, where necessary, additional technical metadata. These records are housed within the Object Management System (OMS) database. A pivotal development in managing this data was the "OMS Viewer," a web-based interface developed by Applications Developer Jo Gilham in 2023. This tool allows staff to access, insert, and update object-level records without using SQL.

The interface streamlines the ingestion process by allowing ADS archivists to import and create new digital objects via comma-separated values (CSV) files containing file names and core metadata (Figure 2). On upload, the archivist is prompted to map spreadsheet headings to the corresponding database fields, a workflow that generates new object representations (which may consist of single or multiple files) and assigns unique identifiers to each. Beyond initial ingestion, the tool facilitates bulk edits; archivists can export existing metadata into a CSV file, modify the data, and reimport it. This controlled environment increases efficiency for large-scale changes and mitigates the risks of human error inherent in running manual SQL statements against a live database.

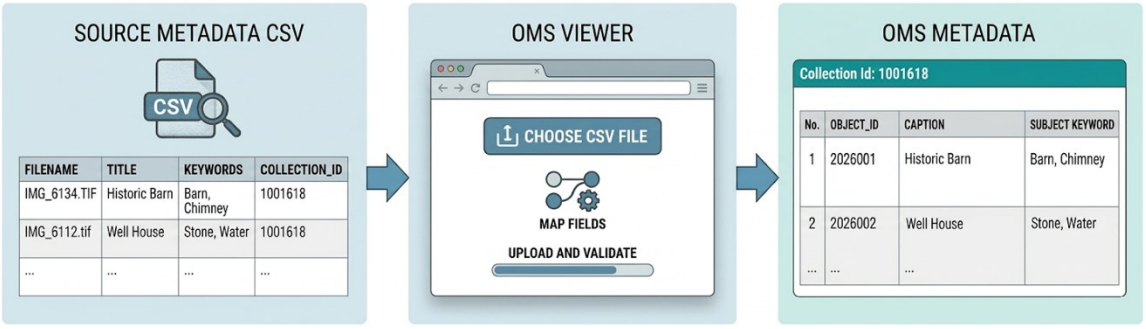

Figure 2. Object metadata import workflow when using the Object Management System (OMS) Viewer.

This custom interface has transformed the internal workflow. It has significantly streamlined onboarding, as new team members can interact with ADS databases without requiring extensive knowledge of the underlying database structure. Furthermore, the tool incorporates built-in controls to safeguard against data loss, ensuring high integrity. Beyond internal management, the OMS Viewer facilitated essential data cleaning post-archiving, improving consistency and interoperability across HS2 archives. This standardisation proved crucial to the functional success of HS2 search interfaces; for instance, the normalisation of subject keywords directly enables more accurate and comprehensive search results for end users.

Augmenting Infrastructure and Batch Processing

In addition to the OMS Viewer, several existing infrastructure tools were augmented by Software Engineer Paul Young to accommodate significantly larger datasets. These enhancements targeted the "Match Objects" tool, used to pair the three representations that would eventually be provided for each object (Original, Preservation, and Dissemination copies) and the "Add Objects to Groups" tool, which facilitates the creation of object groups for dissemination, typically mirroring an archive's directory structure.

To alleviate pressure on the Collection Management System (CMS), the new "Match Objects (Large Archives)" tool introduces the ability to load selected directories individually, rather than attempting to display every file in a collection simultaneously. Similarly, the updated "Add Objects to Groups" interface represents a significant improvement over its predecessor. The original iteration allowed archivists to group objects by data type or directory, but was limited to processing a single group at a time. In addition, it merely generated SQL statements that the user had to manually execute and commit via third-party database tools. The improved version streamlines this into a dual-pane interface, displaying file store directories on one side and interface groups on the other. This allows archivists to map directories to groups and process multiple batches simultaneously, removing the need for manual SQL intervention.

Despite the success of these web-based enhancements, a subset of datasets still exceeded the performance thresholds of browser-based interfaces. For exceptionally large collections containing over 100,000 digital objects, such as the archaeological excavation data from Fleet Marston in Buckinghamshire (COPA: Cotswold Oxford Pre-Construct Archaeology & HS2 Ltd., 2026), the high volume of elements to be rendered can cause standard interfaces to become unresponsive. In these specific instances, the ADS employs command-line programmes to handle intricate structures where metadata mapping is too complex for standard GUI tools.

These high-capacity programmes are accessed via PuTTY, a Secure Shell (SSH) and Telnet client, enabling archivists to run custom scripts developed by Paul Young. These scripts automate object matching by filestem or directory, facilitate bulk file renaming, and manage documentation of complex relationships between objects. These command-line processes remain an essential fallback for large archives.

Furthermore, programmatic approaches were implemented to address manual, repetitive tasks, significantly accelerating the creation of parent-child relationships between objects and

metadata and web-ready preview images. A key example is a batch image conversion tool, developed by Applications Developer Adam Fox in 2023, which processes image files in accordance with ADS-specific repository operations. Built on existing ADS guidelines for file format normalisation, this tool automates established standards without altering fundamental archival procedures. Using XnView (v7.110) and ExifTool (v12.65), the tool automates the creation of preservation and dissemination versions of files (e.g., converting JPG to TIF and vice versa), generates preservation versions of photogrammetry images, and produces optimised previews for the web interface.

While archivists remain cautious, ensuring that significant properties are retained by performing post-processing spot checks, this tool has empowered the team to process deposits with greater speed and fewer manual errors. By retaining original directory structures and allowing for the simultaneous processing of multiple collections, it has transformed a previously labour-intensive workflow. These solutions, now applied to large and small datasets, have strengthened long-term institutional resilience and set a new standard for managing high-volume archaeological data without compromising standards.

The Collaborative Core: In-House Development and Domain Expertise

A fundamental driver of these technical advances was the unique internal synergy between the ADS collections team, archivists, and developers. Unlike many organisations that rely on externalised IT services, which often result in a disconnect between technical execution and domain-specific needs, the ADS benefits from an integrated, in-house development team. This proximity is vital; the developers either possess a background in archaeology or have developed a comprehensive understanding of the discipline through years of close collaboration. This shared domain literacy ensures that challenges are met with solutions that are archaeologically informed rather than data-agnostic. Because the developers are embedded in the archival team, they can translate complex preservation standards into internal tools that streamline the archival workflow. Simultaneously, they leverage their expertise in user interface and web accessibility to design bespoke interfaces for complex data types, from spatial and Geographic Information System (GIS) layers to sensitive osteological records. This symbiotic relationship allowed for the rapid, iterative design of systems that meet high-level digital curation standards and address the practical, day-to-day realities of archaeological data management at scale.

The Human Factor: Communication and Continuity

While new and improved tools addressed the internal throughput of HS2 data, technical automation reached its limit when confronted with the logistics of data deposition. Managing the unprecedented volume of deposits required a shift in how the ADS communicated with depositors.

Of note, the hindrance caused by communication delays was underestimated. For many depositors, archiving is the final, often under-resourced step in a long project pipeline. However, as a FAIR-compliant repository, the ADS does not act as a passive recipient. Active curation is required to safeguard data integrity and maximise reuse potential. Shifting this paradigm required the ADS to collaborate with depositors and emphasise that deposition is a shared process built on mutual responsibility.

Resolving issues such as missing metadata, corrupted images, or broken external references in Computer-Aided Design (CAD) and GIS files required proactive communication. Direct and transparent feedback was provided, explaining the rationale behind required changes, such as policy alignment or archival structure. Expectations were managed appropriately, stating the current status of an archive to maintain momentum after initial tasks concluded. As previously mentioned, concrete resources were provided, including specialised tables with accepted formats and metadata templates for every data type and direct links to the ADS Guides to Good Practice (ADS, n.d.), giving depositors an authoritative reference to use internally.

Administrative and Archaeological Logic, Continuity and Care

A consistent theme throughout this project has been the tension between flexibility and control. It was necessary to keep deposits open and forgiving, which meant accepting hard drives, email attachments, and cloud storage links to ensure the data was received. However, this flexibility resulted in poor version control and inconsistencies. These issues challenged the ability to ensure data integrity and provenance, a concern particularly acute when working with sensitive data, such as human remains.

A key lesson from this project is that organising correspondence and documentation is a skill in its own right. During the deposition process, information was often received about multiple datasets in a single email thread. However, the extent to which this would later complicate information retrieval and task tracking was underestimated. Establishing stricter protocols, such as requiring depositors to limit each email to a single site code, would have significantly streamlined information retrieval and saved valuable time. The "Uploads" tab in the CMS, which stores copies of correspondence alongside the relevant dataset, proved indispensable for collection handovers, allowing team members to assess a collection's status and identify data issues. Linking project metadata directly to the collection record was essential for long-term continuity.

HS2 also revealed the limits of expanding capacity solely through labour distribution; an early strategy of dividing archive processing by data type proved counterproductive. While efficient for larger archives; it led to a siloed understanding of the data and a duplication of effort. This was particularly time-consuming when depositors had moved on or original data creators were no longer available. This demonstrates that digital preservation is not only about labour distribution, but also continuity, context, and care. As Zoldoske (2024) noted, metadata becomes both harder to create and less reliable the further removed one is from its origin, a reality that applies to archivists and depositors.

Reimagining the Archive: Discovery and Reuse

The HS2 Digital Footprint

At the project's outset, the size of HS2's Phase One data was anticipated to be approximately 15 terabytes. This figure was quickly surpassed; data from the largest site totalled 2.8 terabytes on deposit, with others averaging 12.55 gigabytes each. Data were deposited from more than 360 sites along the line from London Euston to Birmingham, spanning a diverse range of data types from standard text, spreadsheets and images to specialised three-dimensional (3D) models, GIS, geophysics, CAD data, audiovisual records and relational databases (HS2 Ltd., 2023a). The ADS had managed large digital archives before, such as the multisite archive for the Channel Tunnel Rail Link (Foreman, 2018). HS2 was exceptional because of its volume and complexity.

Figure 3 shows the unique nature of the HS2 data compared with the ADS's other collection types. Although the 360+ HS2 archives are only a fraction of the ADS's total number of collections, the volume and size of the files from these archives have resulted in data from this project representing a significant share of its total holdings.

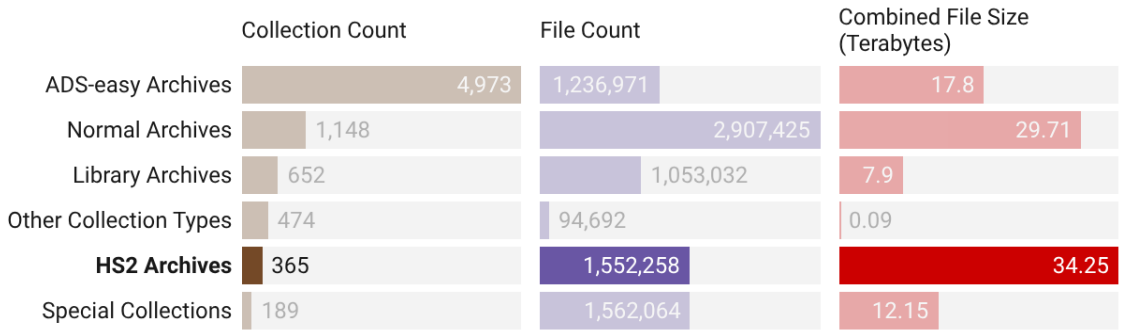

Figure 3. The Archaeology Data Service's (ADS) total holdings by collection type, showing number of collections alongside the total number of files and combined size of files within those collections. Collection types include data deposited using the ADS-easy submission system, collections related to the ADS Library and datasets not deposited using the ADS's submission systems (Special Collections or "Normal Archives").

In addition to technical variations in scale, with some archives containing only a few PDF reports and others terabytes of high-resolution photogrammetry, collections in the HS2 Historic Environment Digital Archive vary enormously in content. Archaeological intervention types encompass everything from building recordings to landscape-scale assessments. To maintain coherence across such a vast array of data, the ADS implemented rigorous controlled vocabularies, ensuring that precise subject terms for monument types, event types, and object classifications captured the specific archaeological context of each site while remaining interoperable within the larger project framework.

This diversity encompasses the sociopolitical context of the project by preserving contrasting perspectives. For example, while an archive containing data from historic setting recording of Birmingham's heritage assets includes promotional videos highlighting the project's public benefits (HS2 Ltd., COPA: Cotswold Oxford Pre-Construct Archaeology, & Cotswold Archaeology, 2024), a separate recording in Coleshill captures the local resistance. The latter includes protest ephemera, such as "STOP HS2" posters, found on a building demolished during enabling works for the HS2 project (COPA: Cotswold Oxford Pre-Construct Archaeology, Cotswold Archaeology, & HS2 Ltd. 2025). Together, these elements showcase the archive as a comprehensive record of the landscape's physical transformation and the cultural tensions surrounding it.

From Archaeological Intervention to Digital Collection

Infrastructure projects such as HS2 challenge existing workflows and prompt questions on the definition of a "site" or "project." In UK commercial archaeology, a "site" typically denotes a specific geographical area of construction rather than archaeological features that categorise a settlement. This pragmatic definition is necessary for field data, but conflicts with ontological debates on the fluidity of archaeological boundaries (e.g., Dunnell, 1992).

The data gathering process heavily informed how the HS2 data were deposited. The project was divided into several phases, work packages, and sites. The ADS generally disseminates data in collections corresponding to how it was deposited, resulting in one collection per site. In HS2, it became clear that adaptable curatorial and dissemination practices would be required to effectively present such a diverse dataset. To aid navigation between connected datasets, HS2 data was disseminated using a top-down approach, with an overarching "parent collection" for the HS2 project. Most collections are site-specific and a small number contain scheme-wide survey and landscape assessment data spanning multiple areas.

Limitations of this intervention-based model are illustrated by work undertaken at St James's Gardens Burial Grounds in London. The site was investigated using various methodologies across multiple phases, resulting in multiple datasets deposited in the ADS. These include trial trenching between 2017 and 2022 (MOLA Headland Infrastructure & HS2 Ltd., 2025a), an extensive open area investigation between 2018 and 2022 (MOLA Headland Infrastructure & HS2 Ltd., 2025b), a survey of a monument to the Christie family (MOLA Headland Infrastructure, MOLA Northampton, & HS2 Ltd., 2023), and a volunteer-led transcription project (MOLA Northampton & HS2 Ltd., 2023), each with its own persistent identifier and discovery metadata.

Archive structures that reflect how, when, and why data was collected fragments the record for end-users. Archiving data in a way that contextualises its origin while facilitating holistic research remains an ongoing challenge.

Results: Disseminating Complex Data

As of March 2026, the 362 HS2 collections for which data have been deposited from the initial London to West Midlands phase of work along the HS2 route have been archived and are available online (HS2 Ltd., 2023a). Over 600,000 files have been accessioned, and at 34 TB, the total preservation footprint of HS2 data constitutes just over one-third of the ADS's total holdings. As aforementioned, such a significant volume of data relating to a single project presented novel challenges for the ADS. Disseminating such a significant volume required an explicit connection to the wider HS2 scheme while ensuring collections remained accessible as standalone resources. The FAIR principles (Wilkinson et al., 2016) were applied to the aggregate HS2 dataset and to the individual collections in it.

Enhanced Collection Discovery

All HS2 collections were indexed to be searchable alongside other ADS collections. For example, a search in the ADS Data Catalogue for "statues in London" would return a building survey of the statue of Robert Stephenson carried out during the HS2 Historic Environment Programme (HS2 Ltd., Mott MacDonald, MOLA Headland Infrastructure, 2024). This removes the need for users to engage with the wider project context when conducting unrelated research. However, a unified map and search interface in the HS2 Parent Collection (HS2 Ltd., 2023a) also allows all HS2 data to be queried in one place.

A bespoke interactive map has been developed to visualise scheme-wide spatial data, integrating intervention areas, archaeological features, and significant finds (Figure 4). Users can perform granular queries, such as filtering for Medieval features or specific object types such as brooches, to navigate the dataset intuitively. Currently populated with primarily evaluation data, it will expand to include post-excavation analysis data in the coming years. Operating as part of the "parent collection," it provides a high-level entry point that allows users to discover data without prior knowledge of site names or archival structures. Interactive pop-ups provide essential context, including object types, site codes, and direct links to the relevant HS2 collections.

Complementing the map, a unified HS2 Search interface facilitates discovery across the entire corpus. By bridging the OMS and the CMS, the tool aggregates object- and collection-level metadata into a single point of access. This allows users to search the full breadth of the HS2 Historic Environment Programme, from high-level collections to individual objects and files. Architecturally, the system uses ColdFusion and SQL-driven database views to dynamically retrieve results. This approach eliminates the need for redundant data storage, ensuring a lean and efficient infrastructure.

To refine these results, users can utilise a comprehensive set of facets that return objects strictly matching selected criteria. These include location (county and named place), subject matter (archaeological monuments, features, or specific objects), temporal data (relevant archaeological or historical periods), administrative metadata (intervention types, creation dates, and data creators), technical identifiers (ADS Object IDs, Collection IDs, and site codes), and data specifications (filenames and data types).

This cross-collection search functionality mitigates the "silo" effect typical of traditional archival structures by bridging the gap between the deposit (the collection) and the individual record (the object). This approach significantly improves the accessibility and discoverability of diverse datasets within the HS2 programme, enhancing accessibility for professional researchers and the general public.

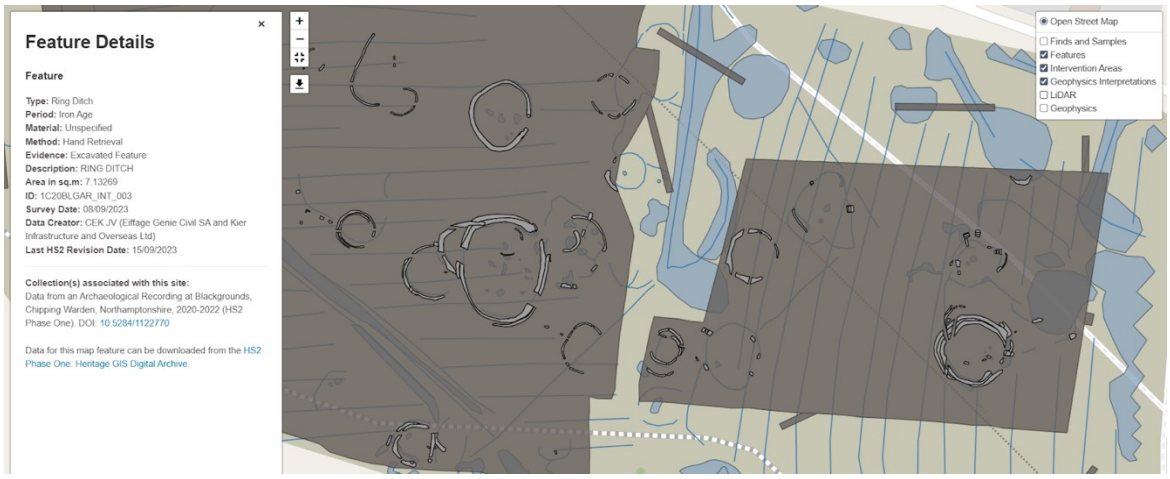

Figure 4. Capture of the High Speed Twp (HS2) parent collection map (HS2 Ltd., 2023b), showing a group of Iron Age ring ditches near Chipping Warden. Map layers include archaeological features, intervention areas, and geophysical data interpretations.

Bespoke Interfaces and Visualisation

Certain complex collections, such as the St James's Gardens burial data (MOLA Headland Infrastructure & HS2 Ltd., 2025b), required bespoke query interfaces to address their unique requirements. This specific collection contains detailed records for 11,000 burials excavated from a Victorian burial ground in Camden, London, between 2017 and 2021. Rather than providing a standard downloads page where users view files grouped by data type or a basic structure, a dedicated "Special Collection" interface was created. This system, powered by the ADS Special Collection database and populated with data from collections, allows users to query and access specialist burial and bioarchaeological data directly within the archive.

Search options in this interface include burial type, burial location, and finds associated with the burials, as well as specific skeletal age, sex, and pathology data. The data can also be searched spatially, with a map of the burial ground linking through to burial data (Figure 5). The resulting data pages display these details for each burial alongside site photographs, which are dynamically linked using burial IDs. By providing these specialist search and display options, the ADS enables researchers to interrogate the burial data directly in the browser interface. This approach significantly enhances access to and reuse of the archive by removing the need for users to download and manually reconstruct the database to perform complex analyses.

Figure 5. Capture of the map in the St James's Gardens Burial Ground excavation archive (MOLA Headland Infrastructure & HS2 Ltd., 2025b).

This bespoke approach also allowed ethical concerns surrounding the dissemination of sensitive content to be addressed. The St James's Gardens Burial Ground included individuals interred as recently as the 1850s (Blanks et. al., 2022) and public concerns were raised over the exhumation (BBC, 2017). To facilitate a bespoke solution for this dataset, the bioarchaeological data were used to identify human remains containing features such as preserved soft tissue, which users accessing the data may not expect to see. These records were tagged in the ADS OMS database, and relevant images in the archive interface are now initially displayed as blurred, with users having the option to click to view the content. This approach was informed by specialist subject knowledge within the ADS and consultation with data users (Bohling, 2025). The dissemination of sensitive content is a growing concern in the sector (e.g., Márquez-Grant & Errickson, 2017), and the bespoke interface solutions used for several of the large HS2 cemetery archives have acted as a pilot that will inform how the ADS disseminates sensitive content in the future.

Enhanced Data Visualisation

Tools for enhanced data visualisation for specialist data types were also improved or developed concurrently with the HS2 Historic Environment Programme. This included new preview options for GIS data, with interactive previews of shapefiles added to archive interfaces. These previews are generated automatically from GIS data, replacing the static images of GIS formerly created manually by archivists. In addition to saving time for archival staff, this interface enhancement allows end users to view shapefiles, rendered via the Leaflet JavaScript library, directly in a web map in the browser. This removes the need for users to download files and use specialist GIS software to view the data.

To improve the end-user experience, a multilayered 3D viewer was also created. In 2025, Applications Developer Adam Fox, in collaboration with Marco Callieri from Italy's Consiglio Nazionale delle Ricerche, developed a customised version of 3D Heritage Online Presenter (3DHOP) (Potenziani et al., 2015). Integrated as a ColdFusion custom tag using a new JavaScript framework, this tool provides an in-browser preview of 3D models directly on the object information page. This development builds on HS2's early emphasis on incorporating audio and video assets into the archive, ensuring a rich multimedia record.

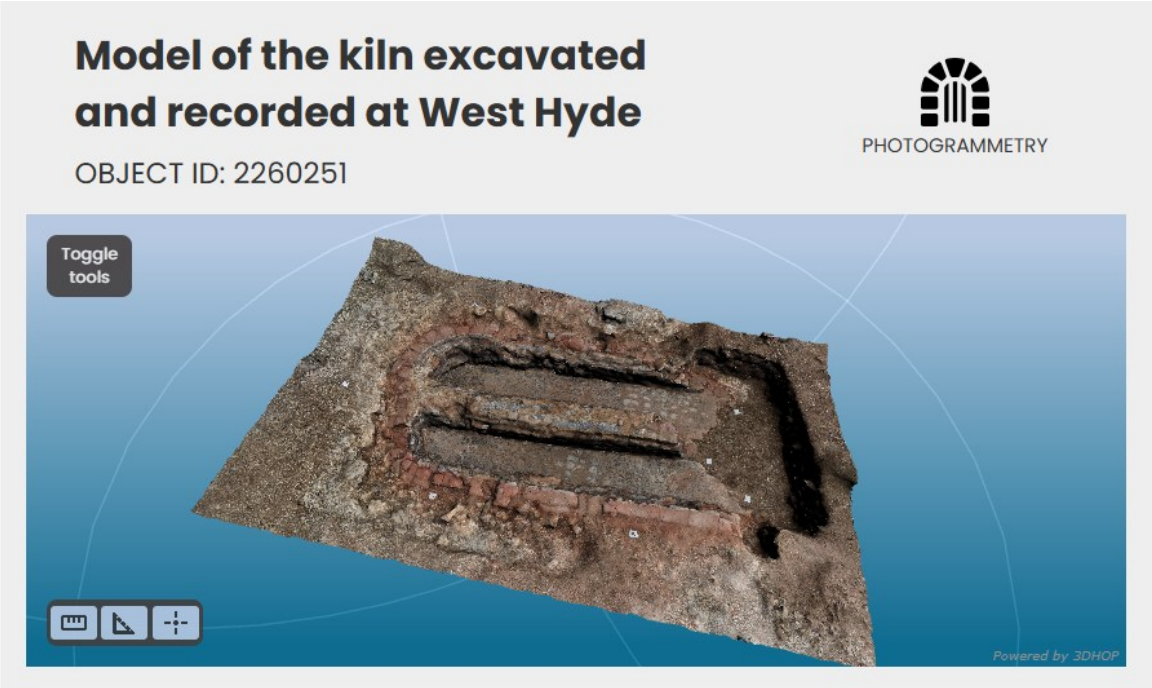

Figure 6. Capture of a 3D model displayed using the 3DHOP Viewer on the ADS object information pages. The model is from an archaeological recording of a kiln at West Hyde (HS2 Ltd., MOLA Headland Infrastructure & MOLA Northampton, 2025).

Curated Collections

One of the most successful outcomes has been the reuse framework developed for HS2 and the creation of "Curated Collections," which are thematic resources designed to aggregate data from various HS2 archives. These collections aim to make complex archaeological datasets accessible to a broader audience while encouraging the reuse of digital assets.

Applications Developer Jo Gilham developed a new template that differs from the standard ADS archive format. A curated collection brings together objects from different collections and displays them as "cards" on a simplified archive page (Figure 7). Each card features enhanced titles and descriptions, illustrated by a representative image that can be customised. These collections are defined in a new dedicated table in the Special Collections database, and each card links directly to the original object information page. This relational structure ensures that any updates made to object metadata are automatically reflected in the curated collection, maintaining data integrity across all dissemination methods. In the simplified presentation, the user is only a couple of clicks away from the technical metadata, the original record, and download options.

The scope of these collections ranges from peculiar burial practices to assemblages of everyday items, such as combs and toys (ADS, 2026a; 2026b), recovered during the HS2 Historic Environment Programme. By grouping thematically similar objects and providing detailed reuse statements, the potential (and the limitations) of the data can be communicated to researchers and the public. For example, a curated collection of everyday objects can turn individual finds into an assemblage that supports the understanding of personal identity and material culture in the past. Using this curated approach, the ADS provides interpretative tools to unlock the full reuse potential of the HS2 Historic Environment Digital Archive.

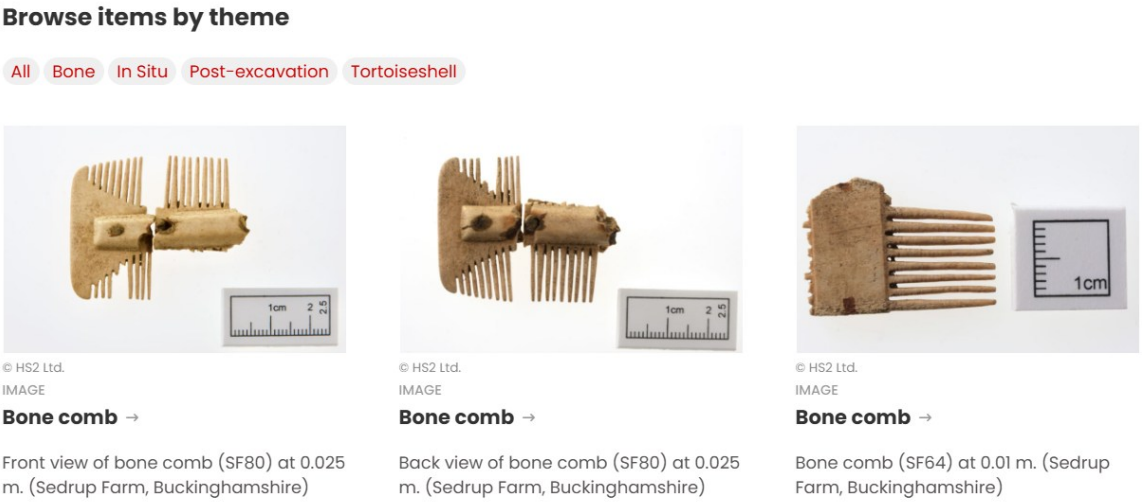

Figure 7. Capture of objects displayed in the "Combs in Context" curated collection (ADS, 2026a).

In summary, enhanced interfaces and data visualisation tools have significantly improved the accessibility of HS2 data. High-level search options allow data to be viewed and queried in its entirety. Bespoke "Special Collection" interfaces allow users to query specialist data, such as bioarchaeological records. Interface enhancements allow users to view GIS data or 3D models without specialist software. Finally, curated collections provide an alternative entry point to the HS2 data, allowing users to browse subsets grouped thematically. These options to search and view complex and specialist archaeological data lower the technical barriers for end users accessing and reusing the HS2 Historic Environment Digital Archive.

Conclusion and Reflections: Keeping the HS2 Digital Archive on Track

At the time of writing, the HS2 Historic Environment Digital Archive is an ongoing project. However, the challenges faced to date have arguably impacted the operations and ways of working more than any other undertaking during our 30-year history. The complexities of dealing with an unprecedented volume of work and data were magnified by the scale of the project. Working within specific external requirements provided an opportunity to enhance workflows, develop new technical tools and rethink our understanding of often overlooked commonplace notions, such as what is a "collection" or a "site." Although there may be no definitive answers to all the questions that emerged from this experience, archiving digital archaeological data from the HS2 project has pushed the boundaries of how we work.

On reflection, the ADS has been transformed in its technical capacity and in its understanding of the infrastructural, organisational, epistemological, and ethical dimensions of large-scale digital curation. Challenges were real and often overwhelming; however, they offered a proving ground in which future approaches to archaeological data management and archiving could be tested. This has led to new tools, adaptable procedures, better workflows, and more nuanced perspectives on the value of curated data. Perhaps most importantly, it has prompted reflection on how to deal with situations that surpass capacity, how to tackle constantly evolving scenarios, who data is curated for, and how to ensure that access and understanding are sustained long-term. This reflection mirrors broader sectoral calls for a culture shift that values sustainable, standardised digital preservation (Richards, in press) and reinforces the need to actively design archives as dynamic spaces for specific audiences (Opgenhaffen, 2022).

More broadly, HS2 has brought a long-standing sectoral challenge into sharp focus. As early as the 1970s, figures such as Frere (1975) and Cunliffe (1990) expressed concerns about the publication of rescue excavations, noting that archaeology was "drowning in data." Large-scale infrastructure projects such as HS2 may be said to accelerate this historical trend. While the establishment of digital repositories such as the ADS provides essential infrastructure to manage these outputs (Richards, 2017), the volume of modern data remains a significant pressure point. It highlights a critical distinction: high volumes of data do not automatically equate to high-quality information.

For instance, on several HS2 sites, rapid capture methods such as structure-from-motion photogrammetry were used to document complex features and minimise time on site (Sparrow, Bain, Kimber, & Wilson, 2024). However, this speed can come at a cost, and if priority is not given to recording metadata and context at the point of capture, the ultimate interpretative value of the digital record can be permanently diminished (Zoldoske, 2024; Urankar, Krajšek, & Lipovec, 2024). Quantity must be balanced with quality; large datasets require metadata that is rich to the same extent and showcases the intention behind their creation as well as foresight for their discovery, accessibility, and usability. The allure of state-of-the-art technologies must not override the necessity for purpose-driven data creation, especially in contexts involving human remains or culturally sensitive materials (Márquez-Grant & Errickson, 2017).

Strategic Reflections and Operational Evolution

The HS2 project was undertaken during a period of notable growth for the ADS, which occurred alongside an increasing volume of non-HS2 archaeological data being deposited. Navigating this combined workload required strategies to be refined. The lessons learned from the first phase of HS2 are now directly informing procedures for handling large infrastructure projects and standard data deposits.

One of the most significant shifts has been the closer integration of developer expertise in the archival workflow. By allocating dedicated developer time to support our archivists, manual workarounds towards programmatic fixes have been reduced. This collaboration has fostered a deeper institutional understanding of how technical automation can preserve archival integrity.

While continuing to maintain the practice of splitting larger archives by data type for larger projects, strategies for managing the sheer scale of these collections have been refined. We are actively moving towards embedding automation tools, such as the successes seen in automated batch processing of data, directly into archiving procedures to streamline and increase consistency in the preservation workflow.

Large-scale projects are highly susceptible to cognitive fragmentation. This has been addressed by implementing stricter communication protocols, such as requiring depositors to use discrete communication threads for different sites. This ensures that vital metadata is not lost amid the "administrative noise" or the high volume of correspondence generated by large-scale projects.

The scale of HS2 also exposed the limits of traditional file-sharing and storage permissions. Instances of duplicate datasets due to access restrictions highlighted that legacy methods were unsustainable. This has prompted a move towards more robust, centralised access management within internal holdings, ensuring team members have appropriate rights to view and edit data without compromising integrity.

Beyond HS2: A New Standard for Archiving

The unprecedented workload of the HS2 archive was a trial by fire for our capacity; it arrived with significant funding and a large degree of professional autonomy. This institutional trust granted flexibility to make high-level appraisal and archival decisions, create tools that streamline complex preservation workflows, and design enhanced interfaces for complex data types, such as multilayer 3D viewers and interactive map interfaces. These approaches, although largely born out of the pressures of the HS2 Historic Environment Programme, are now implemented across all ADS deposits, large and small.

Of note, these lessons have heavily influenced the development of the ADS's new Ingest system, designed to replace the legacy ADS-easy platform. For example, this system incorporates a dedicated messaging function that automatically links pre-deposit correspondence with its respective dataset. By automating these administrative links, greater transparency is provided, and time is saved for both archivists and depositors.

The HS2 Historic Environment Digital Archive has functioned as a catalyst for a more resilient ADS. As the next stage of HS2 data is prepared, including extensive post-excavation analysis, the strategies forged in the first stage ensure that the archive is not merely kept on "on track," but that a new trajectory for digital curation in the 21st century is actively being set.

References

ADS. (n.d.). *Guides to good practice*. Archaeology Data Service. https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice

ADS. (2022a). *Instructions for depositors*. Archaeology Data Service. https://archaeologydataservice.ac.uk/help-guidance/instructions-for-depositors/

ADS. (2022b). *File-level metadata*. Archaeology Data Service. https://archaeologydataservice.ac.uk/help-guidance/instructions-for-depositors/file-level-metadata/

ADS. (2026a). *Combs in context: Finds from the HS2 archaeology programme*. Archaeology Data Service. https://archaeologydataservice.ac.uk/archives/collections/view/1008960/

ADS. (2026b). *Traces of play: Figurines and dolls uncovered on the HS2 route*. Archaeology Data Service https://archaeologydataservice.ac.uk/archives/collections/view/1008967/

Aryankhesal, F. F. (2023). HS2 *Phase one: Heritage GIS digital archive* (Data paper). *Internet Archaeology*, HS2 Publications. https://doi.org/10.11141/ia.hs2.1

BBC. (2017, August 23). *London Euston HS2 exhumation plan prompts memorial service*. www.bbc.co.uk/news/uk-england-london-41020759

Blanks, A., Goodman, L., Catherine, D., Betts, I., Hartle, R., Davidson, H., ... Fowler, L. (2022). *Post excavation assessment at St James's Gardens*. High Speed Two Ltd. https://doi.org/10.5284/1122288

Bohling, S. (2025). Online dissemination of 3D bioarchaeological data: An exploration of ethics, user preferences, and contextualisation in an official digital repository setting. *Internet Archaeology, 69*. https://doi.org/10.11141/ia.69.9

COPA: Cotswold Oxford Pre-Construct Archaeology, Cotswold Archaeology, & High Speed Two Ltd. (2025). *Data from a historic building and setting recording at Capitol Joinery Timber Yard, Middleton House Farm, Coleshill, Warwickshire, 2020–2023* (HS2 Phase One) [data set]. Archaeology Data Service. https://doi.org/10.5284/1135581

COPA: Cotswold Oxford Pre-Construct Archaeology, & High Speed Two Ltd. (2026). *Data from an archaeological recording at Fleet Marston Cottages and Putlowes, Aylesbury, Buckinghamshire, 2021* (HS2 Phase One) [data set]. Archaeology Data Service [distributor] https://doi.org/10.5284/1122727

Cunliffe, B. (1990). *Publishing in the city. Antiquity, 64*(244), 667–671. https://doi.org/10.1017/S0003598X00078595

Dunnell, R. C. (1992). The notion site. In J. Rossignol & L. Wandsnider (Eds.) *Space, time, and archaeological landscapes* (pp. 21–38). Springer. https://doi.org/10.1007/978-1-4899-2450-6_2

Foreman, S. (2018). Channel Tunnel Rail Link Section 1. York: Archaeology Data Service [distributor]. https://doi.org/10.5284/1000230

Frere, S. S. (1975). *Principles of publication in rescue archaeology: Report by a working party of the Ancient Monuments Board for England*. Committee for Rescue Archaeology.

Gaffney, V., & Exon, S. (1999). From order to chaos: Publication, synthesis and the dissemination of data in a digital age. *Internet Archaeology, 6*. https://doi.org/10.11141/ia.6.3

HS2 Ltd. (2017, September 6). HS2 Phase one historic environment research and delivery strategy. Birmingham: High Speed Two Ltd. www.gov.uk/government/publications/hs2-phase-one-historic-environment-research-and-delivery-strategy

HS2 Ltd. (2023a). *HS2 historic environment digital archive* [data set]. Archaeology Data Service. https://doi.org/10.5284/1113008

HS2 Ltd. (2023b). *HS2 Phase one: London to West Midlands historic environment digital archive* [data set] Archaeology Data Service. https://doi.org/10.5284/1110091

HS2 Ltd., COPA: Cotswold Oxford Pre-Construct Archaeology, & Cotswold Archaeology. (2024). *Data from historic landscape recording at Birmingham Communications and Communities, Birmingham, West Midlands, 2022–2023 (HS2 Phase One)* [data set]. Archaeology Data Service. https://doi.org/10.5284/1122732

HS2 Ltd., MOLA Headland Infrastructure & MOLA Northampton. (2025). *Digital archive from archaeological recording of a kiln at West Hyde Embankment, Hertfordshire, July 2019 (HS2 Phase One)* [data set]. Archaeology Data Service. https://doi.org/10.5284/1135491

HS2 Ltd., Mott MacDonald, & MOLA Headland Infrastructure. (2024). *Data from a building survey at Grade II listed statue of Robert Stephenson, Camden, 2020 (HS2 Phase One)* [data set]. Archaeology Data Service. https://doi.org/10.5284/1119901

Hills, C. (1993). The dissemination of information. In J. Hunter & I. Ralston (Eds.), *Archaeological resource management: An introduction* (pp. 215–224). Sutton.

Márquez-Grant, N., & Errickson, D. (2017). Ethical considerations: An added dimension. In D. Errickson & T. Thompson (Eds.), *Human remains: Another dimension. The application of imaging to the study of human remains* (pp. 193–204). Academic Press. https://doi.org/10.1016/B978-0-12-804602-9.00015-1

MOLA Headland Infrastructure & HS2 Ltd. (2025a). *Digital archive from archaeological excavations at St James's Gardens Burial Grounds, Euston, Camden, Greater London, 2018–2022 (HS2 Phase One)* [data set]. Archaeology Data Service. https://doi.org/10.5284/1122289

MOLA Headland Infrastructure & HS2 Ltd. (2025b). *Digital archive from archaeological trial trenching at St James' Gardens, Euston, Camden, Greater London, 2017-2022 (HS2 Phase One)* [data set]. Archaeology Data Service. https://doi.org/10.5284/1122717

MOLA Headland Infrastructure, MOLA Northampton, & HS2 Ltd. (2023). *Data from a building survey at the monument to the Christie family, St James's Gardens, Euston, Camden, 2019 (HS2 Phase One)* [data set]. Archaeology Data Service. https://doi.org/10.5284/1119904

MOLA Northampton & HS2 Ltd. (2023). *Data from stories of St James's Burial Ground, Euston, Camden, Greater London, 2021-2023 (HS2 Phase One)* [data set]. Archaeology Data Service. https://doi.org/10.5284/1118189

Opgenhaffen, L. (2022). Archives in action. The impact of digital technology on archaeological recording strategies and ensuing open research archives. *Digital Applications in Archaeology and Cultural Heritage, 27*: e00231. https://doi.org/10.1016/j.daach.2022.e00231

Potenziani, M., Callieri, M., Dellepiane, M., Corsini, M., Ponchio, F., & Scopigno, R. (2015). 3DHOP: 3D Heritage Online Presenter. *Computers & Graphics, 52*, 129–141. https://doi.org/10.1016/j.cag.2015.07.001

Richards, J. D. (2004). Online archives. *Internet Archaeology, 15*. https://doi.org/10.11141/ia.15.7

Richards, J. D. (2017). Twenty years preserving data: A view from the UK. *Advances in Archaeological Practice, 5*(3), 227–237. https://doi.org/10.1017/aap.2017.11

Richards, J. D. (in press). Digital archiving and preservation. In E. Barker, O. Bobou, & R. Raja. (Eds.) *The Oxford handbook of digital classical studies*. Oxford University Press.

Sparrow, T., Bain, K., Kimber, M., & Wilson, A. S. (2024). Visualising heritage: Using 3D immersive technologies to innovate, document and communicate rich narratives for HS2. *Internet Archaeology, 65*. https://doi.org/10.11141/ia.65.7

UK Parliament. *High Speed Rail (Preparation) Act 2013*, c. 56. www.legislation.gov.uk/ukpga/2013/31/contents

Urankar, R., Krajšek, J., & Lipovec, B. (2024). Archaeologists, hunter-gatherers of digital data. *Internet Archaeology, 65*. https://doi.org/10.11141/ia.65.2

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*(1): 160018. https://doi.org/10.1038/sdata.2016.18

Zoldoske, T. (2024). Metadata for discovery: Planning for an information network. *Internet Archaeology, 65*. https://doi.org/10.11141/ia.65.6