Collecting Information and Developing Narratives: the use of data on HS2 Phase One, UK

John Halsted

A visual representation of the stones from the recently unearthed Roman road at Blackgrounds, Northamptonshire, being recorded with modern technology. Image credit: HS2 Ltd

Large Infrastructure projects create vast amounts of data during the course of programmes of archaeological investigation, from the description of an archaeological deposit to complex three-dimensional survey data. It is key for future research and the completion of the archaeological programme that the data support the questions we wish to answer.

This paper will consider the range of data generated from HS2 Phase One, and the potential of those data in the process of analysis and interpretation and their broader spatial and research context . The paper will also consider which data are key for

different stages of the project lifecycle, and the extent to which the process of data capture may influence the narratives that are developed.

1. Introduction

[HS2 Phase One](#) represents the largest single programme of historic environment work undertaken in the UK. The data collected from this programme will, therefore, be considerable. It is necessary to consider how those data may be approached, how a series of research questions may be addressed and the potentially different scales of analysis that may be employed. This paper builds on a presentation given at a session on 'big data', major infrastructure projects and archaeology at the 2022 [European Association of Archaeologists](#) conference and aims to outline the types of data generated by HS2 Phase One and the potential ways in which this may influence archaeological narratives.

The Historic Environment Research and Delivery Strategy (HERDS; HS2 Ltd [2017](#)) sets out a series of specific research objectives and it is important to consider how the data we have collected during the programme of survey and investigation may be used in addressing those questions. The strategy covers all aspects of the historic environment from palaeo-environmental data to historic buildings. This paper will focus primarily on archaeological themes.

2. The Historic Environment Research and Delivery Strategy (HERDS)

The HERDS sets out a broad strategy for historic environment research within the HS2 Phase One scheme (HS2 Ltd [2017](#)). It is underpinned by a series of supporting technical standards reflecting industry best-practice for both fieldwork methodology and data delivery, including GIS data schema and guidance documents.

A series of Specific Objectives were set out, which included research questions or Knowledge Creation objectives which were considered at route wide, regional and locally specific scales across archaeological periods. These objectives ranged from settlement location in earlier prehistory, regional patterns in later prehistoric and Romano-British settlement and trends in medieval settlement and agriculture through to cemetery management and burial practice.

The delivery mechanism for field surveys and investigation were set out in Project Plans and Location-Specific Written Scheme of Investigation (LSWSI) documents which detailed the agreed methodologies. The Project Plans set out fieldwork methods in response to the Specific Objectives and the known historic environment context. LS-WSIs set out the programme of works and methodologies within a given

area which was usually determined by the construction land parcels or other practical considerations.

3. The potential of the data and scales of analysis

HS2 Phase One is a linear scheme c.200km in length traversing a diverse number of geologies, topographies, regions and archaeological character areas in the midlands and southern England. The route has the potential, therefore, to identify many local and regional variations in material culture consumption, settlement morphologies and palaeo-environments across large timescales. At broader regional and national scales, the railway can be seen as a relatively narrow transect across the landscape, especially when compared to other big-data syntheses which have been recently undertaken in Britain and Europe, such as EngLaId, The Rural Settlement of Roman Britain or the Roman Hinterland Project in Italy (Gosden and Green 2021; Allen *et al.* 2017; Attema *et al.* 2022). This narrow transect may limit the scale of study when compared with these more extensive research projects. Nevertheless, there remains clear potential for both comparative analysis within the scheme, utilising the various forms and scales of intervention and for the future use of data for testing existing models and regional syntheses.

The data generated, in common with many other archaeological projects, range from the ultra-small-scale scientific data of human bio-molecular or palaeo-environmental data, for example, through various classes of artefacts, archaeological feature types and structural remains recorded at often large landscape-scale excavation sites. These multi-hectare excavations form only individual components of investigations across the 200km scheme. When viewed at the scale of the project those sites can themselves appear relatively small scale.

The archaeological excavations are a product and a component of a wider series of investigations. These include non-intrusive survey data including those from scheme-wide LiDAR and geophysical survey and the more localised and bespoke use of multi-spectral imaging, through to broad scale trial trench evaluation or topsoil sampling. The results from these investigations may also lend themselves to landscape scale comparative analysis. Any one of these classes of data can be studied and interpreted at a number of different scales depending upon the research questions being asked.

These data clearly, therefore, have the potential to be examined at multiple scales of analysis both within individual sites, inter-site comparative analysis and beyond through broader contextualisation. The data also have the potential to be examined in terms of methodological effectiveness for site identification and the ability to discern patterns of activity across landscapes and through time. In order to further consider how those data may contribute to the formulation of archaeological narratives, it is necessary to outline what data have been collected and how those data have been collected and categorised.

4. Categories and types of data

The data generated from a project of this scale is wide and varied and ranges from primary fieldwork records and on-site data collection, including the increasing use of digital tablet-based recording systems and databases, through to specialist data tables forming part of post excavation assessment, textual reporting and interpretative illustrations. The principal components of the data from the project include:

- Context data within spreadsheets, Harris matrices and reporting;
- Specialist artefactual, palaeo-environmental and scientific data - within reports, spreadsheets and databases;
- Non-intrusive survey data; and
- Spatial data-within GIS data deliverables and attribute tables.

Certain categories of data such as specialist reporting and associated spreadsheets will be accessible as either static tables within reports or as digital files as a component of a digital archive held by the [Archaeology Data Service](#) (ADS). Whilst considerable effort has been made to standardise specialist data reporting through the supply chain of organisations, these data formats will inevitably require further consolidation and standardisation if they are to enable inter-site comparative analysis at a scheme-wide level, across what have been three principal contract areas.

For spatial data deliverables it was recognised at an early stage of the project that standardisation was required and that a framework for delivering data would need to include the ability to record archaeological attribute data within specific GIS feature classes.

5. The GIS spatial data schema and potential for analysis and interpretation

A series of data schema were defined at the outset of the project which covered the key anticipated areas of delivery (see Ayrankhesal, [this volume](#)). The principal categories of data that are captured include:

- the site boundaries and extents (within Project Plan and Location-Specific Written Scheme of Investigation data);
- the extent of archaeological interventions (those areas excavated by archaeologists); and
- archaeological features and individual archaeological objects.

These were linked to databases for archaeological objectives (HERDS Specific Objectives) and document datasets (links to associated fieldwork reports for example) within a relational database. A list of activity types (fieldwork and survey

techniques) was generated which link to each of the data schemas and are captured within the data, alongside unique site codes.

Within the data schema are a number of elements which will help to categorise the results of fieldwork investigation for the purposes of visualising their spatial location and extent and to aid in querying those data for archaeological analysis. Where standardisation was required for Feature Class attributes, the Forum on Information Standards in Heritage (FISH) thesauri of terms was used.
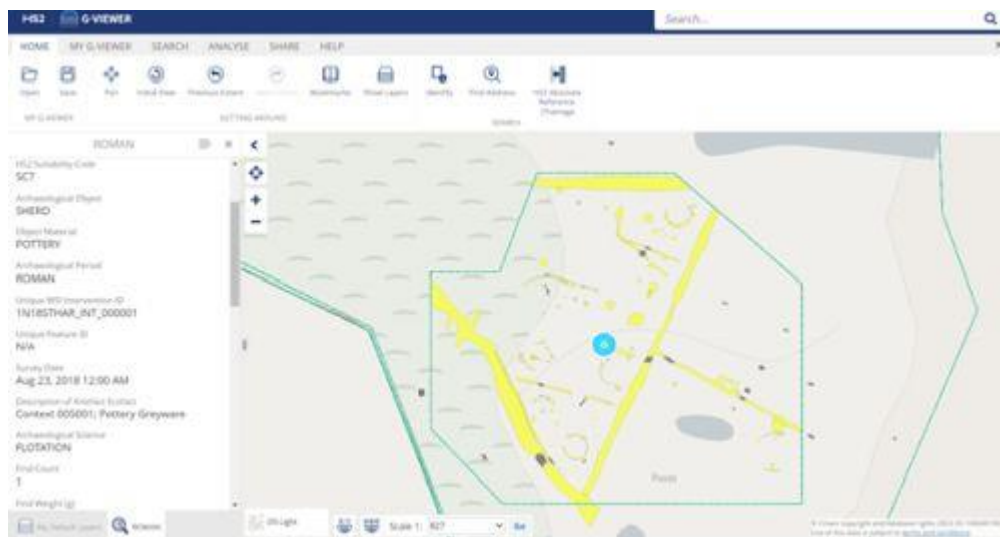
.

Figure 1: HS2 GIS data: colour-coded Romano-British features with an object selected, near Ryknild Street, Staffordshire. Image credit: HS2 Ltd

All archaeological features recorded in the field are able to be recorded within the GIS data, along with their spatial extent and maximum depth data (Figure 1; Figure 2). The latter may enable volumetric calculations of material excavated alongside archaeological intervention data also captured within the GIS and detailed contextual records contained within spreadsheets. Volumetric data has been highlighted as a necessary means of providing more meaningful comparative analysis of levels of material culture between archaeological sites, particularly for the Romano-British period (Fulford and Holbrook 2011).

Figure 2: Park Street and Freeman Street, Birmingham, archaeological feature data displayed in GIS. Image credit: HS2 Ltd

The facility to record the spatial location of archaeological objects was also built into the HERDS GIS schema. This enabled the recording of individual objects for either field surveys (such as metal-detecting) or the recording of certain classes of finds on site at the discretion of the excavators. At St James's Gardens post-medieval burial ground in London, the spatial recording of depositum plates has been undertaken, for example. This will assist in both the identification of named individuals and family groups and the potential differences in modes of burial across social classes in the Victorian period (Figure 3). The spatial recording of registered finds at the Romano-British roadside settlements at Fleet Marston (Akeman Street), Buckinghamshire and in the vicinity of the Roman road at Ryknild Street, Staffordshire for example (Figure 1 and Figure 4) may help to elucidate functional areas and the patterns of discard of those living adjacent to and in the hinterland of these arterial routeways. It may be possible to further develop this category of spatial data to facilitate further thematic spatial analysis during the post-excavation stages.

Figure 3: St James's Gardens, London, burial ground excavation and distribution points for depositum plates within GIS. Image credit: HS2 Ltd

The facility to capture and search by archaeological period has been included within the HERDS GIS schema and where possible these data have been ascribed (Figure 1). This will be a valuable tool at the post-excavation analysis stage allowing users to visualise the spatial extent of locations with specific period-based evidence. Clearly as the programme of post-excavation progresses and further chronological analysis takes place, data relating to archaeological periods may require revision.
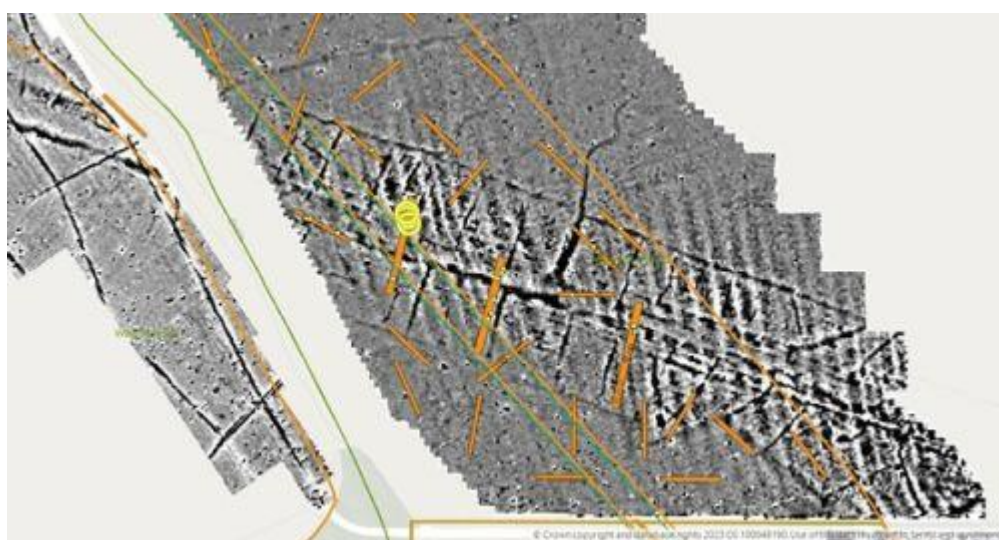

Figure 4: Fleet Marston, Bucks, selecting Archaeological Object data with trial trenching locations and geophysical survey results in GIS. Image credit: HS2 Ltd

The ability to capture Specific Objectives spatially within the data is also a useful means of visualising the spread of locations where objectives may be being

addressed. This will provide a useful mechanism for data to be inputted and displayed during the post-excavation stage of the project. Being able to display objectives within GIS data was also envisaged as a practical aid to the management of the fieldwork programme, in terms of which objectives were potentially being addressed where. The iterative nature of archaeological fieldwork and data delivery limited the usefulness of this element of the schema during the fieldwork stage.

6. Stages of spatial data delivery

For such a large-scale archaeological programme, data are inevitably delivered in stages at different points in the life of the project. Despite agreed programmes for data delivery, data can only be delivered once fieldwork and an appropriate level of assessment has taken place at any given location. Proposed fieldwork locations and designs will always have to be updated with archaeological data once investigations have actually taken place.

Although broadly progressing from survey and evaluation to investigation, those investigations are undertaken within the context of complex programmes of work. The programme and schedule of those works may be dictated by a variety of factors, such as ecological habitat mitigation or preliminary engineering works programmes where access to certain locations has to be prioritized and the timing of archaeological investigations has to be coordinated to take this into consideration.

In this sense the dataset available for analysis is to some extent fragmented until the later stages of fieldwork completion, no matter how frequently data are delivered. Exceptions to this may include geophysical survey data or remote sensing data such as LiDAR which are captured at large scales in advance of subsequent fieldwork or the use of data captured from external sources such as a route wide paleo-environmental desk-based assessments with broad study areas (Howard and Hopla 2017; Brown *et al.* 2017). Any analysis of data obtained can, therefore, only begin to take place at a project-wide level, towards the end of the data delivery process. Decision making during the programme is inevitably, therefore, taken on the basis of data captured within the scheme to date, including interim reporting, alongside contextual data from broader non-intrusive survey or desk-based assessment. How these processes of data capture, delivery and decision making ultimately influence the archaeological narratives that are produced at the post-excavation stage is something for future consideration.

7. Data re-assessment and analysis-potential

Non-intrusive datasets have a significant influence upon the selection of locations for further archaeological investigation, they provide results at large scales and are easily incorporated within geo-spatial applications. Although it is recognised that the results from such surveys may not always be able to answer questions from all archaeological periods, they inevitably influence the location and frequency of

evaluation trenching for example, and often the final site selection for further investigations. A large linear project such as HS2 Phase One highlights the variability in the effectiveness of such surveys, whilst also providing opportunities to ground truth areas of positive anomalies alongside testing apparently 'blank' areas.

A pilot study and subsequent re-appraisal of LiDAR data for the northern section of the route (including parts of Warwickshire and Staffordshire; Cox *et al.* 2020) included more advanced data processing and visualization, which resulted in a number of additional landscape features mostly relating to medieval or post-medieval agriculture and boundaries. The data captured by this re-appraisal could be examined further through an assessment of any ground-truthing of these anomalies through intrusive investigations. The data may also be seen as a resource in itself, particularly for contextualising known medieval and post-medieval settlement.

A mid-stage review of the effectiveness of geophysical survey in the northern section of the route against interim trial trenching results highlighted a number of factors potentially resulting in the variability of the survey dataset (Ovenden and Appleby 2020). In areas where geophysical survey results appeared less effective at predicting the presence of archaeology several factors including a possible lack of magnetic contrast between feature fills and the natural geology have been discussed. At locations where human activity did not result in the deposition of significant magnetically enhanced residues or where levels of artefactual deposition were low, geophysical signatures may be reduced. Whilst it is also possible that natural soils and their magnetic susceptibility may be inherently low.

There is clearly potential to undertake further research into the effectiveness of non-intrusive techniques across varying geologies and topographies and potential for analysis against subsequent intrusive investigation results from both trial trenching and open area archaeological recording. Such work would enable a closer understanding of both the presence and absence of archaeological sites and the processes for the identification of those sites and the types of sites that are subsequently identified for further investigation. The appraisal of evaluation strategies and the influence of geophysical survey data is a theme that is currently being explored through PhD research, for example (Higham, forthcoming), and evaluation has been recently assessed by the wider industry (CIFA 2022).

8. Creating narratives from data

A large transect through the landscape provides an opportunity for data to be used for inter-site comparative analysis and to address landscape level objectives alongside other scales of analysis which will be necessary to address the local, regional and route wide specific objectives set out in HERDS.

Non-intrusive surveys and intrusive trial trenching are a significant component of identifying 'sites' on a linear scheme and are a significant factor in the identification of locations for further intrusive investigation, alongside evidence identified through desk-based assessment. These locations of open area investigation will form the principal component in addressing HERDS specific objectives and will inevitably be associated with the greatest level of associated data.

Whether drawn from the spatial data deliverables, or from associated specialist tables from within digital archives, associated unique site codes and other unique identifiers will enable inter-site comparisons to be made (for more detailed discussion see Ayrankhesal, this volume). This will be essential for those HERDS Specific Objectives where regional distinctiveness, for example, is looking to be drawn out. Where volumetric data has been captured for material excavated, this comparative analysis may be considered to be more reliable, since any differences in sample sizes between sites can be factored in.

All such identified sites do, nevertheless, reflect specific foci which were also an integral part of varying scales of human movement, social interaction and economic activities and a 'site' may be seen as a node of concentrated activity within broader landscape areas. Where our data plot excavated sites, therefore, these locations will only ever be one component of a more widely settled and traversed landscape with certain activities leaving more readily identifiable traces than others. For some archaeological periods, sites of any kind are difficult to identify, either through their infrequent presence or for inherently ephemeral modes of occupation, particularly for the earlier prehistoric periods. It is important to highlight, therefore, that programmes of blank area testing, including plough-soil test-pitting, were enacted across specific areas of the scheme in an attempt to capture less visible archaeological remains. Any analysis of the data at broad scales should, therefore, not only assess the sites identified and excavated (with their attendant concentrations of material culture and palaeo-environmental evidence,) but also the relative frequencies of less intensive activities across the landscape. This will enable a more holistic narrative of settlement and change over time across the areas with which the scheme has intersected.

The survey data collected by the project can also enable enhanced discussion of archaeological character areas. Where multiple data sources such as Digital Terrain Models captured from LiDAR data, along with geological data and extensive geophysical survey can be combined with intrusive investigations and field surveys, these can feed into landscape-level analyses of the past. Attempts to discern patterns in settlement across time and space, can be made both within landscape contexts and against the considerations of potential biases in the data (Gosden and Green 2021; Cooper and Green 2017). If used to its full potential spatial and GIS data could considerably aid the discussion of such themes and be an essential tool in analysing how people interacted with each other, their surroundings and established places. Artefactual data and comparative analysis of material consumption may add to a more holistic landscape scale approach with the aid of the spatial, survey and topographic data captured. Such analysis can form a useful basis for discussion in the context of social, political and economic narratives.

The results can clearly lend themselves to route wide or regional objectives, but that is not to preclude the detailed examination of individual sites, which can also lead to broader-scale insights into the past (cf. Gosden and Kirsanow 2006). At burial grounds, for example, data can be examined at a detailed site scale, from the snapshot in time reflected in the placing of personal possessions with the individual at the point of burial and the selection of burial plots and grave furniture as a reflection of social status at death. At a wider scale, stages within an individual's life can be explored through osteological and isotope studies, and through examining

biographical sources, family relationships and the study of genetic histories with ancient DNA.

Spatial feature and artefactual data can also be combined with cartographic and burial records to provide new understandings into both cemetery management and burial practices across time, social class, gender and age, for example. The ability to locate individuals spatially will be enhanced through subsequent specialist osteological and artefactual analysis, and contextual studies including the ongoing citizen science project Zooniverse, for example. Recent projects have also highlighted the usefulness of bringing cartographic sources within GIS -based analysis for the study of historical periods (Trepal *et al*. 2021). For our burial ground sites e.g. St James's Gardens in London, good cartographic sources exist and will be essential for combining with the excavation data for a study of cemetery management and development.

9. Conclusions

The data generated by HS2 Phase One has the potential to make a significant contribution to the range of period-based knowledge creation objectives established at the outset of the project, from the analysis of individual sites to the assessment of broad route-wide patterns. A systematic and consistent spatial data delivery held within GIS format will be a key feature for the study and retrieval of data for the project. This will, however, also need to be examined in tandem with the contextual and specialist reporting data held within the project digital archive, since those data submitted within specified schemas are only one element of a process of assessment and analysis. Having clear and systematic spatial data for site location and interventions will allow those supporting data to be easily tied to spatial location for analysis. The data held for those interventions will enable route-wide comparative analysis and form a basis for both generating and testing archaeological narratives.

The need for a closer attribution of data to archaeological periods is a reflection of the iterative process of data delivery and reflects the process of archaeological research through its interim, assessment and analysis stages.

In the longer term, the data will be a legacy resource for future research projects and have the potential to integrate with other datasets for a fuller contextualisation of the route. The testing of existing models and independent academic research is already capitalising upon the data held by the project. The outcomes of such research will be a significant contribution to knowledge alongside the results of the forthcoming programme of post-excavation analysis.

Bibliography

Allen, M., Lodwick, L., Brindle, T., Fulford, M., Smith, A. 2017 *The Rural Economy of Roman Britain*, Britannia Monograph **30**, Society for the Promotion of Roman Studies: London.

Attema, P.A.J., Carafa, P., Jongman, W.M., Smith, C.J., Bronkhorst, A.J., Capanna, M.C., DE Haas, T.C.A., VAN Leusen, P.M., Tol, G.W., Witcher, R.E. and Wouda, N.A. 2022 'The Roman Hinterland Project: Integrating Archaeological Field Surveys around Rome and Beyond', *European Journal of Archaeology* **25**(2), 238-258. https://doi.org/10.1017/eaa.2021.51

Brown, A.D., Hopla, E., Farrington McCabe, A., Generalski-Sparling, S. 2017 *Geo-Archaeological Desk Based Assessment (GDBA) review of the geo-archaeological potential of High Speed Two Phase One*, Atkins/ EDP report for HS2 Ltd. 1D037-EDP-EV-REP-000-000031

CIFA 2022 *valuation Strategies (Evals 1): understanding current practice and encouraging sector engagement*, WSP for the Chartered Institute of Archaeologists. https://www.archaeologists.net/sites/default/files/projects/EVALS%201%20Final%20Report%20for%20publication.pdf [Last accessed: January 2023]

Cooper, A. and Green, C. 2017 'Big Questions for large complex datasets: approaching time and space using composite object assemblages', *Internet Archaeology* **45**. https://doi.org/10.11141/ia.45.1

Cox, C , Jarvis, A and Appleby, J. 2020 *Detailed Desk Based Assessment: EIA LiDAR survey re-appraisal (Area North Extended Visualisations)*, LM/ DJV report for HS2 Ltd, 1EW04-LMJ_DJV-EV-REP-N000-029004

Fulford, M. and Holbrook, N. 2011 'Assessing the contribution of commercial archaeology to the study of the Roman period in England 1990-2004', *Antiquaries Journal* **91**, 323-345. https://doi.org/10.1017/S0003581511000138

Gosden, C. and Kirsanow, K. 2006 'Timescales' in G. Lock and B. Molyneaux (eds) *Confronting Scale in Archaeology: issues of theory and practice*, Springer: New York. 27-37. https://doi.org/10.1007/0-387-32773-8_3

Gosden, C. and Green, C. 2021 *English Landscapes and Identities: Investigating landscape change from 1500BC to AD1086*, Oxford University Press: Oxford. https://doi.org/10.1093/oso/9780198870623.001.0001

Higham, R. forthcoming 'Evaluating archaeological evaluation trenching strategies using GIS', PhD thesis for the University of Brighton.

Howard, A., and Hopla, E. 2017 *Scheme-wide palaeo-environmental Detailed Desk Based Assessment*, Atkins/ EDP for HS2 Ltd, 1D037-EDP-EV-REP-000-000033

HS2 Ltd 2017 *HS2 Phase One Historic Environment Research and Delivery Strategy*, https://www.gov.uk/government/publications/hs2-phase-one-historic-environment-research-and-delivery-strategy [Last accessed 1 January 2023].

Ovenden, S. and Appleby, J. 2020 Detailed Desk Based Assessment for review and assessment of geophysical survey, LM/ DJV report for HS2 Ltd, 1EW04-LMJ_DJV-EV-REP-N000-029002. https://doi.org/10.5284/1104426

Trepal, D., Lafreniere, D. and Stone, T. 2021. 'Mapping Historical Archaeology and Industrial Heritage: the historical spatial data infrastructure', *Journal of Computer Applications in Archaeology* **4**(1), 202-213. https://doi.org/10.5334/jcaa.77