

# Paper and pixels: historic maps as a multifaceted resource

Rebecca C. Roberts <sup>a</sup>, Junaid Abdul Jabbar <sup>a</sup>, Huw Jones <sup>b</sup>, Hector A. Orengo <sup>c</sup>, Marco Madella <sup>d</sup>, and Cameron A. Petrie <sup>\*a</sup>

<sup>a</sup> *McDonald Institute for Archaeological Research, University of Cambridge, rcr45@cam.ac.uk, ja738@cam.ac.uk, cap59@cam.ac.uk*

<sup>b</sup> *University Library, University of Cambridge*

<sup>c</sup> *ICAC, University of Taragona*

<sup>d</sup> *Universitat Pompeu Fabra*

\* Corresponding author

**Keywords:** Historic maps, Archaeology, Digital workflow, South Asia

## Abstract:

Historical maps depict landscapes as they once were, and include evidence for a wide range of information, including settlement distribution, land use, natural resources, transport networks, toponymy, and other natural and cultural data within an explicitly spatial context. Maps produced before the expansion of large-scale mechanized agriculture reflect a landscape that is lost today, and there is a growing awareness that these historical maps record information about the distribution of archaeological sites, many of which are now no longer visible.

This paper will present the ongoing work of the Mapping Archaeological Heritage in South Asia (MAHSA) project, led by the University of Cambridge. We focus here on some of the problems and potential solutions relating to the use of digitised historic Survey of India maps for archaeological research, and their relationship to the libraries which house them. We explore the issues encountered when working to follow best practice in meeting the dual aims of: a) meeting the archaeological research aim of extracting, analysing, and recording geospatial data from historic maps sheets; and b) producing an accessible, interoperable resource that links across different repositories and media formats. This case study stems from coordination between the MAHSA Project and two UK Legal Deposit libraries. The aim is to obtain information that is archaeologically relevant, and to make digitised and georeferenced Survey of India maps available openly for the wider public and research community, while also serving the cataloguing and research needs of the participating organisations.

Pakistan and western India are extremely rich in archaeological and cultural heritage sites, which span in date from the earliest villages, through several phases of urbanism, the rise and fall of numerous historical states and empires, and up to the colonial and modern periods. Today, many areas are densely occupied and undergoing rapid development, so many sites are at risk, typically from factors including erosion, large-scale development, looting, and the expansion of extensive irrigation agriculture and the concomitant levelling of large tracts of land. Archaeological site destruction has been observed in the field and is ongoing, and the level and rate of site loss is largely unmonitored.

The MAHSA project is developing heritage management tools and expertise that will work towards addressing these threats to archaeological sites and monuments. The project is documenting the endangered archaeology and cultural heritage of the Indus River Basin and the surrounding areas, and will publish this information in an Open Access geospatial database using the Arches online database platform, which uses the CIDOC-Conceptual Resource Model ontology (<http://www.cidoc-crm.org/>). It aspires to become sustainable by being the default management database for local heritage stakeholders through a programme of collaboration and training.

Digitising historic maps is a major component of the MAHSA project workflow, which integrates three approaches to identify archaeological and cultural heritage sites, document them, and then make these data openly available. These are: i) collating and systematising the existing published data on archaeological and cultural heritage sites; ii) using historic maps, remote sensing and automated site detection methods to identify and document previously unidentified archaeological and cultural heritage sites; and iii) collaborating with local stakeholders and offering training in GIS, methods of site detection, recording of sites on the ground, and documentation (Figure 1).

Previous research has shown that the colour maps issued by the Survey of India from 1905 to 1941 document the location of thousands of mound features that are often archaeological sites (Petrie et al. 2019). Generated as a legacy of Britain's colonial past, paper copies of these maps are held in UK Legal Deposit libraries, which house hundreds of these historic map sheets, many of which remain available only in paper form, or as text-catalogued digital images (as opposed to georeferenced map objects). The extent of the overlap between the collections is currently under review, but

together these collections represent a vast analogue dataset that currently exist largely as bibliographic objects with their own historic and visual value. If these maps are fully digitised and georeferenced, however, they have the potential to offer detailed geospatial data about archaeological sites and monuments which predates much of the development and land use changes seen in recent decades. These maps also encode information relevant to a wide range of other disciplines.

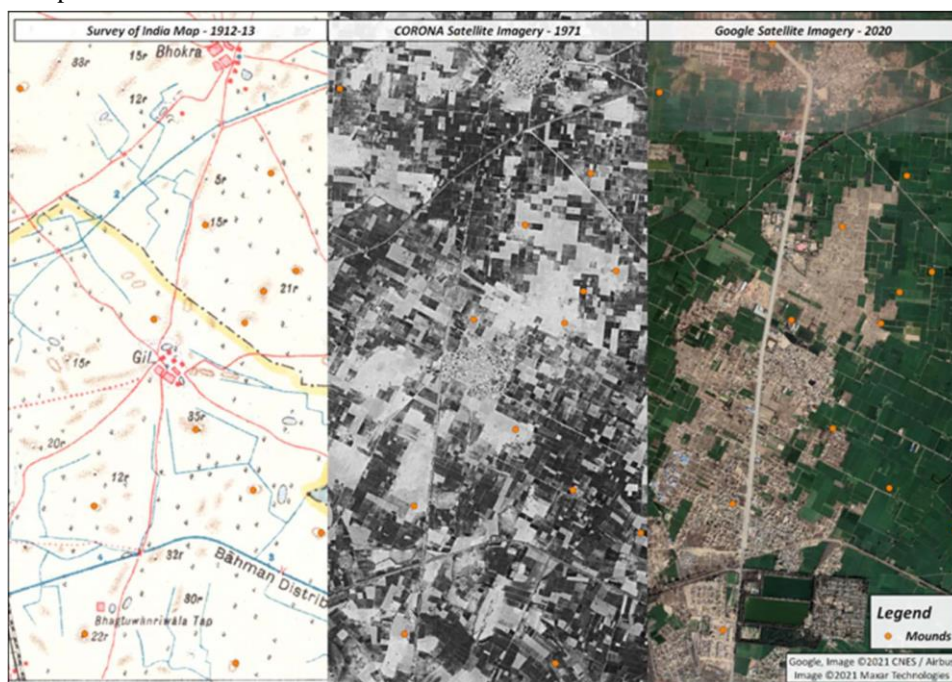


Figure 1: Historic maps as part of the MAHSA-project workflow of site detection and monitoring: digitised map sheet, historic satellite imagery, recent satellite imagery. Showing urban expansion in relation to archaeological mound features.

As part of the workflow, MAHSA is also making use of machine learning approaches to detect specific features and words that are documented on maps that have a specific archaeological relevance (Garcia-Molsosa et al. 2021). Although this represents a very specific application of machine learning, these approaches can be readily adapted to a wide range of other applications.

Once digitised and georeferenced, historic maps exist as a multifaced resource: as images in their own right; as holders of bibliographic and historic information; as signposts to larger datasets; and as containers of geospatial data (in both paper and digital form). With paper copies held, curated, and distributed across nationally important library collections, the digitisation process must therefore serve a wide variety of end users, with the resulting output of digitised maps following best practice in meeting the data standards of all stakeholders, including those in digital spaces (Hyland, Atemezing, and Villazón-Terrazas 2014). As public bodies, UK Legal Deposit libraries are working to make collections increasingly available online for both public consumption and research purposes. The MAHSA project is therefore working to meet these multiple angles of data creation, curation, and dissemination.

One way in which these requirements can be met is to enable digitised historic maps to exist as data objects that adhere to the principles of Linked Open Usable Data (LOUD) in their generation, curation, analysis, and dissemination. This also contributes to their compliance with open access FAIR data principles, aiming to make maps-as-data Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016). The MAHSA project is approaching this problem by working to make georeferenced digitised historic maps available through the International Image Interoperability Framework (IIIF). IIIF facilitates open access to images, by making use of JSON-LD, linked data and standard W3C web protocols (Babcock and Di Cresce 2019). The final step in this workflow is the creation of a geospatial search portal, which is able to point users to the appropriate library holdings. Historic maps will be linked/related as information resources to the individual heritage resources in the MAHSA Arches database, pointing back to library holdings, and enabling historic maps to be found, accessed, and used by a wide variety of users.

We use this case study to examine the implementation of best practices in both workflow and data standardisation when working with multiple stakeholders. We evaluate aspects of effectiveness, efficiency and satisfaction of such requirements, and explore how to feed the lessons learned during so far with our own particular case study back into the improvement cycle for LOUD-compliant data solutions for digitised historic maps.