Archeologia e Calcolatori
36.1, 2025, 469-480
doi 10.19282/ac.36.1.2025.26

BUILDING AN ECOSYSTEM OF DIGITAL RESOURCES ON THE WRITTEN HERITAGE OF ANCIENT ARABIA

1. Introduction

The Arabian Peninsula preserves a vast corpus of pre-Islamic inscriptions in various Semitic languages, written in alphabetic scripts from both the South Semitic (i.e. the local Arabian) and the Phoenico-Aramaic families. In Southwestern Arabia, corresponding to present-day Yemen and neighbouring areas, early kingdoms emerged in the early first millennium BCE. The region's role in the incense production and trade connected it to wider cultural and economic networks. The Ancient South Arabian writing tradition alone accounts for approximately 15,000 inscriptions (Avanzini 2016; Stein 2020). Even more numerous are the inscriptions in the local languages and scripts from Central and North Arabia, with several thousand found in oasis settlements along the trade routes, in addition to tens of thousands of graffiti left by nomads and travellers on the rocks of the vast arid landscapes stretching from northern Yemen to southern Jordan and Syria (Macdonald 2000; al-Jallad 2018).

The Digital Archive for the Study of pre-Islamic Arabian Inscriptions (DASI; https://dasi.cnr.it) currently provides on open access the digital editions of nearly 8800 ancient epigraphic texts from the Arabian Peninsula. Maintained at the CNR-Institute of Heritage Science, DASI is the first of a series of online tools that, since the last decade, have made openly accessible a huge number of annotated textual sources from ancient Arabia (Rossi 2023). Since 2023, in the frame of the Humanities and Cultural Heritage Italian Open Science Cloud project (H2IOSC; www.h2iosc.cnr.it), DASI has been a fundamental case-study within a pilot project of the E-RIHS infrastructure aimed at improving discoverability and accessibility of resources in the domains of digital archaeology and digital epigraphy (H2IOSC WP7.4; Rossi, Salvador 2024). The activities related to DASI case-study aim to optimise the archive as an up-to-date digital tool for the critical edition of epigraphic texts, particularly those belonging to under-resourced textual traditions such as the Semitic corpora attested in pre-Islamic Arabia. This contribution seeks to critically outline the progress made in this area.

2. DASI ecosystem through its history

The initiative of an electronic corpus of the ancient Arabian inscriptions was launched in 1999 at the University of Pisa by Prof. A. Avanzini, focussing on the Corpus of South Arabian Inscriptions (CSAI). Published online

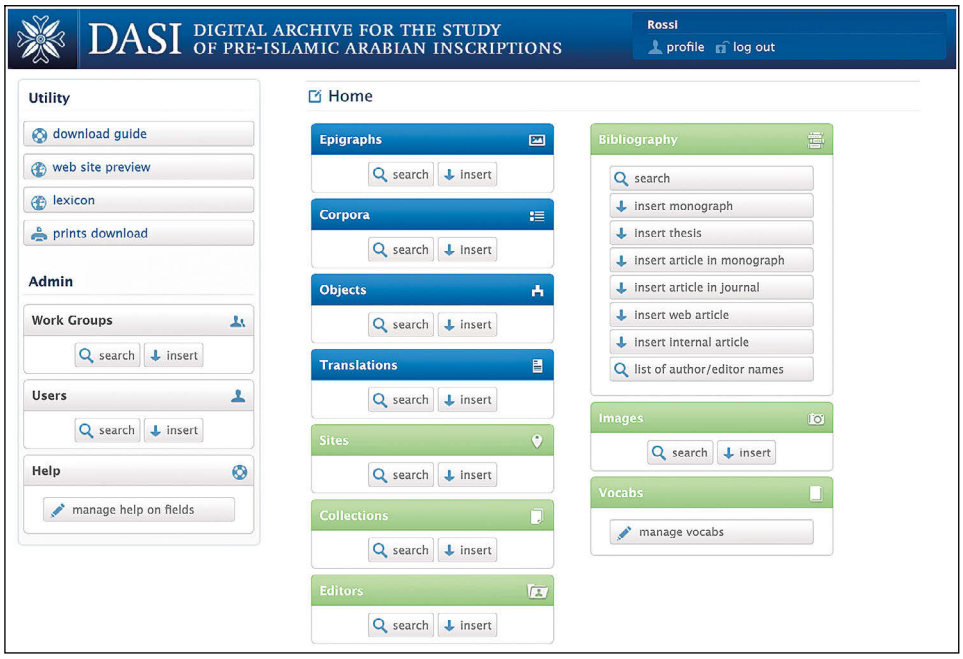

Fig. 1 – DASI-Digital Archive for the Study of pre-Islamic Arabian Inscriptions: homepage of the data-entry interface.

in 2001 as an SGML-based corpus, later XML-based, CSAI was gradually implemented with contents over the years until an ERC Advanced Grant in 2011 (GA 269774) supported the creation of a more complex system, called the Digital Archive for the Study of pre-Islamic Arabian Inscriptions (DASI). The objective of this project was to expand the textual corpus by completing the cataloguing of Ancient South Arabian inscriptions and, where possible, including North Arabian and Aramaic inscriptions produced in the region[1]. Moreover, it aimed at re-engineering the IT system in order to align it with the good practices of digital philology at that time and provide advanced analysis tools.

[1] To this aim, collaborations were established with other centres specializing in the latter textual traditions, which led to the optimisation of some shared vocabularies and the inclusion of some hundreds such inscriptions in DASI. The subsequent funding obtained by these institutions for the development of archives specifically dedicated to the edition of those corpora prompted a halt in the integration of non-Ancient South Arabian texts into DASI, with a view to later fostering interoperability with such initiatives. These are the Online Corpus of the Inscriptions of Ancient North Arabia (OCIANA, https://ociana.osu.edu/) and the Digital Corpus of the Nabataean and Developing Arabic Inscriptions (DiCoNab, https://diconab.huma-num.fr/).

Thanks to the involvement of the IT specialists of the Scuola Normale Superiore di Pisa, a relational database was set up for the management of information on texts, supports, and their provenance, besides the bibliographical and visual apparatus. A web-based user-friendly interface was implemented for data manipulation, ensuring an effective management of collaborative work (Fig. 1). This includes a module for annotating the textual transliteration of the inscriptions according to the TEI-EpiDoc standard. A front-end website was launched in 2013, allowing for indexing, browsing and querying the epigraphic texts and their metadata on open-access. Finally, an OAI-PMH repository was deployed, providing the inscriptions' files according to the TEI-EpiDoc and Dublin Core standards, besides the Europeana Data Model. Before the end of the funded project, a pilot spin-off initiative was launched to create lexica for the Ancient South Arabian languages, which face the typical challenges of under-resourced languages[2].

A few years after the end of the funded project, DASI was migrated to the CNR, which has committed to ensure its technological and scientific maintenance. The number of epigraphic records has been growing, although at different speeds in relation to the availability of fundings for data-entry personnel. Indeed, DASI's objective is not limited to register epigraphic data according to the available editions of the texts, but DASI itself aims at providing the most updated edition at the time of the inscription's inclusion in the archive – potentially improved with respect to existing literature and always updatable. As of now, DASI's solid technological system has efficiently served for twelve years the editorial work of its team, involving more than a hundred collaborators, and has allowed consultation by users worldwide as well as data reuse by aggregators such as Europeana and Trismegistos. At the same time, new needs have emerged, particularly of a scientific nature, demanding the development of new tools.

In 2018 DASI became one of the data providers for the Maparabia project, a collaborative CNRS-CNR initiative funded by the French ANR (ANR-18-CE27-0015), aimed at developing tools for the spatial and semantic integration of the vast body of textual, archaeological and geographical data on ancient Arabia, available in different archives. Besides a tool for exporting DASI data towards the central geodatabase – now browsable through the WebGIS of the Digital Atlas of Ancient Arabia (https://ancientarabia.huma-num.fr/atlas) – the project allowed the creation of the Gazetteer of Ancient Arabia, a geographical thesaurus that reuses DASI data to enrich information about places in the region, with particular attention to toponymic and contextual data (https://ancientarabia.huma-num.fr/gazetteer; De Santis *et al.* 2021).

[2] For an overview of the various aspects of DASI system, see Avanzini *et al.* 2018 and the references therein.

Building on the experience of the Gazetteer, a prosopographical database was recently developed thanks to a grant from the FOE (Fondo ordinario per gli enti e le istituzioni di ricerca, of the Italian Ministry Education, University and Research) as part of an initiative of the Ancient Near East Research Group at the CNR-ISPC. The Prosopography of Ancient Arabia (ProsA) leverages DASI's annotated names of persons, social groups and deities from a historical perspective, through a system that ensures continuous synchronization of the prosopographical database with the textual archive (Rossi, Buono forthcoming).

3. DASI within H2IOSC: data model enrichment

The reuse of textual and contextual data for linguistic and historical research soon highlighted the need of including in DASI categories of written sources that had been excluded from the archive or only sporadically recorded[3]. Some have been traditionally regarded as 'minor' or 'utilitarian' epigraphic documents, such as graffiti and *instrumenta inscripta*; others as only marginally epigraphic, like *legendae* on coins (in the realm of numismatics) and texts on small wooden sticks whose domain intersects manuscript studies (archival documents, correspondence, school exercises, etc.)[4]. In the frame of the H2IOSC project, we carried out a feasibility study to evaluate the extent to which DASI's data model could already accommodate specific information on these materials, determine which components needed to be revised, and identify new features required for their integration. It should be emphasised that the goal is not to create, for instance, a numismatic corpus or an inscribed pottery collection, but rather to facilitate the broader reconstruction of the ancient Arabian civilisations through the integration of the linguistic, iconographic and historical information conveyed by these different types of written sources. Reflection on the cataloguing requirements of such additional materials has also provided an opportunity to enhance the digital representation of the material aspects of written heritage – too often overlooked in philological studies, yet crucial for a holistic understanding of these sources and particularly relevant to the E-RIHS infrastructure (https://www.e-rihs.it/). For this reason, most of the interventions were aimed at improving the description of the textual support and of the visual dimension of writing.

[3] DASI was originally optimized for inscriptions conventionally referred to as 'monumental': mainly produced by scribal schools associated with socio-political or religious authority, they result from a formalized process of textual production, usually involving multiple agents and specialized personnel. They are usually publicly accessible inscriptions, belonging to well-defined textual typologies, such as commemorative texts, building inscriptions, legal records, dedicatory inscriptions, funerary texts.

[4] The category of 'epigraphy' itself is conceptually problematic (see Panciera 2012; Grossi 2016) and proves to be particularly fluid when examined across different textual cultures.

3.1 *Text appearance and execution technique*

Originally, a single attribute was used in DASI to record the writing technique, with a vocabulary that included only the basic types found in South Arabian documentation: 'incised', 'relief', and 'ink'. These correspond to text produced by subtracting material within the shape of the letters (thus

Fig. 2 – Examples of Ancient South Arabian inscriptions' text appearances obtained by different execution techniques: a) cast bronze openwork stamp seal (BM 1993,0126.1); b) agate stamp seal with engraved inscription and its relief impression (BM 120361); c) cast bronze hand with engraved inscription (BM 139443); d) bronze altar with cast inscription in relief (BM 135323); e) stone inscription with engraved texts – one of which results in relief (BM 125130) (photos © The Trustees of the British Museum).

appearing in negative), by removing material around the letters (resulting in a positive, raised appearance), and by adding ink, respectively. Over time, terms such as 'openwork' (e.g., for seals) and 'cast' (for raised letters created through metal casting) were added. With the inclusion of a wider range of inscribed objects, however, these categories were no longer sufficient to account for the full variety of writing techniques. More importantly, it became necessary to distinguish between two separate aspects: the 'execution technique' used to produce the inscription, and the resulting 'text appearance' with respect to the written surface of the support.

The former is more commonly used in epigraphic databases, for which standardized vocabularies have been developed, and is especially relevant for certain types of supports, such as inscribed pottery and graffiti. However, in our case, a mere shift to this vocabulary would not have done justice to the historical relevance that the visual appearance of the text plays in the study of Arabian epigraphy. For instance, relief inscriptions on stone in South Arabia are typical of specific periods and political-cultural contexts: their appearance is an important historical datum. Moreover, it should be underlined that the two types of information do not always correlate directly: the same execution technique can produce different visual results (as is the case with engraving or stamping, which can produce negative/positive results depending on the appearance of the matrix text of the seal) and, conversely, different techniques can lead to similar appearances (as is the case with relief, which can result from both casting and incision) (Fig. 2)[5]. Finally, the indication of the text appearance complements the interpretation of photographic documentation, as determining from photographs whether a text is rendered as positive or negative can be very challenging, hindering the correct identification of letter shapes.

3.2 *Text position*

Another important aspect of the materiality of writing is the 'text position' on the support. This feature is particularly significant for two-faced objects such as coins, where the placement of texts and decorations on the obverse or reverse carries interpretive implications (Fig. 3), and wooden sticks, which often bear separate texts on different sides. It is also relevant for archaeologically significant materials like pottery, where the placement of the inscription on specific parts of the vessel can offer functional indications

[5] Cf. the opening note of the EAGLE's Execution Technique vocabulary: «A major classification problem is faced on this respect by epigraphists. In same (sic!) cases the distinction between the method used to produced (sic!) a text is used as a principle for a definition, in some other the name of the writing technique is based on the result obtained, no matter of the method that was used» (https://www.eagle-network.eu/resources/vocabularies/writing/).

Fig. 3 – Ancient South Arabian silver coin with figurative elements and monograms (BM 2002,0101.1826, photo © The Trustees of the British Museum).

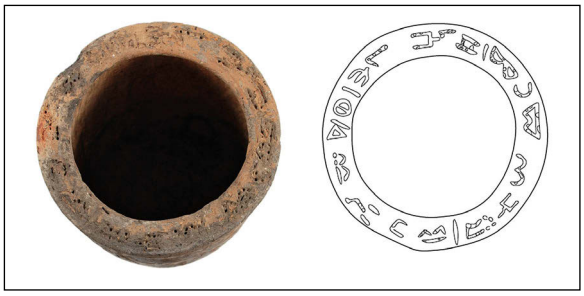

Fig. 4 – Ancient South Arabian vertical inscription running counterclockwise, realised *ante-cocturam* on a jar rim (MAA 1951.589) (facsimile by G. Buono; photo © Museum of Archaeology and Anthropology, University of Cambridge).

and shed light on the object's production and consumption processes (Fig. 4). This indication proved particularly easy to manage thanks to the separation between the entity containing information on the text and the one concerning the support in DASI data model. In the case of a support bearing two distinct texts – for instance, one on the obverse and one on the reverse – a separate record is created for each text, including the corresponding metadata, while both are linked to the same support record. If, on the other hand, a single text extends across different sides of the support, only one epigraphic record is created.

3.3 *Text direction*

Finally, we decided to expand the terms of the vocabulary recording the direction of the text. Until now, this attribute primarily served in DASI to indicate whether a text runs right-to-left (the standard direction in Arabian

scripts), boustrophedon, or left-to-right. However, the inclusion of new materials requires additional terms to account for the various directions a text may follow on different supports. To give a few examples, a text may be written vertically (with one letter placed above the other), especially on narrow surfaces such as pottery or tool handles, or along the rims of vessels, where the circular shape of the support also necessitates specifying a clockwise or counterclockwise direction (Fig. 4). Graffiti often exhibit the most irregular layouts, with mixed, wavy or even spiral formations. To avoid excessive complexity and the proliferation of new attributes, we decided to include all these diverse specifications within the 'text direction' attribute.[6] This also applies to a distinctive form of text production frequently attested in Arabian – particularly South Arabian – epigraphy: monograms. By monograms, we refer to non-linear combinations of letter shapes that convey textual information while also serving a decorative function (Fig. 3). Typically, monograms represent names – of persons, social groups or gods – and may appear either as standalone markers or alongside linear texts. They are formed by combining all or part of a name's letters, and are often challenging to decipher, as a single letter shape (usually geometric) may include, or appear to include, other letter shapes, creating interpretive ambiguity. It is worth noting that DASI already allowed for the documentation of monograms as decorative elements of the support record, with the composing letters or resolved name recorded in a dedicated attribute. As part of the recent upgrade, we have decided to include this important source of onomastic information – when decipherable – transcribed as linear (and thus searchable) text within the epigraphic record, while specifying its nature as a monogram (i.e., non-linear text) under the 'text direction' attribute.

3.4 *Support description*

The four-level hierarchic vocabulary for the description of the type of support proved sufficiently granular to cope with the description of further textual supports like coins, ceramics, sticks, as well as elements of the natural landscape such as the rocks hosting graffiti, by simply expanding the list of controlled terms. However, two additional attributes had to be introduced to accommodate coin-specific information of cultural rather than material

[6] This parameter should not be confused with the orientation of individual letters – a feature not currently recorded in DASI, and which, if needed, should be marked at the level of the individual sign within the textual encoding, rather than as metadata for the entire text. In fact, although letters with a defined orientation (i.e., those not symmetrical along a vertical axis) typically face left when the text direction is right-to-left, and viceversa, this correspondence is often disrupted, such as in graffiti, where – due to the writer's inattention or limited competence – a leftward text may be composed partially or entirely of right-facing letters. Therefore, the two attributes – the orientation of the whole text and that of individual signs – may not coincide.

relevance: 'mint' and 'denomination'. Given DASI's strong focus on the decorative elements of the support, which provide essential visual context for a holistic understanding of the epigraphs, the addition of a repeatable attribute to describe the 'position of each decoration' on the object has significantly enhanced the system's descriptive capacity, mirroring the attribute for text position in the epigraph record (Fig. 3).

3.5 *Chronological and spatial information*

The enhancement of contextual information also concerned the refinement of the system's proficiency in capturing chronological data, previously limited to the periodisation of South Arabian history, indicated by conventional labels such as A, B, C, etc., along with any precise dates mentioned in the text and referring to one of the ancient local systems of relative chronology. The desire to facilitate searchability of these historical data, to make them more accessible to non-specialist users, and to potentially integrate them into a broader network of digital historical resources, required the inclusion of attributes for recording 'time spans' in terms of absolute chronology, as well as specifying the 'date accuracy'.

As regards spatial data, considerable effort had already been made in the past to improve a comprehensive description of the geographical, archaeological and historical features of each site, while also covering the history of research. Ancient toponymy became a particular focus for data reuse in the framework of the Gazetteer of the Maparabia project. In this regard, the only addition we deemed important to introduce in the latest version of DASI is the site's current name transcribed in the local script, rather than in Roman transliteration alone, as had previously been the case, thus supporting multilingual access and interoperability.

4. Perspectives: towards an ecosystem of FAIR data

The main goal of the activities described here is to serve the disciplinary community that uses DASI for research in the history and languages of the Ancient Near East. The new attributes will be populated systematically starting from the newly recorded texts. For the existing records, the (mostly manual) revision and integration process will require a long time. A new version of the frontend website, aligning with the above-mentioned updates to the data model and other technological upgrades, will be soon released and will progressively incorporate the aforementioned data revisions; priority will be given to the attributes for the 'text appearance/execution technique' and 'text direction' as regards monograms.

As in most comparable projects, producing a curated epigraphic edition in DASI with the required updating is a demanding and time-consuming task.

Precisely for this reason, it is essential to ensure the long-term accessibility and maximum reusability of the data produced. In this context, a process of data FAIRification is underway in the frame of H2IOSC activities. As a first objective, we chose to focus on the extensive indexed bibliography related to the Ancient South Arabian corpus, which includes around 1800 records cited in the editions of inscriptions, translations, objects, sites, and images. We decided to rely on a versatile and widely used tool within the scholarly community for managing and sharing bibliographic data: Zotero (https://www.zotero.org/). Initially, the data stored in DASI was mapped to the BibTeX standard using a Python script. This was possible because DASI's bibliographic information was already organized in dedicated records, structured according to the type of publication (journal article, book chapter, monograph, etc.), with data separated into specific fields (e.g., title, author, date, publisher). The resulting 1800 entries were then batch-imported into a dedicated 'CSAI' Zotero group (https://www.zotero.org/groups/5280688/). Afterwards, the entries were carefully revised in accordance with FAIR practices, which implied adding unique identifiers where available (ISBNs, DOIs, URLs). The Zotero group was made public at the end of 2024, and a PDF version of the dataset was subsequently created using Markdown language and Pandoc, and deposited on Zenodo (Rossi *et al.* 2024).

Ongoing work is focusing on other DASI's resource types. Outcomes such as richer descriptive metadata of FAIRified epigraphic records as digital scholarly editions will be made visible in the updated frontend, while the new API endpoint will enable data interoperability. These efforts aim not only to consolidate DASI's role as a disciplinary tool and digital publication, but also to facilitate reuse of its data across the broader communities of epigraphists, linguists, historians of the ancient world, and cultural heritage specialists. Future activities will focus on the application of semantic technologies across DASI ecosystem to enable deeper engagement with these wider communities. A crucial step will be to assess how far these approaches can be meaningfully applied across epigraphic datasets that deal with diverse cultural, linguistic, and scribal traditions, especially in the domain of less-resourced disciplines.

Irene Rossi
Istituto di Scienze del Patrimonio Culturale - CNR
irene.rossi@cnr.it

Acknowledgements

DASI is the result of the collaboration of several individuals and institutions, whose contributions to both the technological and content development of the archive deserve acknowledgment (cf. the Credits page in DASI website). I am deeply grateful to Chiara Salvador (H2IOSC, CNR-ISPC) and Giulia Buono (CNR-ISPC) for their input on the update of DASI data model; to IT specialist Matteo Gallo, who

implemented all technological developments; to Erica Scarpa and Riccardo Valente (H2IOSC, CNR-ISPC) for their support through DASI's FAIRification process. This research was supported by H2IOSC Project - Humanities and cultural Heritage Italian Open Science Cloud funded by the European Union NextGenerationEU – National Recovery and Resilience Plan (NRRP) - Mission 4 'Education and Research' Component 2 'From research to business' Investment 3.1 'Fund for the realization of an integrated system of research and innovation infrastructures' Action 3.1.1 'Creation of new research infrastructures strengthening of existing ones and their networking for Scientific Excellence under Horizon Europe' - Project code IR0000029 - CUP B63C22000730005. Implementing Entity CNR.

REFERENCES

Avanzini A. 2016, *By Land and by Sea. A History of South Arabia before Islam Recounted from Inscriptions*, Arabia antica 10, Roma, L'Erma di Bretschneider.

Avanzini A., De Santis A., Rossi I. 2018, *Encoding, interoperability, lexicography: Digital epigraphy through the lens of DASI experience*, in A. De Santis, I. Rossi (eds.), *Crossing Experiences in Digital Epigraphy: From Practice to Discipline*, Warsaw/Berlin, De Gruyter Open Poland, 1-18 (https://doi.org/10.1515/9783110607208-002).

De Santis A., Gallo M., Rossi I., Schiettecatte J. 2021, *The digital Gazetteer of Ancient Arabia: An example of reuse and exploitation of annotated textual corpora*, «Umanistica Digitale», 11, 125-143 (https://doi.org/10.6092/issn.2532-8816/13681).

Grossi M. 2016, ΕΓΡΑΨΕΝ ΔΕ ΚΑΙ ΤΙΤΛΟΝ Ο ΠΙΛΆΤΟΣ *(Gv 19,19). Verso una nuova definizione di iscrizione*, «Zeitschrift für Papyrologie und Epigraphik», 197, 85-95 (https://www.jstor.org/stable/43909980).

al-Jallad A. 2018, *What is Ancient North Arabian?*, in N. Pat-El, D. Birnstiel (eds.), *Re-engaging Comparative Semitic and Arabic Studies*, Abhandlunden für die Kunde des Morgenlandes Band 115, Wiesbaden, Harrassowitz Verlag, 1-43.

Macdonald M.C.A. 2000, *Reflections on the linguistic map of pre-Islamic Arabia*, «Arabian Archaeology and Epigraphy», 11, 1, 28-79 (https://doi.org/10.1111/j.1600-0471.2000.aae110106.x).

Panciera S. 2012, *What is an inscription? Problems of definition and identity of an historical source*, «Zeitschrift für Papyrologie und Epigraphik», 183, 1-10 (https://www.jstor.org/stable/23849859).

Rossi I. 2023, *Epigraphic heritage and digital practices: The case of pre-Islamic Arabia*, in H. Hayajneh (ed.), *Cultural Heritage. At the Intersection of the Humanities and the Sciences. Proceedings of the International Humboldt-Kolleg (Jordan 2019)*, LIT Verlag, 509-526.

Rossi I. 2024, with the collaboration of C. Salvador, E. Scarpa, R. Valente, *Corpus of South Arabian Inscriptions bibliography*, Zenodo (https://dx.doi.org/10.5281/zenodo.14274417).

Rossi I., Buono G. forthcoming, *Studying the ancient Arabian society through a digital prosopography approach: Reuse and enrichment of DASI onomastic data in the ProsA database*, in S. Alaura, M. Bonechi, F. Manuelli, T. Pedrazzi, I. Rossi (eds.), *Between Tradition and Innovation. Interdisciplinary Approaches to Ancient Near Eastern Studies*, Documenta Asiana XIII, Roma, Quasar.

Rossi I., Salvador C. 2024, *An observatory of epigraphic resources on the web: The Open Digital Epigraphy Hub*, «Archeologia e Calcolatori», 35.2, 503-523 (https://doi.org/10.19282/ac.35.2.2024.51).

Stein P. 2020, *Ancient South Arabian*, in R. Hasselbach-Andee (ed.), *A Companion to Ancient Near Eastern Languages*, John Wiley & Sons, 337-353 (https://doi.org/10.1002/9781119193814.ch18).

ABSTRACT

The Digital Archive for the Study of pre-Islamic Arabian Inscriptions (DASI, https://dasi.cnr.it/) currently provides open access to the digital editions of nearly 8800 ancient epigraphic texts from the Arabian Peninsula. After presenting an outline of DASI ecosystem through its 25-year history, this paper focuses on the recent enrichment of its data model, carried out within a pilot project of the E-RIHS infrastructure under the H2IOSC programme. The aim was to optimise DASI as an up-to-date tool for the digital critical edition of a broad spectrum of epigraphic sources from ancient Arabia, including graffiti, *instrumenta inscripta*, coins, and inscribed sticks, alongside 'monumental' inscriptions. Most of the interventions targeted the description of the visual aspect of writing and related contextual information, enhancing the digital representation of the material dimension of written heritage, which is often overlooked in philological studies. Ongoing work is targeting the FAIRification of DASI data, which has so far resulted in the sharing of an extensive bibliography of 1800 records through Zotero.