

# Introduction

Epigraphy is a multifaceted discipline. Even more than in manuscript studies or papyrology, a researcher approaching an epigraph should be competent with philology, linguistics, archaeology, history of art, not to speak of history tout-court, being inscriptions studied first of all as primary historical sources. The peculiar nature of the epigraphic document – both textual and physical – has put the reflection on digitization of epigraphs at the crossroads of the discussions and advancements in digital humanities and digital heritage, in addition to computational linguistics.

The digitization of the epigraphic heritage is at an advanced stage. A significant number of projects digitizing inscriptions, of both small and big corpora, with different objectives are either under development, or have been recently completed. Many papers have been written, and several proceedings of meetings and conferences dedicated to this topic have been published.

However, digital epigraphy is not yet considered a proper discipline. Digital epigraphers have acquired their skills in digitization methods and techniques informally, “in the field”, through a progressive refinement of those established in the digital humanities. Scholars interested in digital epigraphy are creating more or less formal networks in order to exchange ideas and suggestions, even in very different historical and geographical domains. Nevertheless, there are still no regular occasions to meet and discuss.

Moreover, this large and across-the-board community does not recognize itself in specific journals. They continue to communicate the results of their scientific and technical activities in journals dealing with traditional epigraphy, or, at best, digital humanities in general.

This book is precisely intended to stimulate debate among those practicing digital epigraphy, by recording the methodological issues they have addressed while carrying out specific projects, the solutions they have applied and the criteria that have led to their choices.

In particular, whereas a consistent number of digital initiatives in the domain of Classical epigraphy have been well represented in the proceedings of conferences organized within the frame of the project EAGLE,<sup>1</sup> other domains – and that of Semitic epigraphy *in primis* – are in a quite different situation. Barriers due to the extreme wealth, and also diversity, of writing systems and languages, and to cultural and historical fragmentation, make confrontation and cooperation difficult.

For this reason, the projects represented in the nineteen contributions collected in this book are intentionally diverse in geographic and chronological context, for script and language, and typology of digital output.

---

<sup>1</sup> See further on in the volume (in particular the contributions by Liuzzo) for detailed bibliography.

## The Experience of DASI Project

The idea of a volume collecting different experiences of projects on digital epigraphy has arisen within the frame of DASI – *Digital Archive for the Study of pre-Islamic Arabian Inscriptions*, an ERC – Advanced Grant funded project led by Prof. Alessandra Avanzini at the University of Pisa, aimed at gathering, in an open-access archive, the curated edition of the epigraphic corpora of pre-Islamic Arabia. These consist of thousands of Ancient South Arabian, Ancient North Arabian and Aramaic inscriptions produced since the beginning of the first millennium BCE until the advent of Islam. The study of these inscriptions is essential in order to fill a significant gap in research on the ancient and late antique Near East.

During the five years of the project, a team (consisting of epigraphers, archaeologists, art-historians, digital humanists and IT specialists) worked together, facing methodological and technological challenges while building upon previous experiences of digitization of inscriptional corpora in Semitic languages and alphabetic scripts.

Basic, common issues concerned the modelling of data in order to best describe the complex nature of the epigraphic source, and the encoding of text for its critical edition. Fundamental issues such as those of compliance to standards, interoperability and data openness were tackled. Moreover, specific methodological and technical challenges were faced when approaching the study of under-resourced languages, such as those of pre-Islamic Arabia, which are documented only by epigraphic sources. Specific, lexicographic tools were designed to enhance the description of the language and thereby reach a better comprehension of the messages conveyed by the inscriptions – ultimately leading to the best possible understanding and dissemination of the history and culture of the peoples inhabiting Arabia in pre-Islamic times.

The DASI project has attempted to make the tradition of studies related to pre-Islamic Arabia less “marginal” than before, making the edition of about 10,000 inscriptions originating from ancient Arabia openly available. It has tried to provide useful tools and suggest new approaches to the study of this rich cultural heritage, and to foster reasoning on best practice by taking account of domain-specific questions. This has led to a constant search for confrontation with other digital epigraphy projects.

This volume, conceived during the post-grant phase of the project, continues the mentioned practice of confrontation, wishing to raise new questions and open further, unexpected research perspectives.

## Concept and Content of the Volume

With this vision in mind, this book gives voice to those who have conceived and carried out diverse projects, ranging: from antiquity to medieval and modern times; from alphabetic to logographic writing systems; from Indo-European to Chamito-Semitic to Ancient American languages; from specific databases and lexica, to aggregators, infrastructures and gazetteers.

Hereafter, summaries of the main characteristics of each project and the topics of the related papers are provided in order to facilitate the readers' orientation.

Chapter 1, by Avanzini, De Santis and Rossi, describes the project DASI – *Digital Archive for the Study of pre-Islamic Arabian Inscriptions*, focusing on the main digital epigraphy themes discussed throughout this volume: text encoding and data modelling, interoperability, and lexicography.

The project RuneS – *Runic writing in the Germanic languages* (Chapter 2) collects texts in different Germanic languages and using different Runic writing systems. This comparative approach to the study of the script has led, as explained in the contribution by Zimmermann, Kazzazi and Bahr, to transcend the existent descriptive systems and enhance the visual documentation of inscriptions, through the tagging of images.

Similarly, *Hesperia – Banco de datos de lenguas paleohispánicas* gathers inscriptions and coins in the different Palaeohispanic languages, written in multiple writing systems. The solutions adopted to register and make searchable both script variants and the different transliterations used in the study tradition, are described by Estarán, Beltrán, Orduña and Gorrochategui in Chapter 3.

The two projects *Sinlequiunnini* (Di Filippo) and *Text Database and Dictionary of Classic Mayan* (Prager, Grube, Brodhun, Diederichs, Diehr, Gronemeyer and Wagner) propose different solutions in the textual data modelling in relation to logo-syllabic writing systems, in particular dealing with languages whose interpretation is highly context-driven, in the first case (Chapter 4), and with a still partially deciphered script, in the second one (Chapter 5).

The *Beta Maṣāḥaft* project (Chapter 6) deals with Ethiopian and Eritrean inscriptions and manuscripts. Bausi and Liuzzo address the issue of encoding in XML the relation among multiple copies of the same epigraphic text in a multilingual context, and of annotating their different scripts.

The CIP – *Corpus Inscriptionum Phoenicarum necnon Poenicarum* (Chapter 7) is the first attempt at carrying out a census of the Phoenician and Punic inscriptions spread in a very wide territory, from the Eastern to the Western Mediterranean. The contribution by Xella and Zamora provides an overview of the criteria they have followed to create a complete edition of the only direct textual sources for the reconstruction of the history and culture of this civilization, in the current absence of any attestation of literary texts.

The OCIANA – *Online Corpus of the Inscriptions of Ancient North Arabia* project (Burt, al-Jallad and Macdonald) is a database mainly designed to catalogue graffiti.

Their curated editions, including transcriptions, transliterations in Latin characters and translations, include encoding with particular attention to grammatical analysis and onomastics (Chapter 8).

As the mentioned projects show, the digitization of the overall epigraphic heritage is often aimed at supporting linguistic study. The *Sabaic Dictionary Online* aims at cataloguing all extant lexical material of one of the Ancient South Arabian languages (Chapter 9). Multhoff provides a sound explanation of the methodological issues concerning the annotation of morphological analysis: treatment of ambiguous forms, homographs, heterographs with identical meaning, variant readings, incorrect forms.

The lemmatizer for the Ancient South Arabian languages, KALAM, performs the automatic detection of morphological attributes (Chapter 10). Ruzicka describes its principles and functioning. The contribution must be considered within the frame of the application of NLP techniques to ancient, under-resourced languages.

The OIMEA – *Official Inscriptions of the Middle East in Antiquity* project (Novotny and Radner) edits all the official inscriptions of ancient Middle Eastern polities in cuneiform script. Texts are geo-referenced and fully lemmatized: lexical and grammatical tagging is carried out in order to create glossaries and allow search of text and translation. Historical research is enhanced by the creation of a map-based interface to access geographical information mentioned in cuneiform sources (Chapter 11).

The project *Karnak* (Biston-Moulin and Thiers) focuses on the epigraphs located *in situ* in the ancient Egyptian temples of Karnak. Therefore particular attention is devoted to the preservation of the relation between the inscriptions and their architectural position. An extensive photographic coverage provides high-resolution orthophotographs flanking the transliterations of hieroglyphic, hieratic and demotic texts. These are the basis for a digital lexicon of the languages documented in the temples (Chapter 12).

The infrastructure of the HPM – *Hethitologie-Portal Mainz* (Chapter 13) provides maintenance and access to several independent digital resources available on Hittitology studies. Müller and Schwemer recall the history of a long-lasting project; the continuous technical updates that have been necessary over time; the specific policies for the attribution of resources, their versioning and intellectual property.

Other projects cope with the establishment of systems to identify, sort and connect digital resources. The interdependence of geographic and chronological entities and their labelling, and the need for ontologies with the objective of structuring this information is exemplified by the project EDV – *Epigraphic Database Vernacular* (Cannata), which collects the vernacular inscriptions produced in Italy from late Medieval to Early Modern Age (Chapter 14).

The *Trismegistos* project (Depauw) aims at implementing an identification system, which attributes an ID to each known ancient inscription. This is a first step to tackle the issue of disambiguating and connecting several editions for the same inscriptions in a LOD environment (Chapter 15).

The objective of the project *PeriodO* (Rabinowitz, Shaw and Golden) is the creation of a Linked Data gazetteer of structured period definitions, which provides links between time periods and geographic and cultural contexts, and translation between absolute dates and relative chronologies. Once applied to digital epigraphy, it will foster interoperability of epigraphic collections and their connection with archaeological datasets (Chapter 16).

Interoperability is fully achieved by the aggregator EAGLE, which collects Greek and Latin epigraphs from many different repositories and makes them available to Europeana. The contribution by Liuzzo focuses on the challenges faced, during and after the end of the project, from the up-conversion to the EAGLE schema of the epigraphic records to the harmonization of the terminologies involved (Chapter 17).

Finally, the EPIDAT – *Database of Jewish Epigraphy* project (Kollatz; Chapter 18), which provides its records to national and European aggregators not specifically focusing on digital epigraphy, and the I.Sicily – *Inscriptions of Sicily* project (Prag and Chartrand; Chapter 19), which, in addition to a consistent amount of previously undigitized epigraphs, provides original editions based on the principles of reuse, linked data and collaboration, demonstrate the potential of records encoded according to the best practice shared by the scientific community.

The volume is provided with an index, listing terms grouped by: Ancient and Modern Regions and States; Languages and Scripts; Concepts of the epigraphic discipline and related digital practice. Finally, two appendices complement the volume. Appendix A presents an annotated webliography of selected online electronic resources cited in the volume, described according to the Dublin Core Metadata Element Set (Version 1.1). Appendix B is intended for disambiguation and definition of selected concepts from the Index of Concepts, by mapping them to the Library of Congress Subject Headings and the Getty Art and Architecture Thesaurus.

## Reading Path

The deliberate heterogeneity of subjects, focuses and approaches to digital epigraphy represented in this volume, allows a non-sequential fruition of the contributions. However, they are grouped into two main subject areas. These areas, which have been part of the research of DASI itself, enclose, in our opinion, the main issues that digital epigraphy should address in developing a methodology able to provide the validity criteria proper to a discipline.

1. The first part of the volume is focused on data modelling and encoding, which deeply influence the possibility to perform searches on texts including *lacunae* and variants.
  - Various scripts, belonging to different writing systems and often not completely deciphered, pose fundamental issues in relation to data modelling and/or encoding, given the high uncertainty in the attribution of

- phonetic, morphological and semantic values to graphemes and sequences of graphemes.
- Data modelling and encoding are also influenced by the will of creating proper critical editions of epigraphs and the specific functionalities required to meet their criteria.
  - Moreover, different languages, often extinct and not completely understood in their morphology and lexicon, need to be studied from the linguistic point of view, before historical, cultural, sociological and much more interpretation can be derived. Lexica and tools for morphological analysis, specifically developed on the basis of the epigraphic collections digitized, and coping with fragmentarily attested languages, are therefore described.
2. Interoperability and aggregation are fundamental to relate data that would otherwise remain separate, in contrast to the reality they refer to. This second part of the volume is dedicated to the initiatives aimed at fostering aggregation, dissemination and reuse of epigraphic materials. It includes:
- the experiences which point out the need, and tools, for interoperability
  - portals providing “annotated” access to several digitization projects, and proper aggregators
  - and projects which, thanks to interoperability, are clear examples of successful dissemination of inscriptions digitized in different projects.

Although the contributions allow multiple keys to interpretation, and the editors encourage a “personal” fruition, this ordering of the papers aims at suggesting a reading path. This path follows the red thread of the dialectical relationship between the need to represent in the digital environment the features of peculiar epigraphic materials in the most effective way, and the need for strategies to share, disseminate, and make data reusable. In other words, the relationship between the compliance with the theoretic tools and the methodologies developed by each different tradition of studies, and, on the other side, the necessity of adopting a common framework in order to produce commensurable and shareable results in digital epigraphy.

In sum, by crossing a wide, even though not exhaustive, range of experiences, this volume intends to point out the methodological issues which are specific to the application of information technologies to epigraphy. It was not conceived to be a prescriptive work; it does not provide answers, but focuses on problems. Eventually, it aims at stimulating interest and discussion around the challenges that the use of IT has been imposing on epigraphy and on how the digital approach is reshaping the very discipline.

## Acknowledgements

The dedication we have put in the preparation of this volume is our personal homage to the Director of DASI project, Prof. Alessandra Avanzini. Under her guidance, DASI has been for us a time of absorbing study, passionate debate, curious experimentation, continuous rethinking and multidisciplinary encounters with the many people of the DASI research groups of the Scuola Normale Superiore and the University of Pisa, who have shared with us daily enthusiasm and hard work.

We take this opportunity to thank the contributors to this volume, who have dedicated, in spite of their many commitments, time and energies to reflect on common issues and challenges.

Our thanks go also to the editors of De Gruyter, and in particular to Katarzyna Michalak, for their constant assistance.

The FP7 post-grant Open Access Pilot project has provided financial support for the publication of the present volume.

Annamaria De Santis and Irene Rossi