

Information Technologies for Epigraphy and Cultural Heritage

Proceedings of the First EAGLE International
Conference



europæana
eagle project

Collana Convegni 26

STUDI UMANISTICI
Serie Antichistica

Information Technologies for Epigraphy and Cultural Heritage

Proceedings
of the
First EAGLE International Conference

a cura di

*Silvia Orlandi, Raffaella Santucci,
Vittore Casarosa, Pietro Maria Liuzzo*



SAPIENZA
UNIVERSITÀ EDITRICE

2014

Copyright © 2014

Sapienza Università Editrice

Piazzale Aldo Moro 5 – 00185 Roma

www.editricesapienza.it

editrice.sapienza@uniroma1.it

Iscrizione Registro Operatori Comunicazione n. 11420

ISBN 978-88-98533-42-8

DOI 10.13133/978-88-98533-42-8



Quest'opera è distribuita con licenza Creative Commons 3.0
diffusa in modalità *open access*.

Distribuita su piattaforma digitale da:

digilab

Centro interdipartimentale di ricerca e servizi
Settore Publishing Digitale

In copertina: Logo EAGLE *The Europeana network of Ancient Greek and Latin Epigraphy*.
Divieto di riproduzione.

*In den Flüssen nördlich
der Zukunft / werf ich das
Netz aus, das du / zögernd
beschwerst / mit von Steinen
geschriebenen / Schatten.
(Paul Celan)*

Foreward

Information Technologies have made possible many important changes in the field of cultural heritage and continue to provide dynamic and exciting media platforms through which new possibilities perpetually emerge. This wave of change has had particularly significant consequences in the field of epigraphy, where a vast array of possibilities for digital content fruition continues to reveal itself, constantly opening doors to new and as-yet-unexplored synergies. Many technological developments concerning digital libraries, media entertainment and education are now fully developed and ready to be exported, applied, utilised and cultivated by the public.

EAGLE is a best practice network co-funded through the ICT Policy Support Programme of the European Commission. EAGLE's goal is to enable digital access to epigraphic resources, while providing guidelines and using metadata standards for searching and browsing.

By creating a seamless and centralised online database, EAGLE is providing access to the epigraphic collections and archives of its project partners, amongst which are many of the leading institutions in the field. An ever-growing part of these resources is becoming accessible through a common, multilingual, easy-to-use portal.

The EAGLE metadata will become part of Europeana, the European portal of millions of digitized items from museums, libraries, archives and multimedia collections.

The EAGLE 2014 conference aims to function as a forum in which progress-oriented individuals and institutions find a place to collaborate and present results. It also aims to provide an overview of the state of the art for epigraphic digital collections within the framework

of the following best-practice themes: digital library tools, education and research facilities, IPR issues, cultural heritage and technologies.

The event featured a variety of workshops, sessions and panels that conform to our standard of quality. We hosted a keynote speaker lineup consisting of two of the most salient voices in the field, both invited to join us from overseas. Susan Hazan is Curator of New Media and Head of the Internet Office at the Israel Museum (Jerusalem). She is a champion of Open Access and had blazed new paths for many national and local museums, while simultaneously leading the User Engagement Group for Europeana Foundation.¹ Tom Elliott is Associate Director for Digital Programs and Senior Research Scholar at the Institute for the Study of the Ancient World at New York University. He is a founding member of the EpiDoc consortium, a pivotal group for the revolution in digital epigraphy that has taken place in recent decades.²

An international committee [p.506] evaluated with a double peer review the proposals gathered from the event's Call for Papers. EAGLE 2014 features more than 30 presentations delivered by specialists from several European and overseas countries.³

Our work on EAGLE has shown us that an interdisciplinary approach to the application of IT to epigraphy is a challenge that must be addressed. It is difficult to meaningfully study all the ramifications of the sometimes troublesome marriage of IT with classical studies. On the one hand, the risk is that humanities scholars may ignore, overlook or oversimplify technical issues; on the other hand, IT people are not necessarily aware of the problems and the needs that are unique to epigraphy and classical studies.

As "a digital bridge to the ancient world", the present volume (based on proceedings from the First International conference of the EAGLE BPN in Paris, October 2014) gathers a very wide range of projects in the field of digital epigraphy. Many *epigraphies* of different periods and contexts are represented and not only in Greek and Latin. Many

¹ The keynote address presentation can be seen here: <http://www.imj.org.il/eagle2014/index.html>

² The transcript of his keynote address can be found in *horothesia* (<http://nblo.gs/10sghm>).

³ See C for details on each contributor's affiliation, and the network page on the EAGLE website <http://www.eagle-network.eu/about/partners/>.

tools are presented to improve research possibilities in several areas of interest; interaction and exploration of new possibilities is in many ways the rule of these proceedings.

This conference and this volume bear witness to a sort of revolution in its young age; it immortalizes a moment in which aims become clear, enthusiasm is still high, and the full range of possibilities has not yet revealed itself. Although our presentation of these proceedings is organized according to thematic blocks, it is not the only path the reader might take through the various contributions.

In part I, *Epigraphic Data: Model, Vocabularies and Interactions*, the reader will find contributions related to technical data management for epigraphs; their harmonization, modeling, search and research. Chapter 1 describes the efforts needed to map different epigraphic databases into a single model undertaken for the EAGLE project itself. Chapter 2 looks instead at ways to interrogate the data to make the search results interesting for a researcher, starting from the experience of Epigraphic Database Bari. These two chapters deal thus with the two fundamentals of data, encoding and querying to make it usable for research.

Chapter 3 deals with the third most issue facing epigraphic data harmonization: the establishment of controlled vocabularies of terms. The perspective given in this chapter come from a non-strictly- Epigraphic project, LIMC-France's Thea. This initiative shares the challenges and objectives with projects like EAGLE. Chapter 4 uses the DASI project to demonstrate the interactive use of properly articulated vocabularies in a modern digital edition of an epigraphic corpus. This chapter also evinces that the powerful research opportunities offered by digital approaches are not only for Greek and Latin epigraphy. When the variables at play are understood and the community of participants is open, the possibilities are vast.

The achievements presented in chapter 5 are of no less importance to understand promise of interactions between traditional epigraphy and modern technologies. The Domitilla Project, already well known in the field, has in fact brought back to life in a spectacular 3D mapping an entire catacomb, allowing (in cooperation with EDB), us to put texts back in their context. Achievements realized thanks to interdisciplinary cooperation are epitomized in chapter 6, which examines the study of Visual Recognition processes carried out by CNR-ISTI. This exciting

work allows machines to help us in research tasks such as the recognition and identification of inscriptions, while providing an extremely useful service to any end user in the field.

The last two contributions of this section focus specifically on the text of inscriptions. Chapter 7 shows how readings can be radically improved through the application of the Morphological Residual Model. Chapter 8 looks at a similar possibility; it looks not to the help of sophisticated algorithms, but a simple image editing tool applied to an inscription from Luna.

This section of the conference proceeding gives an overarching view on developments in epigraphic study that digital tools can bring to the researcher at the levels of data, functions, content enrichment, project structure, contextualization, ease of use and recognition, and rediscovery of texts.

The need for translations of epigraphic documents both for didactic and research purposes as well as for user engagement has become patent in the latest years. A series of extremely interesting projects has taken broached this issue and many questions (and answers) have emerged. The II part of these conference proceedings, *Translating Epigraphy: Challenges and Research Outcome*, looks at these emerging problems and at the pilot projects dealing with translations of inscriptions. Traditionally translations have been underestimated in their scientific and euristic value, but today the importance of this task for research is becoming clear, and not just because of the lower level of knowledge of the languages; on the contrary, it is because of the high level of interpretation involved and for the relevance this reality has to our understanding of the ancient world.

The four contributions in the following section provide a first theorization for this sector of research and scholarly activity. Chapter 9 presents the results of the ongoing online project to publish English translations of Attic Inscriptions. It is the only project currently dedicated specifically to the online publication of a major regional corpus of translated inscriptions for both research and teaching purposes. Chapter 10 introduces a first practical and pragmatic theorization of the translation work, presenting problems and some guidelines for this task on the basis of the inscriptions of Lepcis Magna and Roman Tripolitania. Chapter 11 explores cases and considers the development of translation activity for epigraphists, complementing the propositions

of the previous chapter with attention to specific cases as specific lexicon and elliptical forms of expression.

The section is then closed with the presentation in 12 partnership with Wikimedia Italia and the Perseids project. This collaboration seeks to bring together existing translations with Wikibase so that translation studies may be carried out more effectively. This contribution segues nicely into the following section, as it deals with the users of this content.

Part III *Users, Epigraphy and the social web* deals substantially with user engagement. Against the common intuition that epigraphy is not a topic that can engage anyone besides researchers and specialised amateurs, the reader will find projects focusing on specific classes, museums and people; on many types of epigraphy, designed for all levels of education. In chapter 13 we look at epigraphy in a primary school in Slovenia. Chapter 14 looks at secondary schools in the British system within the Ashmolean Latin Inscriptions' project, while chapter 15 presents the results of a study conducted with university students in Romania. The picture is then completed in chapter 16 by a large survey directed to those who are considered the main users of digital epigraphy. If this set of contributions looks at the world of education and research, the following moves the focus to cultural institutions, especially museums. Chapter 17 explores the objectives and results of #svegliamuseo in bringing the use of social networks to cultural institutions. Chapter 18 reports the accomplishments of #digitalinvasions, a project that began with users' promotion of free knowledge⁴ via social media. Chapter 19 presents the ArcheoWiki project and the many activities that it comprises. This initiative reaches out users, institutions and policymakers to advocate for free knowledge and its dissemination to the public. In the last paper in this section, chapter 20, attention is given to the procedure adopted by the British School at Rome for the resolution of intellectual property right issues.

Having thus covered some of the various aspects of a digital epigraphy project, part IV, *Digital Approaches to Cross-disciplinary Studies of Inscriptions*, looks more directly at some digital epigraphy projects wherein interactions, methodologies and tools are exploited to bring research into this new era, despite skepticism and traditionalism. Among

⁴ Broadly intended as usable and freely accessible data about cultural heritage for machine and human readers.

these projects, special attention is given to the field of 3D modeling for archaeology (already seen in some of the previous contributions). This is the latest horizon of research interactions and thus deserves more careful scrutiny in the arch of our project. Chapter 21 introduces a project set on digitizing the Marmor Parium, thus bridging to the field of ancient Greek historiography. Chapter 22 problematizes a key point of the digital description of inscriptions: the need to harmonize the description of the text with the description of the text-bearing object (or monument). It presents a theoretical paradigm that throws light on many problems found also in other contributions to this volume. Chapter 23, presents activities and ideas behind the Inscriptions of Greek Cyrenaica project. It gives an example of how new methods bring new questions and approaches to the discipline of epigraphy, while keeping attention on long-term continuity in the needs of codification. Chapter 24 announces to the reader an initiative focusing on a very specific corpus of Latin epigraphic poetry from *Hispania* and *Gallia*. It exploits existing tools to bring forward ideas in full continuity with previous scholarship on the poems. Chapter 25 looks at the outcomes of an extremely accurate imaging project, presenting the results obtained for the datation of the inscriptions of Aquincum with deep paleographic analysis.

The final contributions are related to 3D projects and methodologies. Chapter 26 looks at basic low cost possibilities like the Structure From Motion technology, while chapter 27 gives an overview of the potential of models and data used with a specific open access toolkit in the context of the Open-Access Epigraphy project. The direct impact on modeling and imaging research is presented in chapter 28, which complements the results presented in chapter 7 by applying the findings of the Morphological Residual Model to epigraphically relevant questions about objects and texts.

Finally, the two last contributions look at projects from a multi-disciplinary perspective. The EPNet project, in chapter 29, presents one possible way in which rich structured data can be deployed to understand wider controversial issues for the study of the economy of the Roman Empire with an otherwise impossible scientific base. The section is then concluded on a pleasant note in chapter 30 wherein an inscription is translated into real music by the Terpandros ensemble and accompanied by a 3-D modeling of the instrument which would have accompanied the text.

A list of the challenging and vibrant panels [A] held during the conference in Paris can then be found after the papers, followed also by a list of the posters [B].⁵

Our hope is that gatherings such as the EAGLE2014 Conference will further promote dialogue between specialists in different fields. We are confident that these efforts will lead to many fruitful future collaborations.

The EAGLE2014 conference aims to become a place where institutions, industries, the European Commission and Europeana family projects in the areas of cultural heritage can find ample opportunity for networking, debating, sharing ideas as well as best practices.

As editors of the volume and organizers of the conference, it is our pleasure to express our gratitude to the dedicated program co-chairs, committee members and conference support staff who have contributed to making the EAGLE 2014 international conference a sold-out event.

Silvia Orlandi

Vittore Casarosa

Raffaella Santucci

Pietro Maria Liuzzo

⁵ These contributions can also be seen in the digital online exhibition of the posters and in the videos: www.eagle-network.eu/about/events/eagle2014/digital-poster-exhibition/

Contents

PART I – EPIGRAPHIC DATA:	
MODELS, VOCABULARIES AND INTERACTIONS	21
1. A Conceptual Model for Inscriptions	23
<i>Vittore Casarosa, Paolo Manghi, Andrea Mannocci, Eydel Rivero Ruiz, Franco Zoppi</i>	
1.1. Introduction	24
1.2. EpiDoc and modern epigraphy	26
1.3. Analysis in CIDOC-CRM	27
1.4. The Aggregator Conceptual Model	31
1.5. The EAGLE implementation	34
1.6. Conclusions	38
2. Improving Text-Based Search of Inscriptions	41
<i>Michelangelo Ceci, Gianvito Pio, Anita Rocco</i>	
2.1. Introduction	41
2.2. Epigraphic text-based search in EDB	42
2.3. Exploiting lemmatization to improve text-based search	43
2.4. Conclusions and Future Work	47
3. From the LIMC Vocabulary to LOD	51
<i>Anne-Violaine Szabados</i>	
3.1. Introduction	51
3.2. Why the need for a controlled vocabulary?	52
3.3. From a controlled vocabulary to a standards-compliant thesaurus	53
3.4. ISO 25964 -compliant	58
3.5. Linked Data : opening new ways	61
3.6. Conclusion	63

4. Between Harmonization and Peculiarities of Scientific Domains	69
<i>Alessandra Avanzini, Annamaria De Santis, Daniele Marotta, Irene Rossi</i>	
4.1. Studying the epigraphic heritage of ancient Arabia: overview of the project DASI	69
4.2. Describing inscriptions	70
4.3. Publishing digitized inscriptions	84
4.4. Future developments	89
5. Epigraphy, Art History, Archaeology	95
<i>Antonio E. Felle, Norbert Zimmermann</i>	
5.1. EDB and spatial data	95
5.2. The Domitilla Project	97
5.3. Interaction between EDB and Domitilla Projekt	102
5.4. An example: the region of the Flavi Aurelii	103
5.5. Conclusion	112
6. Inscriptions visual recognition	117
<i>Giuseppe Amato, Fabrizio Falchi, Fausto Rabitti, Lucia Vadicamo</i>	
6.1. Introduction	117
6.2. Related Work	119
6.3. Tested approaches	119
6.4. Experiments	121
6.5. BoF with Geometric consistency checks	125
6.6. Comparison	126
6.7. Conclusions and Future Work	127
7. Morphological Residual Model	133
<i>Hugo Pires, João Fonte, Luis Gonçalves-Seco, Maria João Correia Santos, Orlando Sousa</i>	
7.1. Introduction	133
7.2. Data Capture	136
7.3. Methodological Approach	136
7.4. Visualization of M.R.M. results for epigraphic reading: four case-studies	138
7.5. Conclusions and Future Work	141

8. Image Editing Programs as Tools for the Study of Ancient Inscriptions	145
<i>Federico Frasson</i>	
8.1. Photos of inscriptions from Luna	145
8.2. Results	148
8.3. Conclusions	149
 PART II – TRANSLATING EPIGRAPHY: CHALLENGES AND RESEARCH OUT- COME	 153
9. Attic Inscriptions Online (AIO)	155
<i>Stephen Lambert, Finlay McCourt</i>	
9.1. Background to AIO	155
9.2. Target users of AIO	157
9.3. Key features of AIO	158
9.4. Technical aspects of AIO	161
9.5. Progress so far	162
9.6. Future plans and EAGLE context	162
10. Towards an EAGLE Standard in Translating Inscriptions	167
<i>Francesca Bigi</i>	
10.1. Introduction	167
10.2. Translations: an innovation	167
10.3. An EAGLE vocabulary for translations?	168
10.4. A vocabulary: some suggestions	169
10.5. What is to be translated	170
10.6. Names	174
10.7. Offices and formularies	175
11. Translating Greek and Roman Inscriptions	179
<i>Camilla Campedelli</i>	
11.1. Absence of the original context	181
11.2. Elliptical form of expression	181
11.3. Specific cultural lexicon	183

12. The EAGLE Mediawiki	187
<i>Pietro Maria Liuzzo, Andrea Zanni, Luca Martinelli, Lorenzo Losa, Pietro De Nicolao</i>	
12.1. EAGLE and Wikimedia Italia	187
12.2. The EAGLE Mediawiki in place	191
12.3. Networking the Mediawiki	196
12.4. Future developments	197
12.5. Conclusions	199
 PART III – USERS, EPIGRAPHY AND THE SOCIAL WEB	 203
13. Epigraphy as a tool for learning Latin	205
<i>Anja Ragolič</i>	
13.1. The Prežihov Voranc Primary School in Ljubljana	205
13.2. Research camps	206
13.3. Research papers	207
13.4. Research project Epigraphy	214
13.5. Guide to the lapidarium for primary schools	215
13.6. Conclusion	215
14. The Ashmolean Latin Inscriptions Project (AshLI)	221
<i>Jane Masségli</i>	
14.1. A Neglected Corpus	221
14.2. Gauging the Needs of Users	222
14.3. Proposing a Solution: AshLI	223
14.4. The Team	223
14.5. At the end of our first year – what we've learned	228
15. (Digital) epigraphy as viewed by Romanian Archaeology /Classics Students	233
<i>Rada Varga</i>	
15.1. Target groups and method	233
15.2. The survey	234
15.3. The general outline	235
15.4. Epigraphy and digital epigraphy	236

15.5. Conclusions	238
16. Meeting the Needs of Today's Audiences of Epigraphy with Digital Editions	239
<i>Laura Löser</i>	
16.1. A focus on purposes and needs	239
16.2. Aims in support of progressive development	242
16.3. Today's diverse audiences for epigraphy	243
16.4. How to meet diverse needs with modern editorial means	246
16.5. Conclusions	250
17. #svegliamuseo	255
<i>Francesca De Gottardo</i>	
17.1. The Project	255
17.2. The reaction of Italian museums and the #svegliamuseo community	258
17.3. Italian museums online: what has changed since 2013 and future developments	259
18. #DigitalInvasions	265
<i>Elisa Bonacini, Marianna Marcucci, Fabrizio Todisco</i>	
18.1. Prosumers and new ways of cultural heritage dissemination through UGC	265
18.2. #DigitalInvasions: best practice of crowd cultural value co-creation	269
18.3. #DigitalInvasions2014: a massive digital phenomenon from Italy to the world	277
18.4. Cultural policies and socio-digital impacts of #DigitalInvasions projects	278
19. Archeowiki: enhancing archaeological heritage in Lombardy (Italy) with open-source strategies	285
<i>Anna Antonini, Dante Bartoli, Sara Chiesa, Cristian Consonni, Rossella Di Marco, Sara Franco</i>	
19.1. Introduction	285
19.2. Digital archives for museums	286

19.3. Open Source initiatives in the cultural environment: ideas and examples	288
19.4. Wikipedia and Crowdsourcing initiatives in the cultural environment	291
19.5. Archeowiki project	294
19.6. Creative Commons licenses for cultural heritage protected by MIBACT: a viable solution	298
19.7. Chain reaction	302
20. Intellectual Property Right Issues	307
<i>Alessandra Giovenco</i>	
20.1. An overview of Intellectual Property, copyright and copy-left	307
20.2. Are all works protected by copyright? Some thoughts on the Italian IP law	309
20.3. The BSR process for solving IPR questions	310
20.4. Images from the South Etruria collection (inscriptions located in Italy)	313
20.5. Images from the Libya collection (inscriptions located in Libya)	314
20.6. Translations from Libyan inscriptions	315
20.7. Conclusions	315
PART IV – DIGITAL APPROACHES TO CROSS-DISCIPLINARY STUDIES OF INSCRIPTIONS	317
21. Digital Marmor Parium	319
<i>Monica Berti, Simona Stoyanova</i>	
21.1. The Digital Marmor Parium Project	319
21.2. The Digital Marmor Parium	320
22. The Inscription between text and object	325
<i>Emmanuelle Morlock, Eleonora Santin</i>	
22.1. Introduction and purposes	325
22.2. Interaction between text and object: four possible configurations	328

22.3. Defining concepts: key entities for the material and textual dimensions	336
22.4. The encoding strategy of the <i>IGLouvre</i> project	342
22.5. Conclusions and perspectives	347
23. The IGCyr project	351
<i>Alice Bencivenni, Simone Agrimonti</i>	
23.1. The IGCyr project	351
23.2. A laboratory for encoding IGCyr: challenges and problems of a user engagement based markup	359
24. Latin Epigraphic Poetry Database Project	369
<i>Concepción Fernández Martínez, María Limón Belén</i>	
24.1. Introduction: The <i>CLE Hispaniae</i> Project	369
24.2. Objectives of the project	371
24.3. The <i>CLE Hispaniae</i> website: www.clehispania.com	371
24.4. 4. Impending Project: the <i>CLE Galliae</i>	377
25. Paleographic Analysis of the Stone Monuments of Aquincum, Pannonia	381
<i>Nándor Agócs; István Gergő Farkas; Ádám Szabó, Ernő Szabó</i>	
25.1. Introduction	381
25.2. Methodology	384
25.3. The process of the project	384
25.4. Conclusion	392
25.5. The future of paleographical research in Pannonia	393
26. Low-cost Structure from Motion Technology	401
<i>Daniele Mittica, Michele Pellegrino, Anita Rocco</i>	
26.1. Introduction	401
26.2. Methodology	405
26.3. The workflow	409
26.4. Other reconstruction tools: Autodesk 123D Catch and Agisoft Photoscan	411
26.5. 3D model management: visualization and presentation	414
26.6. Conclusion and possible developments	414

27. Open-Access Epigraphy	421
<i>Eleni Bozia, Angelos Barmpoutis, Robert S. Wagman</i>	
27.1. Introduction	421
27.2. Dissemination of 3D epigraphic content	427
28. Travelling back in Time to Recapture Old Texts	437
<i>Maria João Correia Santos, Orlando Sousa, Hugo Pires, João Fonte, Luís Gonçalves-Seco</i>	
28.1. Introduction	437
28.2. Methodology	438
28.3. Travelling back in time to recapture old texts: four case studies	439
28.4. Towards some conclusions	450
29. The EPNet Project	455
<i>José Remesal, Albert Díaz-Guilera, Bernardo Rondelli, Xavier Rubio, Antonio Aguilera, Daniel Martín-Arroyo, Alessandro Mosca, Guillem Rull</i>	
29.1. Setting the focus	455
29.2. An ongoing debate: the economy of Roman Empire	456
29.3. A groundbreaking vision	457
29.4. Innovation	459
30. A Multi-Layered Reserach on an Ancient Cypriot Inscription	465
<i>Michalis Georgiou, Spyros Armostis, Sorin Hermon, Elena Christophorou, Valentina Vassallo</i>	
30.1. Introduction	465
30.2. The Eulalios inscription (AKGDC, E40)	467
30.3. 3D documentation of musical instruments	471
30.4. Metadata for Ancient Cypriot Inscriptions	474
30.5. Conclusions	475
A. Panels	479
A.1. Dealing with the Whole Object: the Archaeological Dimension of Epigraphy	480
A.2. Technology and tradition: a synergic approach to deciphering, analyzing and annotating epigraphic writings	482

A.3. Digital Humanities Publishing and Collaboration Strategies and Frameworks	483
A.4. Mobile Applications in Cultural Heritage	484
A.5. Linked Ancient World Data	486
B. List of Posters presented	487
C. List of authors	490

4. Between Harmonization and Peculiarities of Scientific Domains

Digitizing the epigraphic heritage of pre-Islamic Arabia in the project DASI

Alessandra Avanzini, Annamaria De Santis, Daniele Marotta, Irene Rossi

Abstract

DASI is an ERC-Advanced Grant project aimed at digitizing the pre-Islamic inscriptions from Arabia and fostering best practices for the digitization of the epigraphic heritage related to Semitic languages. This paper describes the content model, the standards chosen, and exemplifies the vocabularies in view of a possible harmonization of data pertaining to the specific domain. The architecture of the system and the tools for encoding and retrieving textual content are also illustrated.

Keywords

Digital epigraphy, pre-Islamic Arabia, vocabularies, text encoding, epigraphic standard.

4.1. Studying the epigraphic heritage of ancient Arabia: overview of the project DASI

From the late 2nd millennium BC until the advent of Islam, the huge territory designated as Arabia was a country of literate civilizations, which are known almost exclusively by their enormous epigraphic heritage in different languages and scripts.¹ However, Arabian studies is yet a very young field. Epigraphic material of pre-Islamic Arabia is only relatively studied and is often difficult to interpret. Furthermore, this cultural wealth today is hardly accessible, due to the complicated socio-political situation of some regions of the Peninsula.

¹ For an overview of the topic, see ROBIN 1991 and MACDONALD 2000.

The five-year project DASI – Digital Archive for the study of pre-Islamic Arabian inscriptions² has been funded by the European Research Council with the objective of getting the whole corpus of ancient Arabian inscriptions inventoried and digitized. At the third year of the project, nearly 7000 Ancient South Arabian inscriptions are accessible for fruition on the web, plus a number of inscriptions in the Aramaic and Ancient North Arabian languages,³ which are available thanks to the cooperation with UMR 8167-Orient & Méditerranée of the CNRS-Paris and the project OCIANA – Online *Corpus of Inscriptions of Ancient North Arabia of the University of Oxford*.⁴

The second objective of the project is to enhance Arabian studies by means of the edition of the material on the web, scientific publications and dissemination, in order to achieve a better overall perception and knowledge of the civilizations of pre-Islamic Arabia and stimulate the protection of their cultural heritage.

The third objective of DASI is to foster methodologies and getting a common understanding of best practices and procedures for the digitalization and preservation of the epigraphic heritage, with a special attention to the specificities of the pre-Islamic Arabian cultures and Semitic languages.

4.2. Describing inscriptions

DASI is the development of the project CSAI – *Corpus of South Arabian Inscriptions*, which had been set up in 2001 by the same research group of the University of Pisa. CSAI was a pioneering attempt to publish the entire Ancient South Arabian (hereafter ASA) epigraphic corpus in digital form.⁵ The digital edition of the inscriptions consisted of XML files, each one grouping the inscriptions belonging to one ASA

² The project, under the directorship of Prof. Alessandra Avanzini of the University of Pisa and with the collaboration of the Scuola Normale Superiore di Pisa for the technical part, was funded within the Seventh Framework Programme “Ideas”, Specific Programme “ERC – Advanced Grant”, Grant Agreement n° 269774. <http://www.dasiproject.eu>

³ DASI –*Digital Archive for the Study of pre-Islamic Arabian Inscriptions*. <http://dasi.humnet.unipi.it>

⁴ <http://krc2.orient.ox.ac.uk/aalc/index.php/en/news/286-the-safaitic-database-online-sdo>

⁵ The project is described in AVANZINI et al. 2000; see also AVANZINI 2008 for the editorial criteria.

sub-corpus. The description model focused on the text, whose editorial interventions and onomastic features were marked.

The clearly text-oriented model of CSAI has undergone changes over time. The digitization of inscriptions preserved in museums and the adoption of the holistic interpretation method, led to the enhancement of the description of the physical support in order to wholly appreciate the interdependency between the inscription and the objects carrying the text, which are often artistically valuable.⁶

4.2.1. A hybrid system for uniform data retrieving

The new approach of DASI is a compromise between the two models applied by most of the projects related to digital archiving and publication of epigraphs:⁷ the database approach and the XML approach. Whereas the latter – experimented in CSAI – had some limitations,⁸ DASI is a hybrid system combining both the technologies for archiving and displaying data. It consists of three main components: a relational database, a data entry and a front end.

The database stores not only metadata, but also text encoded in XML format according to the EpiDoc standard⁹, being the data entry provided with an editing module specifically developed to encode pre-Islamic Arabian inscriptions.¹⁰ In this way, the texts of the inscriptions are still “fully queryable and manipulable” [CAYLESS et al. 2009, p. 26].

On the other hand, the entire content of the database is extracted in XML by a web service, called XMLManager, in order to construct the dynamic sections of the front end. In particular, the text of the inscriptions is indexed by using a customization of the library Berkeley

⁶ The adoption of a multidisciplinary method of study – that includes the linguistic, the philological, the archaeological and the art-historical perspectives as well – was enhanced thanks to the funding of related projects such as CASIS and MENCAWAR, which focused on the direct examination and documentation of the inscriptions preserved in the museums involved.

⁷ See BABEU 2011 for a worldwide overview of single projects relating to digital epigraphy.

⁸ Namely, those relating to the management of multiple users providing data via web and of highly structured data (description of supports, origin of objects, etc.).

⁹ EpiDoc: Epigraphic Documents in TEI XML. <http://sourceforge.net/p/epidoc/wiki/Home/>

¹⁰ The advantages of a database and its web-based data entry system are well known: ease in data insertion via web browsers; use of controlled vocabularies that ensure uniformity of data; data sharing among a large number of users; integrity of data, due to different levels of authorization.

The dual nature of inscribed artefacts, as material objects carrying inscriptions and textual information carried by objects, is represented by two separate but strictly related entities: “Epigraph” and “Object” (Fig. 4.1). Each item digitized in DASI is represented by a physical object linked to one or more epigraphs, each one provided with its own images and bibliography.¹² As for the “Epigraph”, the model allows to overcome the limitations of the paper editions, still partly present in some projects of digital publishing based only on databases.¹³ The focus on the text, that already characterized CSAI, has been strengthened by the enhancement of the searches performed on the texts and a forthcoming lexicon of some South Arabian languages (see 4.4). Moreover the attributes of the entity “Epigraph” provide information on linguistic features, writing, chronology and type of text; in addition to the notes of apparatus, there are general and cultural notes.

Information about an object is not embedded in the attributes of the “Epigraph”, but has its own autonomy, while maintaining close ties with that of the inscription. Indeed, the entity “Object” includes attributes regarding: type of support, materials and dimensions, provenance and archaeological context, and the detailed description of its decorative elements.

In order to highlight the complexity of inscriptions, contextual entities are considered. For instance the entity “Site” has been connected with the entity “Object”. DASI supplies not only the information needed to contextualize artefacts, such as provenance or place of production, but also ancient and modern toponyms, geographical coordinates, country, region, ancient kingdom, archaeological information about the sites, such as monuments, history of studies, archaeological missions and so on.

Finally, the entity “Image” is linked both to “Epigraph” and “Object”: this double relation gives the opportunity to provide a suitable

¹² The data model of DASI lends itself to be represented by the CIDOC CRM conceptual model (<http://www.cidoc-crm.org/>) in view of a possible encoding in computational semantic, while preserving the advantages of the EpiDoc description. The benefits of this approach in terms of ontologies and *linked open data* are well exemplified in DOERR et al. 2010 and ÁLVAREZ et al. 2011.

¹³ BODARD 2008, p. 2: “The objects on which the texts are inscribed, the stelae, statues, wall panels, tablets, and grave monuments, are studied by archaeologists and art historians for whom the written texts are little more than a footnote, if not an inconvenience. This fact has tended to keep inscriptions in an academic limbo – not quite literary text and not quite archaeological object”. See also BABEU 2011, pp. 97-98.

visual documentation for both the object as a whole and the portions where the inscription is placed. This allows to perform the functions of “scale” and “media” [BODARD 2008] and provides more precise information (e.g. several inscriptions on a single support).

4.2.3. Transcription and encoding

The benefits of XML encoding of texts have led to the spread of EpiDoc within the scientific community [ROUECHÉ 2009]. This spread will be further increased if easy-to-use tools are offered to epigraphists. This is the reason why DASI has developed a tool that allows to obtain an EpiDoc text without learning XML.

An XML editor has thus been integrated into the data entry (Fig. 4.2). It does not handle a generic XML, but a set of EpiDoc elements to represent specific phenomena, according to the scientific needs of the project DASI. The encoding is transparent to the user, who is not bound to learn XML syntax and work directly on the code, but only uses menus and buttons to enter the Latin transliteration (UTF-8 set) and the phenomena to be highlighted.

The encoded phenomena are:

- structural phenomena (line breaks and in-word line breaks);
- phenomena concerning the relationship between text and support (broken support, turn of side);
- onomastics (names of individuals and social groups, toponyms, etc.);
- textual portions (signature, eponyms, etc.);
- transcription phenomena (ambiguous and lost characters, lacunae, text omitted);
- symbols (word divider, paraph);
- editorial interventions (restored text, superfluous characters, corrections);
- grammatical phenomena (grammatical analysis of the words).

The development of the editor faced some problems related to the creation of valid and well-formed XML documents. The editor prevents users from entering elements in positions not allowed by the EpiDoc schema and not considered within the project.

The editor verifies that tags are rightly nested in order to avoid overlapping. In this regard, *fragmentation* is the solution [MARINELLI et al. 2008] used by DASI: in case two elements overlap, the last element which was inserted, is divided into several fragments; each of these fragments is provided with an identifier and related with the others through linking attributes. This rule has been adopted because it is not predictable which information level will be encoded first. For example:

<supplied> inserted after <persName>

```
"b<supplied reason="lost" xml:id="e8" next="e9">r </supplied>
<persName type="gender" subtype="m"><supplied reason="lost"
xml:id="e9" prev="e8"> Dr</supplied>ħm</persName>
```

<persName> inserted after <supplied>

```
"b<supplied reason="lost">r
<per-sName type="gender" subtype="m" xml:id="e3" next="e4">
Dr</persName></supplied><persName type="gender"
subtype="m" xml:id="e4" prev="e3">ħm</persName>
```

Even though the EpiDoc guidelines suggest to use the Leiden conventions¹⁴, DASI only partly matches them. In fact, being the Leiden conventions focused on Greek and Latin sources, they are not always suitable to represent the same phenomena in pre-Islamic Arabian texts. For example, they represent the unclear text with a dot under the individual characters; in the transliteration of Semitic texts this would create ambiguity, because the dot is used as a diacritical mark to transcribe some phonemes typical of the Semitic languages (e.g. the dental emphatic phoneme, represented by the South Arabian letter 𐩦, is transcribed with the letter ḍ). Presently this is not a limitation, since the encoding is carried out with XML elements in the digital edition.

¹⁴ General Introduction To The EpiDoc Guidelines. <http://www.stoa.org/epidoc/gl/latest/intro-intro.html>

The screenshot displays the 'Epigraph text' interface. At the top, there's a header 'Epigraph text' with a help icon. Below it, a section titled 'Epigraph text ?' contains a 'Markers' button. The main area is divided into several columns: 'Individuals', 'Geo/Social', 'Gods', and 'Other names'. Each column has a dropdown menu. The 'Individuals' dropdown is open, showing a list of onomastic terms such as 'Proper Name - Male', 'Proper Name - Female', 'Proper Name - Unknown', 'Patronymic', 'Matronymic', 'Name of King with title', 'Name of King without title', 'Name of Queen with title', 'Name of Queen without title', 'Name of Member of Royal Family - Male', 'Name of Member of Royal Family - Female', 'Name of Member of Royal Family - Unknown', 'Name of Member of Royal Family - Patronymic', 'Name of Member of Royal Family - Matronymic', 'Name of King with royal title - Patronymic', 'Name of Queen with royal title - Patronymic', and several entries with Greek letters and numbers. To the right of the dropdowns are input fields for 'Epithet of god', 'Name of building', '# #', 'Foreign text', 'lat', 'Grammatical', and 'Substantive'. Below these fields is a transcription area with a list of entries, each starting with a number and followed by a transcription of a Greek inscription. At the bottom left, there is a button labeled '< > xml/text'.

Fig. 4.2. The textual editor of DASI data entry. Transcription and encoding of an inscription, with application of onomastic tags.

4.2.4. Vocabularies

Several metadata elements are provided through vocabulary fields. The elements requiring lists of controlled terms are related both to the textual aspect and content and to the material object.

A survey of current terminologies intended for description of both epigraphic and artistic objects has been carried out while building specific DASI lists. The “Writing technique” vocabulary, aimed at classifying only the method used to produce the texts, and the “Object material” vocabulary, including only the most common materials, could refer already existing terminologies.¹⁵ As regards the other elements, the

¹⁵ Respectively the EAGLE Vocabulary – Writing (<http://www.eagle-network.eu/voc/writing/index.php>) and the EAGLE Vocabulary – Material (<http://www.eagle-network.eu/voc/material/>). The vocabularies of the EAGLE consortium are among the authority lists and controlled vocabularies suggested by EpiDoc. Moreover, the epigraphic domain, the multilingual and alternative terms and the poli-hierarchical design are the reasons of this choice. The overview of the EAGLE vocabularies is taken from the project deliverable EVANGELISTI et al. 2014.

taxonomies and the tradition of Arabian studies do not allow to totally adhere to existing vocabularies and thesauri, as illustrated below, and have led to the creation of independent vocabularies.

4.2.4.1. Language

The classification of the ancient languages of the Arabian peninsula is one of the still debated topics among scholars of the ancient Near East. Even though general linguistic taxonomies including languages treated in DASI are already available,¹⁶ they do not allow to point out the linguistic peculiarities that the epigraphic sample of DASI shows, depending on the chronological and geographical distribution.

About the Ancient South Arabian¹⁷, the third level of the vocabulary (Tab. 4.1) has been added in order to represent further sub-divisions within the corpora corresponding to the linguistic varieties of the chief languages.¹⁸

In addition to the ASA languages, DASI includes epigraphic materials belonging to the Ancient North Arabian, the Aramaic and the Old Arabic linguistic groups. Similarly to ASA, discrepancies in the distinction between languages and varieties led to implement the independent vocabulary according to the classification of the experts in that linguistic domain.

¹⁶ 2009. Multitree: A digital library of language relationships (<http://multitree.org/>). Ypsilanti, MI: Institute for Language Information and Technology (LINGUIST List), Eastern Michigan University: useful thesaurus providing for each language the ISO 693-3 code, alternative labels, definition, hierarchical relations and bibliography.

¹⁷ Being the authors part of the “Corpus of South Arabian Inscriptions” research group, this paper refers mainly to the ancient South Arabian culture, which developed at the southern edge of Arabia (today’s Yemen) in the late 2nd millennium BC. A monumental and a minuscule alphabets were used to inscribe public and private texts in four different languages (Sabaic, Minaic, Qatabanic, Hadramitic), for a total of around 12,000 inscriptions known to date.

¹⁸ The Minaic inscriptions, for instance, are divided into Central Minaic inscriptions – coming from the motherland of the kingdom of Maʿīn – and the Marginal Minaic inscriptions – found outside southern Arabia. The Marginal Minaic differs from Central Minaic in terms of grammar, lexicon and formulas. The Sabaic documentation has been divided mainly on the chronological basis: the texts until the 4th century BC pertain to Early Sabaic; those testifying for the last two centuries of South Arabian history (4th-6th AD) are Late Sabaic; from the 4th-3rd century BC to the 4th AD, the inscriptions from the Sabaean homeland are considered Central Middle Sabaic, those from the high plateau south of Sana have been grouped as Southern Middle Sabaic and those originating from some areas of the northern regions of Yemen are labelled Northern Middle Sabaic. The peculiar traits of the Ancient South Arabian languages are described in STEIN 2011; see also AVANZINI 2004 for Qatabanic, PRIOLETTA 2013 for Hadramitic, and ROSSI 2014 for Minaic.

Group	Language	Variety
Ancient South Arabian	Hadramitic	
	Minaic	Central Minaic
		Marginal Minaic
		Undefined Minaic
	Qatabanic	Central Qatabanic
		Marginal Qatabanic
		Awsanite
		Undefined Qatabanic
	Sabaic	Early Sabaic
		Central Middle Sabaic
		Southern Middle Sabaic
		Northern Middle Sabaic
		Late Sabaic
		Undefined Sabaic
	Undefined South language	Ancient Arabian

Tab. 4.1. «Language» vocabulary. Selection of items from the vocabulary of DASI showing the addition of the level related to the varieties of the ASA languages.

4.2.4.2. Textual typology

This vocabulary is aimed at organizing inscriptions based on the typology of their texts. The categories it is made by, refer to fixed textual models attested in the inscriptions. These models were defined taking into account the formulary pattern, that is composed of [AVANZINI 2008, pp. 152-153; STEIN 2010b, pp. 264-269]:

- constituent parts: lexicon items (in particular the main verb) and morpho-syntactic features;
- their linear lay-out.

As in the previous vocabulary, the comparison with the existing ones¹⁹ has pointed out that some of the entries have exact matches, others are just related to some terms²⁰, and the remaining ones have no match at all (Tab. 4.2), since the textual typologies reflect the particular use that every civilization makes of writing.²¹

DASI entry	Definition	Match
Construction text	<i>text posted on public or private structures, reminding who constructed or restored them (eventually in order to indicate their possession); the text typically includes verbs meaning “build”, “found” (bny, hḥdt)</i>	Bauinschrift
Dedicatory text	<i>dedication of an object or a person to the divinity for different purposes, the text including the verb “dedicate” (ḥqny, sʕlʕ)</i>	RT Weihinschrift
Writing exercise	<i>text related to scribal practice (copies of sample texts, lists of grammatical forms, etc.)</i>	
Ritual text	<i>text reminding religious rituals</i>	

Tab. 4.2. «Textual typology» vocabulary. The selected terms exemplify the results of the comparison between the DASI vocabulary and existing vocabularies used in the epigraphic domain synthesized by EAGLE – Type of Inscription.

4.2.4.3. Support type

The list of controlled terms referring to the physical supports of inscriptions includes a sort of classification of the ancient Arabian material

¹⁹ EAGLE Vocabulary – Type of inscription. <http://www.eagle-network.eu/voc/typeins/>

²⁰ This kind of relation occurs, for instance, when the preferred label and the scope note of a term explicitly refer to the Greek and Roman specimen, as expected, due to the domain of the network.

²¹ The minuscule inscriptions on wooden sticks offer a unique insight of the private uses of writing in ancient South Arabia. The rich typology of those texts deduced by STEIN 2010a is a clear example of the complex classification of pre-Islamic Arabian inscriptions based on content.

culture. It has been created on the basis of the repertoires focused on individual classes of artefacts and the early attempts at systematizing the history of pre-Islamic Arabian art.²² Thus, as knowledge of the latter advances, the vocabulary is still in progress.

²² See, as examples, the works by ANTONINI 2001, 'ALĪ 'AQĪL et al. 2007, ANTONINI DE MAIGRET 2012, respectively about the stone statues, the bronze artefacts and the history of art in general, and the catalogues of the exhibition on the ancient Yemen, traveling in Europe between 1997 and 2002 (see e.g. SIMPSON 2002).

DASI Term	Sub-term	Definition	Match
Incense burner	circular	objects used to burn aromatic substances, with circular shape	AAT: incense burner [ID: 300198817]
	squared	the censer has a squared shape, sometimes placed on four angular legs	
	with truncated pyramidal base	characterized by a censer placed on a base of truncated pyramidal shape	
	with splayed foot	short cylindrical censer on a splayed foot	
Stela	aniconic stela	typical of the Qatabanian area, they are constituted by an inscribed base supporting an upper and plain slab. The top can be either convex or concave	AAT: stelae [ID: 300007023] EAGLE: Stele (http://www.eagle-network.eu/voc/objtyp/lod/250)
	with eyes or stylized face	particular and homogeneous group from the Jawf area and few samples from Qatabani, representing eyes or a schematic facial outline	
	with figure in relief	isolated human or animal figures, usually realized in high relief	
	with narrative scene	decorated by narrative scenes, usually composed by multiple elements depicted in low relief	
– continued on next page			

Table 4.3 – continued from previous page

DASI Term	Sub-term	Definition	Match
	<i>with framework</i>	<i>refined by a decorated framework in low-relief with abstract-figurative elements or with geometric motifs, associated in various and different combinations</i>	

Tab. 4.3. «Support type» vocabulary. Sample from the vocabulary illustrating the degree of correspondence between the list of terms of DASI and the EAGLE – Object Type vocabulary, focused on supports of inscriptions, and the Getty AAT, which is instead a general thesaurus for artistic objects.

Presently, it results to be compatible with the Getty – AAT and the EAGLE – Object Type vocabulary²³ for artefacts whose identification is strictly related to the inscriptions they carry. However DASI includes further terms intended to represent the peculiarities of the ASA art.²⁴

4.2.4.4. Decoration

In spite of its title, the “Decoration” vocabulary does not include the decorative motifs surrounding the inscriptions only. Due to the importance and complexity of the representations, resembling sometimes proper figurative scenes, the vocabulary has been implemented so as to enable the description also of figurative subjects and iconographies.

The standard classification systems already available,²⁵ being focused on the western visual culture, presently does not consider some of the figurative subjects distinguishing the ASA artefacts. Among them, the symbols are commonly found in ancient Near East context as well: used to express a concept in a synthetic and striking way, they could be part of the decoration but are closely linked to the inscription too. Their description is of great importance, since their study and interpretation is still at the beginning.²⁶

²³ The Getty Research Institute – Art & Architecture Thesaurus Online (<http://www.getty.edu/research/tools/vocabularies/aat/>); EAGLE Vocabulary – Object Type <http://www.eagle-network.eu/voc/objtyp>

²⁴ For instance the term “stela” has been split into five entries referring to as many types of stela: the “aniconic stela”, the stela “with eyes or stylized face”, “with figure in relief”, “with narrative scene” and “with framework”, whose different physical characteristics correspond to different dating, zones of provenance/creation and functions.

²⁵ IconClass (<http://www.iconclass.nl/>) is the most suitable among the standard classification systems for subjects in visual works: through hierarchical paths and queues of keys it allows to identify single elements and the context and the significance of scenes.

²⁶ As an example, the so-called “Totschläger”, a sort of curved stick, is attested in contexts related to the main Sabaeen god Almaqah; the simultaneous analysis of its representations, the text and the figurative subjects could light up the nature of this relation. See: GROHMANN 1914 and GAJDA 2012 for issues about ASA symbols; BLACK et al. 1992 for an overview of the ancient Near East figurative subject and the above mentioned ANTONINI DE MAIGRET 2012 dedicated to ASA art.

DASI Entry	Iconclass notation	EAGLE term
Recumbent antelope	25F24(ANTELOPE)(+534) <i>hoofed animals: antelope</i> (+ <i>squatting, crouching animal</i>)	
Head of antelope	25F24(ANTELOPE)(+33) <i>hoofed animals: antelope (+ head of an animal)</i>	
Human gesture – hand raised	31A2511 <i>arm raised upward</i>	
Crescent moon and disc		
Totschläger		
Guilloche	48A98131(+1) <i>ornament ~ circle and derived from circle, e.g.: guilloche (+ surface pattern ~ ornaments)</i>	Guillochis

Tab. 4.4. «Decoration» vocabulary. Examples of: figurative subjects corresponding to Iconclass notations, decoration fitting with terms from the EAGLE – Decoration vocabulary and DASI entries with no concordance.

4.3. Publishing digitized inscriptions

4.3.1. Content browsing

The front end of DASI²⁷ publishes data distinguished into objects, epigraphs and sites, by different levels. The level determines the amount of material that can be accessed via indexes and tools.

The top level, including all the materials archived and published, allows content browsing through a Google map in the homepage. Here are positioned the sites from which users can consult the related records.

The “Project” sub-levels, corresponding to the major linguistic groups (e.g. Ancient South Arabian, Ancient North Arabian), consists in turn of “Corpora” and “Collections”, that include materials homogeneous for language and deposit respectively. Both provide some textual tools, the indexes of inscriptions, objects and sites, in addition to the above mentioned map, and a list of bibliographic references. Each index gives access to data by an alphabetic list of items that can be

²⁷ DASI – Digital Archive for the Study of pre-Islamic Arabian Inscriptions. <http://dasi.humnet.unipi.it/>

screened through a set of filters on metadata. The filters are almost the same as those used to refine the textual search results (see below 4.3).

Data describing each item – be it an inscription, an object or a site – are then displayed in a specific card. Each type of card collects not only the information concerning the pertaining entity, but also pieces of information relating to the related entities which are considered fundamental to complete its description.

CIAS 47.11/o 1/F 72 Q 269; Ja 122; CSAI I, 125



INSRIPTION INFORMATION

Language	Ancient South Arabian » Qatabanic » Central Qatabanic	
Alphabet	Ancient South Arabian	
Script typology	Monumental writing	
Writing technique	Incision	
Measure of letters	2.5	
Chronology	Period	C
Textual typology	Dedicatory text	

TEXT

marker

1

Br't d't byt Rtd'I bn S²hz s¹q—

2

nyt d't Hmym 'itr Ygi slmt d—

3

hbn hg-n tkrbt-s¹ l-wfy-s¹ w-wf—

4

y 'dn-s¹ w-mqm-s¹ w-qny-s¹ qzrt

5

'm d-Rbh w rs²wt 'm d-D←ymt

6

b→Wrw'I Gyln Yhn'm |

Fig. 4.3. Upper part of an epigraph card in DASI front end site. Sigla, image, metadata and transcription of an inscription. Further details are given below the text: translation, information on the support and its provenance, bibliography.

For example, an epigraph card includes the text of the inscription, its translation(s) and other data concerning the text (Fig. 4.3), and also the basic information on its support and on its origin/provenance context, provided with the links to the related “object” and “site” cards. This representation of the content contributes to highlight the autonomy of the various entities and allows a rapid consultation while providing complete information.

4.3.2. Textual tools

The major effort in the publication of the catalogued material has been made to provide tools for the study of the epigraphic text. Indeed the reconstruction of lexicon and grammar of the languages of the Arabian peninsula is one of the most urgent needs for a full comprehension of the inscriptions and, consequently, of the culture and history of these civilizations.

The *textual search* is one of the most sophisticated functions of the DASI system, which is a powerful mean for the scientific study of the inscriptions. The system allows to search for strings of characters and display results as:

1. text portions
2. distinct words.

For instance, entering the pattern “*t *ml*”²⁸, the system retrieves all the texts consisting of two words, the first one ending with “t” and the second one containing “ml”. The results:

²⁸ The wildcard * allows to search for: complete words or words with a given string of characters in any position (beginning, end, inside). The search finds also words beginning, ending or containing lacunae.

- in the case 1., are displayed as the text portions “bnt yqhmlk”, “bynht ʔmlk”, “gdr̥t mlk”, etc.
- in the case 2., are grouped by the words “ʔmlk”, “bnt”, “bynht”, “gdr̥t”, “mlk”, “yqhmlk”, etc.

The examples above show that the system assumes the boolean operator AND between adjacent strings of characters. The filter *Word range* sets the maximum number of words that can intervene between the first and the last words to be searched for. The system retrieves effective results if the value of the range is $r \geq n-1$, where n is the number of strings typed in the research box. For example, if one searches for the three strings “b* *tm *m*”:

- with $r=2$, the system retrieves the following contexts: “S^{2c}rm **b-dtm** n^cmt w-b”, “s^{2ft}-s¹ **b-dtm** s¹mnt ʔtr̥t”, etc.
- with $r=3$, the system retrieves the following contexts: “w-mqtr̥nhn d̥-b ʔtwtm w-ms¹l̥mn d̥-dhbn d̥-b”; “S^{2c}rm **b-dtm** n^cmt w-b”; “s^{2ft}-s¹ **b-dtm** s¹mnt ʔtr̥t”, “s^{3b}-s¹ **b-dtm** s¹wfy Hwkm ^cmdkr̥”, etc.

The textual search is performed by the Berkeley DB XML on the texts al-ready tokenized. This process segments the text on the basis of rules that con-sider a space, a divider symbol (<g type=“wordSeparator”/>) or a clitic com-ponent (<milestone unit=“clitic”/>) as the elements distinguishing one word from another: if one searches for the strings “f ʔtww”, the system will retrieve all three of the forms “f ʔtww”, “f | ʔtww”, “f-ʔtww”.

The system is also able to search for textual variants. DASI uses both in-line and external apparatus methods to encode variants. The latter is preferred when the editor chooses one reading or onomastic interpretation only: the variants are recorded in separate notes and cannot be searched. On the contrary, variants are coded in the text of the inscriptions when, from a point of view of the reading or of the interpretation of the text, they have the same probability of being true. Therefore a textual search can be performed on them. For example, in a context $w_1 w_2 ((w_{2.1})) ((w_{2.2})) w_3$, the system can retrieve $w_1 w_2 w_3$ but also $w_1 w_{2.1} w_3$ and $w_1 w_{2.2} w_3$. This output is obtained by adding specific attributes to words while indexing, which are:

- the number of the word (w_2 , $w_{2.1}$, $w_{2.2}$ have $n=2$)
- the level (L0: w_2 ; L1: $w_{2.1}$; L2: $w_{2.2}$; all the other words are on the L0)
- the id of the group of variants, as there might be more than one in a text (w_2 , $w_{2.1}$, $w_{2.2}$ have $id=1$).

The example shows the principles followed:

- each variant of a specific group adds a level to the text: in the example there are 3 levels,
- the search combines the variants of all the levels with the level 0 of the other words.

The system manages a series of filters to refine the search on the texts of the inscriptions (onomastic categories, textual portions) and on the metadata (language, textual typology, support, decoration, site, etc.).

The search function generates a list of results presented in a concise and legible manner. It shows their distribution among languages and their onomastic category (if any) at a glance, allows to view the contexts in which they appear and then to access the inscriptions, where the text searched for is highlighted in bold.

In addition to the dynamic textual search tool, DASI provides lists of words in alphabetical order, which can be generated according to the level of corpora at which the lists are accessed (e.g. the whole ASA corpus, or the Qatabanic corpus, or one of its sub-varieties, etc.; Fig. 4.4). There are two different *lists of words*: one includes the words beginning with a letter, which can be complete or contain a lacuna inside or at the end of the word; the other one lists the words beginning with a lacuna and therefore lacking the first letter that would assign them to an index letter in alphabetic order. Both lists are created though a textual search following the same principles of tokenization showed above, but with no string searched for. Each of these lists can be refined to give only the lexical or the onomastic results, as a whole or by a specific onomastic category.

TOOLS

Q Textual search

Word lists

Map

INDEXES

Corpora map

Collections map

Epigraphs

Objects

Sites

Bibliography

Help

Editorial Criteria

Word list occurrences

10 total occurrences for word **m'ḥg**

activate image preview

language	epigraph	context
Early Sabaic	MAFRAY-Husn Āi Sālih 1	» 3Y gb b-fdldm b-m'ḥg N m ⁴ (n) l-m'ḥg smn «
Early Sabaic	MAFRAY-Husn Āi Sālih 1	» b-m'ḥg N m ⁴ (n) l-m'ḥg smn g-Wdm w- «
Early Sabaic	MAFRAY-Husn Āi Sālih 1	» s ² 2mwy yd' bn m'ḥg N mnn' d ¹⁰ S ² wn s ² - «
Early Sabaic	MAFRAY-Husn Āi Sālih 1	» s ² -hw yd' n bn m'ḥg s ¹¹ mn' d-g Ws ² bm «
Early Sabaic	MSM 3611	» w-hwdr w-hqb m(ḥ)ḡḡ-hmy Byhn l-ms'qt «
Early Sabaic	RES 3902 bis N° 131	» Lzn] 2bny jw-kl m'ḥg-hw Yfd g-3ys'qyn «
Central Middle Sabaic	Ghul Hūl 2	» w-hwr 2n w-syhn m'ḥg-hmw Dhbm ms'qt s ² - «
Central Middle Sabaic	Ir 7	» S ² htj k-y twm dy m'ḥg-hmw g-Yfd s'qym «
Central Middle Sabaic	RES 3911	» s ² qr 2[...]s ² w-mqh m'ḥg-hmw Dbyn w-hry «
Central Middle Sabaic	Ry 404	» (g)- M (g)nm w-rjd w m'ḥ(g)-hmw w-brt-hw «

Fig. 4.4. Example of occurrences of a word retrieved through the word list in DASI front end site. The occurrences of the word *mr*² in the Qatabanic corpus are listed, with the sub-corpora they belong to, the sigla of the inscriptions and the occurrence's context in the inscription.

4.4. Future developments

Since the enhancement of knowledge of the ancient Arabian civilization is one of the main objectives of DASI, the future efforts of the project are directed toward the interoperability of records and the harmonization of controlled vocabularies.

On the first side, DASI has committed to make available its data in different export formats. Surely records will be entirely distributed according to the EpiDoc standard that is the one presently assuring the valorization of texts and allowing the digital edition of inscriptions.

As regards the vocabularies, DASI will refer to external terminologies, as recommended by the EpiDoc guidelines themselves, as far as possible. To some extent it will try to interact with other projects and contribute concepts not included in already existing vocabularies.

Finally, the creation of digital *lexica* of the ASA languages is already scheduled. In fact, this is one of most urgent needs for the studies on Ancient Arabian epigraphy, as the amount of linguistic material has increased exponentially in the last years, calling for updating of established dictionaries. The lexicon will consist of a list of roots; each root will be entered into the system and provided with its definition. The words deriving from it and attested in the inscriptions will then be

linked to the pertaining root and analyzed in grammar, with the aid of the context of its occurrences provided by the word list.

Bibliography

- ‘ALĪ ‘AQĪL, A. and S. ANTONINI (2007). *Bronzi sudarabici di periodo pre-islamico*. Roma: IsIAO, Paris: Académie des inscriptions et belles lettres. ISBN: 9788885320437 (cit. on p. 80).
- ÁLVAREZ, F.-L., J.-L. GÓMEZ-PANTOJA, and E. GARCÍA-BARRIOCANAL (2011). “From relational databases to linked data in epigraphy: Hispania epigraphica online”. In: *Metadata and Semantic Research*. Communications in Computer and Information Science. Izmir: Springer, pp. 225–233. URL: http://link.springer.com/chapter/10.1007/978-3-642-24731-6_24 (visited on 07/29/2014) (cit. on p. 73).
- ANTONINI DE MAIGRET, S. (2012). *South Arabian art: art history in pre-Islamic Yemen*. Paris: De Boccard. ISBN: 9782701803258 (cit. on pp. 80, 83).
- ANTONINI, S. (2001). Repertorio iconografico sudarabico 1. Roma: Istituto italiano per l’Africa e l’Oriente ; Académie des inscriptions et belles lettres (cit. on p. 80).
- AVANZINI, A. (2004). *Corpus of South Arabian inscriptions I-III: Qatabanic, Marginal Qatabanic, Awsanite inscriptions*. Pisa: Edizioni Plus/Università di Pisa. ISBN: 9788884922632 (cit. on p. 77).
- (2008). “Criteri editoriali per la pubblicazione dello CSAI”. In: *Egitto e Vicino Oriente* 31, pp. 145–156 (cit. on pp. 70, 78).
- AVANZINI, A., G. MAZZINI, and D. LOMBARDINI (2000). “Corpus of South Arabian Inscriptions. La pubblicazione integrale del corpus sudarabico qatabanico”. In: *Bollettino del centro di ricerche informatiche per i beni culturali* 10, pp. 73–83 (cit. on p. 70).
- BABEU, A. (2011). *Rome Wasn’t Digitized in a Day: Building a Cyberinfrastructure for Digital Classics*. Council on Library and Information Resources. ISBN: 9781932326383. URL: <http://www.clir.org/pubs/reports/pub150/pub150.pdf> (cit. on pp. 71, 73).
- BLACK, J. and A. GREEN (1992). *Gods, Demons and Symbols of Ancient Mesopotamia: An Illustrated Dictionary*. University of Texas Press. ISBN: 9780292707948 (cit. on p. 83).

- BODARD, G. (2008). "The Inscriptions of Aphrodisias as Electronic Publication: a user's perspective and a proposed paradigm". In: *Digital Medievalist* 4. URL: <http://www.digitalmedievalist.org/journal/4/bodard/> (cit. on pp. 73, 74).
- CAYLESS, H., C. ROUECHÉ, T. ELLIOTT, and G. BODARD (2009). "Epigraphy in 2017". In: *Digital Humanities Quarterly* 3.1. URL: <http://www.digitalhumanities.org/dhq/vol/3/1/000030/000030.html> (visited on 02/21/2013) (cit. on pp. 71, 72).
- DOERR, M., K. SCHALLER, and M. THEODORIDOU (2010). "Integration of complementary archaeological sources". In: *Beyond the artefact. Digital Interpretation of the Past. Proceedings of CAA2004*. Ed. by F. Niccolucci and S. Hermon. URL: http://proceedings.caaconference.org/files/2004/09%5C_Doerr%5C_et%5C_al%5C_CAA%5C_2004.pdf (cit. on p. 73).
- EVANGELISTI, S., P. M. LIUZZO, and H. VERRETH (2014). *Content Harmonisation guidelines, including GIS and terminologies*. URL: <http://www.eagle-network.eu/about/documents-deliverables/> (cit. on p. 76).
- GAJDA, I. (2012). "Liens entre symboles et divinités dans les inscriptions sudarabiques". In: *Dieux et déesses d'Arabie images et représentations: actes de la table ronde tenue au Collège de France (Paris) les 1er et 2 octobre 2007*. Ed. by I. Sachet and C. Robin. Paris: De Boccard, pp. 447–459. ISBN: 9782701803067 (cit. on p. 83).
- GROHMANN, A. (1914). *Göttersymbole und Symboltiere auf süd-arabischen Denkmälern*. Vienna: In Kommission bei Alfred Hölder (cit. on p. 83).
- MACDONALD, M. C. A. (2000). "Reflections on the linguistic map of pre-Islamic Arabia". In: *Arabian Archaeology and Epigraphy* 11.1, pp. 28–79. ISSN: 1600-0471. DOI: 10.1111/j.1600-0471.2000.aae110106.x. URL: <http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0471.2000.aae110106.x/abstract> (visited on 07/17/2014) (cit. on p. 69).
- MARINELLI, P., F. VITALI, and S. ZACCHIROLI (2008). "Towards the unification of formats for overlapping markup". In: *New Review of Hypermedia and Multimedia* 14.1, pp. 57–94. URL: <http://www.tandfonline.com/doi/abs/10.1080/13614560802316145> (visited on 07/29/2014) (cit. on p. 75).
- PRIOLETTA, A. (2013). "Remarks on some processes of assimilation and innovation in the language and culture of Hadramawt during its ancient history". In: *Aula orientalis: revista de estudios del Próximo Oriente Antiguo* 31.1, pp. 93–108. URL: <http://dialnet.unirioja.es/servlet/articulo?codigo=4721113> (visited on 07/29/2014) (cit. on p. 77).
- ROBIN, C. (1991). "Introduction". In: *Revue du monde musulman et de la Méditerranée* 61.1, pp. 9–12. ISSN: 0997-1327. DOI: 10.3406/remmm.1991.1502 (cit. on p. 69).
- ROSSI, I. (2014). "The Minaeans beyond Maʿīn". In: *Languages of Southern Arabia. Papers from the Special Session of the Seminar for Arabian Studies*

- held on 27th July 2013*. Ed. by O. Elmaz and J. Watson. Supplement to the Proceedings of the Seminar for Arabian Studies 44. Oxford: Archaeopress, pp. 111–124. ISBN: 9781905739813 (cit. on p. 77).
- ROUECHÉ, C. (2009). "Digitizing inscribed texts". In: *Text editing, print and the digital world*, pp. 160–168 (cit. on p. 74).
- SIMPSON, S. J., ed. (2002). *Queen of Sheba: treasures from ancient Yemen*. British Museum Press (cit. on p. 80).
- STEIN, P. (2010a). *Die altsüdarabischen Minuskelinschriften auf Holzstäbchen aus der Bayerischen Staatsbibliothek in München*. Epigraphische Forschungen auf der Arabischen Halbinsel 5. Tübingen: Wasmuth (cit. on p. 79).
- (2010b). "Literacy in Pre-Islamic Arabia: An Analysis of the Epigraphic Evidence". In: *The Qur'ān in Context. Historica and Literary Investigations into the Qur'ānic Milieu*. Leiden: Brill. Ed. by A. Neuwirth, N. Sinai, and M. Marx. Leiden-Boston, pp. 255–280. URL: <http://booksandjournals.brillonline.com/content/books/10.1163/ej.9789004176881.i-864.58> (visited on 07/29/2014) (cit. on p. 78).
- (2011). "Ancient South Arabian". In: *The Semitic Languages, An International Handbook*. Ed. by S. Weninger. Berlin, Boston: De Gruyter Mouton. ISBN: 9783110251586. URL: <http://www.degruyter.com/view/product/175227> (visited on 07/29/2014) (cit. on p. 77).

Acknowledgement

DASI is the result of the work of a large group of collaborators. We wish to acknowledge their effort. A special thank goes to Prof. Amos Bertolacci and Mr. Umberto Parrini, scientific responsible and technical coordinator for the part of the Scuola Normale Superiore, and Dr. Alessia Priolella, scientific coordinator for the part of the University of Pisa.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 269774.