



A. De Santis, I. Rossi
Scuola Normale Superiore di Pisa

Is still Arabia at the margins of
digital epigraphy?
Challenges in the digitization of the pre-
Islamic inscriptions in the project DASI

6th EAGLE International Event
Off the beaten track. Epigraphy at the borders

Bari, September 24th to 25th 2015



Is still Arabia at the margins of digital epigraphy?



DASI - Digital Archive for the Study of pre-Islamic Arabian Inscriptions

- ❖ **Grant:** ERC – Advanced Grant, 7th Framework programme
- ❖ **Duration:** 5 years (2011-2016)
- ❖ **Principal Investigator:** Alessandra Avanzini
- ❖ **Host institution:** University of Pisa – Dip. Civiltà e Forme del Sapere
- ❖ **Additional participant:** Scuola Normale Superiore di Pisa
- ❖ **Objects:** pre-Islamic inscriptions from the Arabian peninsula



UNIVERSITÀ DI PISA



SCUOLA
NORMALE
SUPERIORE



Is still Arabia at the margins of digital epigraphy?



Ancient North Arabian



Aramaic



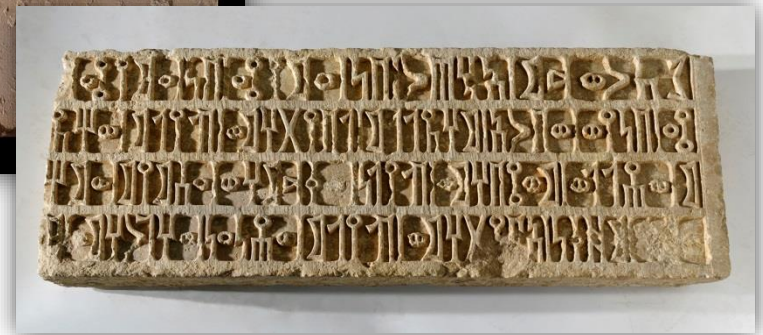
Ancient South Arabian

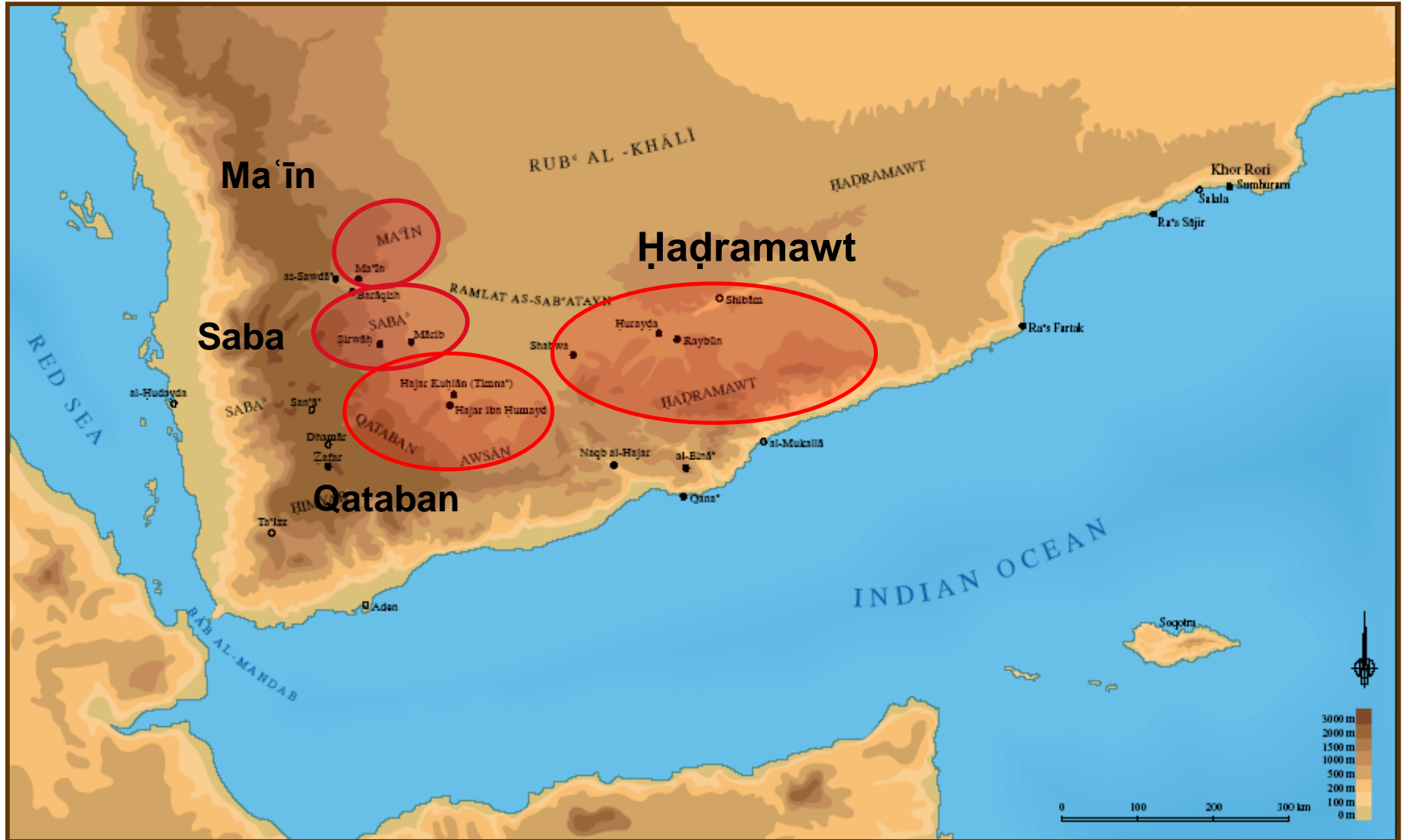


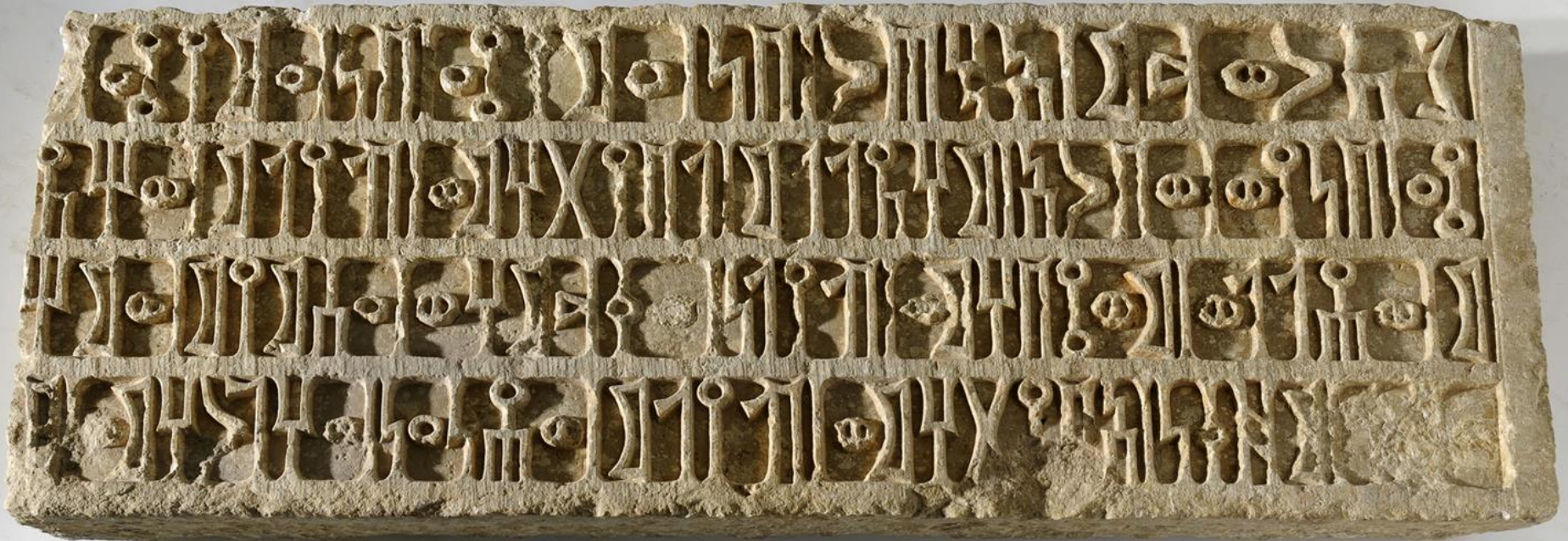
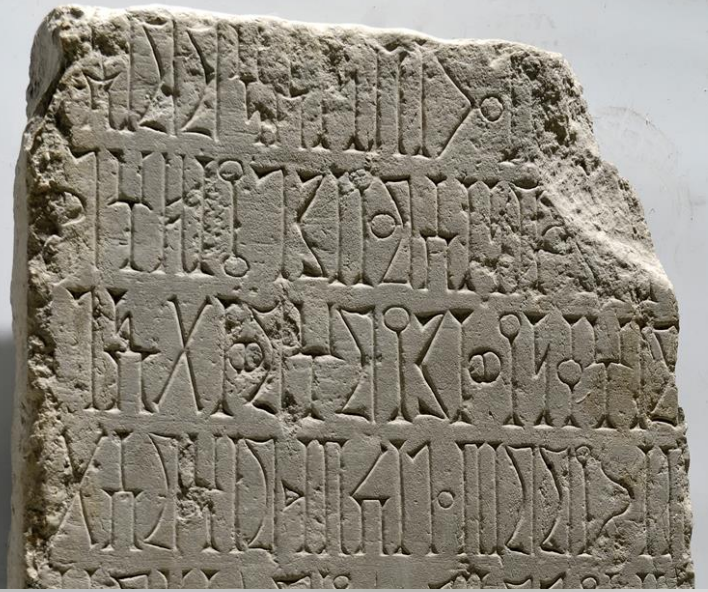
MEDINA - Mediterranean Network for the valorization and fruition of the inscriptions preserved in museums

funded by EU with the ENPI – CBCMed programme:

- South Arabian collection of the Museo Nazionale d'Arte Orientale di Roma
- Phoenician inscriptions of the National Museum of Beirut
- Nabataean inscriptions of the Yarmouk University Museum

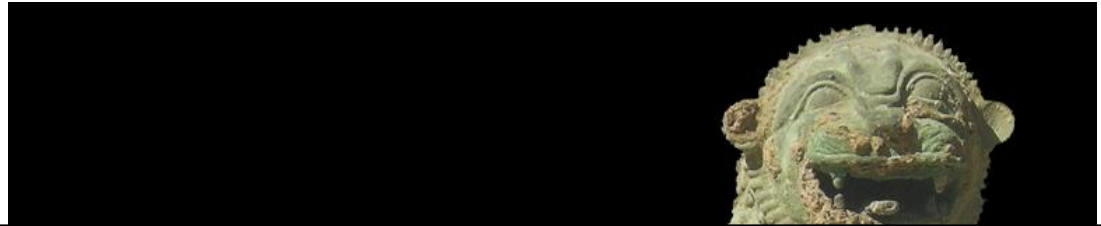








Is still Arabia at the margins of digital epigraphy?





DASI – Specific objectives

- ❖ continuing the mass digitization of the pre-Islamic Arabian epigraphic heritage;
- ❖ enhancing historic and cultural knowledge of Ancient Arabia;
- ❖ strengthening the linguistic study of texts (reconstruction of lexicon and grammar)



The experience of CSAI

Website's home

Encoding method

CORPUS OF SOUTH ARABIAN INSCRIPTIONS

Home Download font Editing criteria Help Contacts News

Digital edition of the South Arabian epigraphic corpus

The objective of the CSAI project, which began in 1999 is to collect the published Ancient South Arabian inscriptions. New corpora have been added in recent years: the catalogue of the British Museum inscriptions, and those of a number of Yemeni museums. For the first time, texts written in minuscule writing are also included into the database. Up to the present, the published corpus of non-monumental texts from south Arabia is relatively small, but a big edition of new texts is going to be published.

→ continue...

South Arabian Inscriptions

- All the inscriptions
- Single corpora by language
 - Hadramitic Inscriptions
 - Minaic Inscriptions
 - Marginal Minaic Inscriptions
 - Qatabanic Inscriptions
 - Awsanite Inscriptions
 - Marginal Qatabanic Inscriptions
 - Sabaic Inscriptions
 - Ancient Sabaic Inscriptions (work in progress)
 - Middle Central Sabaic Inscriptions (work in progress)
 - Raydanite Inscriptions (work in progress)
 - Late Sabaic Inscriptions (work in progress)
 - Undated Sabaic Inscriptions (work in progress)
 - Inscriptions of Undefined Language

South Arabian Minuscule Texts

- South Arabian Minuscule Texts

Other Semitic Inscriptions

- Corpus of Phoenician Inscriptions from the Beirut National Museum
- Corpus of Inscriptions from Jordan

Demo version

- Try the demo version of CSAI

```

344 <TU TYPE="COMPLETE">
345 <EPHH>
346 <EPHT>
347 <TIT>al-Jawf 04.11</TIT></EPHT>
348 <EPHB>
349 <BIB>Arbach and Schiettecatte 2006: 24, pl. 5/12</BIB>
350 </EPHB>
351 <EPHC>
352 <L VAL="Minaic"/>
353 <P FILE="aljawf0411.jpg"/>
354 <E FILE="aljawf0411.html"/>
355 <A VAL="Ns2n"/>
356 <M VAL="as-Sawd&#x0101;a["/>
357 <C VAL="B"/>
358 <T VAL="Dedicatory_text"/>
359 <D VAL="Sjana&#x0101;a[, The National Museum"/>
360 <G>This text could seem a fake for some anomalies: the irregular writing, the odd form of patronymic, the expression k+s1ma]t at the t
361 <SUPPORT_TYPE><OBJECT><STONE_INSCRIPTION/></OBJECT></SUPPORT_TYPE>
362 <SUPPORT_INFO><GENERAL_INFO><MEASURES>h. 28, w. 32, th. 10, lh. 4</MEASURES></GENERAL_INFO></SUPPORT_INFO></
363 </EPHC>
364 <CULT><ENTRY TOPIC="Society" KEYWORD="Law">If we maintain this text original, it turns out to be interesting as it shows a legal fo
365 <EPH>
366 <EL N="1"/><PROP GENDER="M">H]ma]t</PROP> bn <PATR>A]m<BW/>nyhn<PATR>
367 <EL N="2"/>k+s1ma]t <PROP>Y<BW/>hrmik</PROP>
368 <EL N="3"/>w+<TRIB>Ns2n</TRIB> h] <BW/>t]d
369 <EL N="4"/><TEON>A] <UREAD PTR="B"/>t] <UREAD PTR="E"/>tr</TEON> dj+<EPIDIV>Grb</EPIDIV>
370 <EL N="5"/>nfs1+s1 w+wld+s1 w+<EL N="6"/>qny+s1 <WVS/>
371 </EPH><EPRH><REM>L. 2. Yhmik might be the name of a king of Nashsh&#x0101;n. </REM>
372 <REM>Ll.3-4: the root h]T]D] is never attested. We cannot exclude that it is an error for r]d] or s1h]d]. </REM></EPRH></TU>
  
```



DASI SYSTEM

✕ Relational database

- metadata
- XML text

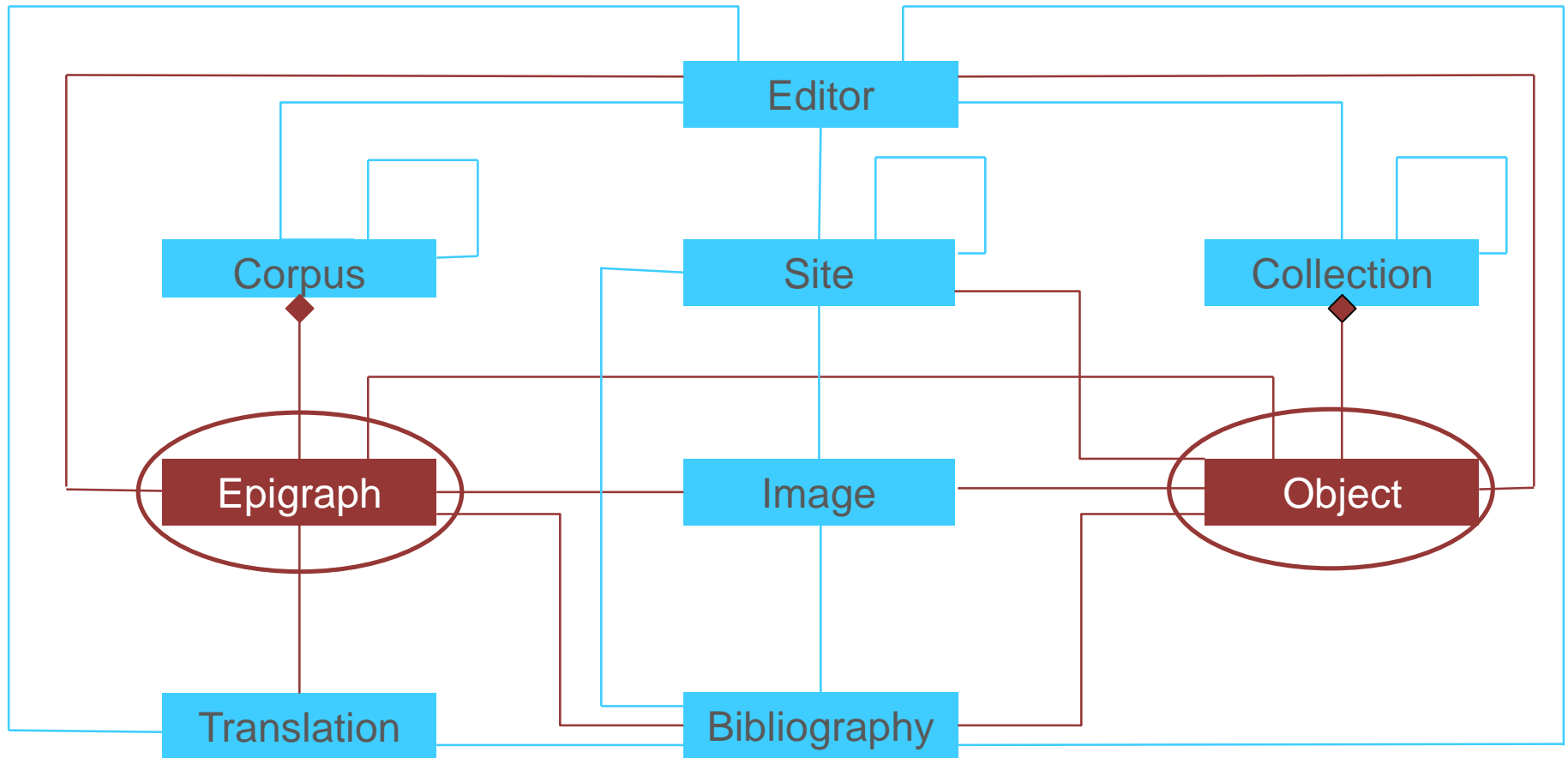
✕ Data entry

✕ Front end

- content of the database is extracted in XML by the web service XMLManager, in order to construct the dynamic sections of the front end



DASI Conceptual model





“Epigraph” with DASI textual editor module

CARDS

- Epigraphs
- Corpora
- Objects
- Translations
- Sites
- Collections
- Editors
- Bibliography
- Images

home page

download guide

web site preview

prints download

✎ Edit epigraph

Main data | Text and apparatus notes | Other notes | Relations | Images

Epigraph text

Epigraph text ?

▶ T Markers

Individuals	Geo/Social	Gods	Other names
Proper Name - Male	Nisbe	Epithet of god	
Textual portions	Structural	Foreign text	
Compound Name	Broken Support	lat	

Button bar for critical phenomena: () [] [...] <> <.> {} ** (()) - - | +

Special Characters

1. ▶Br't◀ gt byt ▶Rtd'l4 bn ▶S'hz◀ s'q—
2. nyt ▶gt Hmym ◀ ttr ▶Yol◀ slmt g—
3. hbn hg-n tkrbt-s' l-wfy-s' w-wf—
4. y 'gn-s' w-mqm-s' w-qny-s' qzr
5. ▶'m◀ g- ▶Rbhw◀ rs'wt ▶'m◀ g-▶D◀-ymt◀
6. b-▶▶Wrw'l◀▶Gyln◀▶Yhn'm◀◀ |

```
<lb n="5"/><name type="theonym"> m</name> g<milestone unit="clitic"/><name type="epithet" subtype="divine">Rbhw</name> rs'wt
<name type="theonym"> m</name> g<milestone unit="clitic"/><name type="epithet" subtype="divine">D<milestone unit="turnSide"
rend="l"/>ymt</name>
<lb n="6"/>b<milestone unit="clitic"/><rs type="complex"><persName type="royal" subtype="withoutTitle" role="king"
n="1">Wrw'l</persName> <persName type="royal" subtype="withoutTitle" role="king" n="2">Gyln</persName> <persName
type="royal" subtype="withoutTitle" role="king" n="3">Yhn'm</persName></rs> <g type="wordSeparator">|</g>
```



DASI – Encoding of text

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



“Object” entity

Object identification	
Status	approved
User	(import)
Work group	Csai
Object id	518
Object title	CIAS 47.11/o 1/F 72

Support description	
Support type	Object » Artefact » Base » Of statue with dedicatory inscription
Measures	Stone base: h. 17, w. 26, th. 5.8; bronze statuette: h. 18

▲ Materials (Number of records: 2)	
Material	Bronze
Material	Stone

▲ Decorations (Number of records: 1)	
Decoration Id	190
Figurative subject	Human figure
Human gender	Woman
Part of human body	Complete figure
Human gesture(s)	Sitting, Arms stretching out
Human clothes and attributes	Short sleeved tunic, Sandals
Human jewellery	Diadem, Necklace, Ankle band





“Site” entity

Site identification	
Status	approved
User	()
Site id	67
Modern name	Hajar Kuhlān
Ancient name	Tmna ¹
Site data	
Latitude	15° 01' 11"
Longitude	45° 48' 16"
Country	Yemen
Governorate	Shabwa
Geographical area	wādī Bayhān
Chronology	Probes carried out by the Italian Mission have revealed that the settlement has been inhabited since the beginning of the 1st millennium BC. A hiatus has been detected and dated between the 6th and the 5th centuries BC, while the apogee was during the 5th – 1st centuries BC. During the 1st century AD Timna ¹ suffered two major destructions, all archaeologically documented (fire traces) and dated thanks to the finding of imported pottery (mainly Arretine ware and Near Eastern Sigillata). At the second half of the 1st century AD the site has probably already lost its capital status, which was transferred to the nearby site of Hajar ibn Humayd. The kingdom was soon afterwards definitely conquered by Ḥaḍramawt.
General description	Extension: 23 ha. The site was encircled by a defence wall measuring about 680 x 420 m, and whose length has been estimated in 1850 m. The mound has a height of 20 m in respect to the surrounding valley (to be partly attributed to a natural elevation). It has remained abandoned since its last destruction in antiquity, and only its SE edge is now occupied by a modern settlement.
Classical sources	Strabo , Geographia, 16.4.2 (Tamna) [citing Erathostenes of Cyrene] Plinius the Elder, Naturalis Historia, 6.32.153; 12.32.64 (Thomna) Claudius Ptolemaeus, Geographia, 6.7.37 (Thumna) [?]
Identification	1895: E. Glaser
Travellers	1899: G.W. Bury 1936: S. Perowne



City walls and gates

South-West gate: Two massive bastions (11 x 5 m) frame the entrance. They have both an ample vane at their interior and they have been realized using imposing granite blocks (e.g. 1.70 x 2.80 m) which have been also incised with important legal inscriptions (e.g. RES 3878, 3879) or recalling construction activities. The passage is 5.50 m large (externally) and 4.20 m (at the interior of the town) and was originally paved and with two benches on each lateral side, while a small staircase was put against the eastern bastion. No traces of a closing device of the doorway have been detected. **North-West gate:** Only construction inscriptions (MAFYS-Timna 2-5 and MAFYS-Timna 1) allow so far the reconstruction of this gate, which has not been excavated scientifically and has moreover suffered an intense pillaging. According to the texts, these gates were formed by a massive granite basement (like that still visible in the SW Gate) which sustained an upper structure made in wood and mud bricks. Limestone and calcite-alabaster blocks enriched this upper part.



South Gate



South Gate



Is still Arabia at the margins of digital epigraphy?



DASI front-end

DASI DIGITAL ARCHIVE FOR THE STUDY OF PRE-ISLAMIC ARABIAN INSCRIPTIONS

DASI Home Projects

TOOLS

- Map
- Help
- Editorial Criteria

Corpus of South Arabian Inscriptions

Online Corpus of the Inscriptions of Ancient North Arabia

DASI

Digital Archive for the Study of pre-Islamic Arabian Inscriptions

DASI is a five-year project directed by Alessandra Avanzini of the University of Pisa, which has been funded by the European Community within the Seventh Framework Programme "Ideas", Specific Programme "ERC – Advanced Grant".

DASI seeks to gather all known pre-Islamic Arabian epigraphic material into a comprehensive online database, with the aim to make available to specialists and to the broader public a wide array of documents often underestimated because of their difficulty of access.

By means of a digitization process through a hybrid data entry/xml system developed by the Scuola Normale Superiore di Pisa according to international encoding standards, DASI gives access at present to nearly 7,500 Ancient South Arabian inscriptions recorded by the University of Pisa team under the direction of Alessandra Avanzini.

Thanks to the fruitful collaboration with other two major European centres for the study of the epigraphic documentation of the Arabian Peninsula, also sections of the two corpora of Ancient North Arabian inscriptions (supervision by Michael C.A. Macdonald, University of Oxford) and Aramaic inscriptions (supervision by Laila Nehmé, UMR 8187, CNRS-Paris) are being prepared for consultation.

[Project details »](#)

Projects

Corpus of South Arabian Inscriptions

The ASA (Ancient South Arabian) corpus is one of the most interesting collections of epigraphic documents of the Semitic world, first and foremost for its vastness. With its over 15,000 inscriptions, it is the first-hand, written documentation of the culture that flourished in South Arabia from the late second millennium BC to the sixth century AD. At present, CSAI contains a collection of some 7,500 texts digitized by the team of the University of Pisa directed by A. Avanzini.



Textual search

The screenshot shows a web interface for a textual search tool. On the left is a sidebar with 'TOOLS' (Textual search, Word lists, Map) and 'INDEXES' (Corpora map, Collections map, Epigraphs, Objects, Sites, Bibliography, Help, Editorial Criteria). The main area is titled 'Textual search' and contains a search input field with the text '*šlm* dħb*' and a 'search' button. Below the input field is explanatory text about wildcard usage. Underneath, there are radio buttons for 'Showing results as:' with 'word list' selected and 'text portions (experimental)' unselected. To the right of these is a 'Word range:' dropdown menu. Below that is a 'Filter by' dropdown menu set to 'any'. At the bottom, there is a vertical list of filter categories: Onomastic categories, Textual portions, Language and Script, Chronology of the inscription, Textual Typology, Sites, Objects, and Decorations.



Textual search

Results

Search results for `*šlm* dhb*`
12 different words in a total of 144 occurrences found

1 d s

4 words start with d

word	language	lexicon/onomastics	occurrences
d hb	Qatabanic		
		lexicon	1
d hbhn	Hadramitic		
		lexicon	1
d hbm	Qatabanic		
		lexicon	6
	Sabaic		
		lexicon	2
d hbn	Hadramitic		
		lexicon	1
	Qatabanic		
		lexicon	23



Textual search

DASI Home / CSAI / Textual search / Results

TOOLS

- Textual search
- Word lists
- Map

INDEXES

- Corpora map
- Collections map
- Epigraphs
- Objects
- Sites
- Bibliography

- Help
- Editorial Criteria

Results

Search results for *šlm* ḡhb*

11 different text portions in a total of 72 occurrences found

1 5

3 text portions start with 'šlm ḡhb

text	language	occurrences
'šlm ḡhbhn	Ḥaḍramitic	1
'šlm ḡhbn	Ḥaḍramitic	1
'šlmm ḡhbm	Qatabanic	1



TOOLS

- Textual search
- Word lists
- Map

INDEXES

- Corpora map
- Collections map
- Epigraphs
- Objects
- Sites
- Bibliography

Help
Editorial Criteria

Textual search

šlm dħb* ḥg *krb* s*

search

Type words to search preceded and / or followed by a wildcard (*), so as to obtain results respectively constituted by all the words that end, begin or contain within them that sequence.
The wildcard character can be used only once at each end of the search string and also once within the word.

Showing results as:

word list
 text portions (experimental)

Filter by

any

Word range:

4

2
3
4
5
6
7
8
9
10
11
12
13
14
15

Onomastic categories

Textual portions

Language and Script

Chronology of the inscription

Textual Typology

Sites

Objects

Decorations



Tokenization rules

❖ Spaces

❖ Tags indicating:

- divider symbols (<g type="wordSeparator"/>: mlk|hw)
- geometric signature (<g type="paraph">: #)
- line breaks (<lb n="n"/>)
- clitic components (<milestone unit="clitic"/>: mlk-hw < *mlkhw*)
- variants (<rdg>: mlk((ml'))hw)

E.g.: SEARCH: “f ’ tww” > “f ’ tww”
“f| ’ tww”
“f- ’ tww”, etc



TOOLS

Q Textual search

Word lists

Map

INDEXES

Corpora map

Collections map

Epigraphs

Objects

Sites

Bibliography

Help

Editorial Criteria

General word list

Filter words by: lexicon only

all (no filter)

lexicon only

onomastics only

7757 different words filtered by lexicon only

. ' ˆ b d ḍ ḍ f g ḡ h ḥ ḥ i k l m n q r s s¹ s²

s³ š t ṭ ṭ w y z z

164 words start with š

1 - 20 21 - 40 41 - 60 61 - 80 81 - 100 101 - 120 121 - 140 141 - 160

word	language	lexicon/onomastics	occurrences
šlmtn	Sabaic	lexicon	23
šlmtnhn	Sabaic	lexicon	3
šlmtynn	Sabaic	lexicon	1

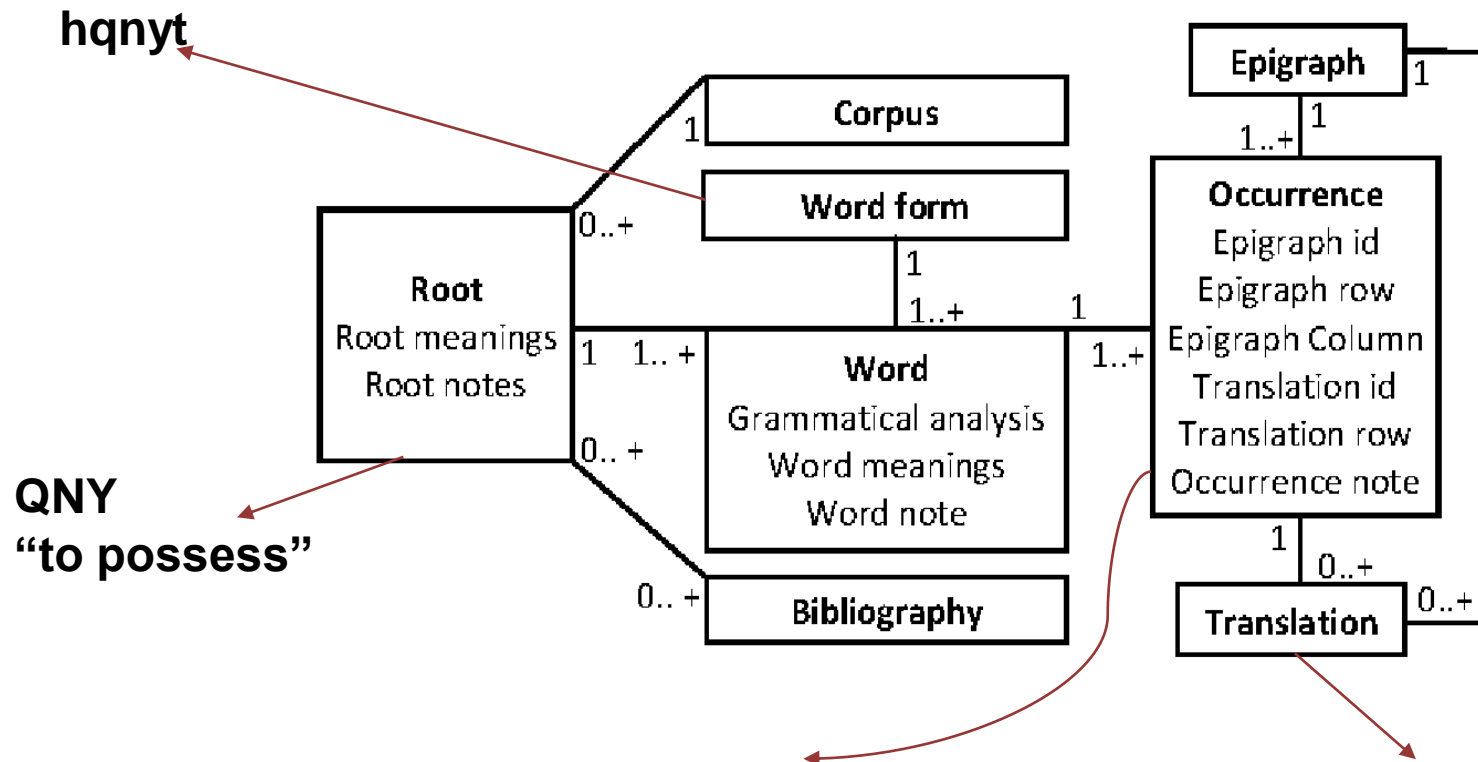


DASI Lexicon

The screenshot shows the DASI website interface. At the top, the header includes the DASI logo and name, the user name 'Rossi', and links for 'new prints', 'profile', and 'log out'. The left sidebar contains a 'Utility' menu with 'download guide', 'web site preview', 'lexicon' (circled in red), and 'prints download'. Below this are 'Admin' sections for 'Work Groups', 'Users', and 'Help'. The main content area is titled 'Home' and features several panels: 'Epigraphs', 'Corpora', 'Objects', 'Translations', 'Sites', 'Collections', and 'Editors', each with search and insert buttons. On the right, there is a 'Bibliography' panel with search and insert buttons for monographs, theses, articles in monographs and journals, web articles, and internal articles, plus a 'list of author/editor names' button. Below that is an 'Images' panel with search and insert buttons, and a 'Vocabs' panel with a 'manage vocabs' button.



Conceptual model of DASI Lexicon



CIAS 39.11/o 3 n° 1 » w-rtdw **hqnyt**-hmw

CIAS 39.11/o 3 n° 10 » Ddh(m)d `mh bn Hdw(t) **hqnyt**
`lmqh

“and committed their dedication”

“Ddhmd servant of the family
Hdw(t) dedicated to `lmqh”



LEXICON | CSAI CORPUS OF MINAIC INSCRIPTIONS

Admin

Edit

Preview

Recycler

Welcome Rossi | [Close Lexicon](#)

Preview

13 different roots in a total of 24 words and 153 occurrences

◌ ◌ B **Ḍ** M Q

3 roots start with **Ḍ** and include 4 words, 10 occurrences

root	meanings	notes	bibliography	words
ḌBH	sacrifice			2 words
ḌHB ⁽¹⁾	bronze			2 words
ḌHB ⁽²⁾	watered valley			2 words

[^ top](#)

word form	occurrences
ḍhb	<input type="checkbox"/> ḌHB (watered valley) / Substantive - Construct - Masculine singular 2 occ. <input type="checkbox"/> ḌHB (bronze) / Substantive - Unclear - Masculine singular 1 occ.
ḍhbn	<input type="checkbox"/> ḌHB (bronze) / Substantive - Emphatic/defined by article - Masculine singular 3 occ. <input type="checkbox"/> ḌHB (watered valley) / Substantive - Emphatic/defined by article - Masculine singular 1 occ.



LEXICON | CSAI CORPUS OF MINAIC INSCRIPTIONS

Admin

Edit

Preview

Recycler

Welcome Rossi | Close L

[Preview](#) / [Words](#) / [Occurrences](#)

ḡhbn

root: ḡHB | meanings: bronze

grammar: Substantive - Emphatic/defined by article - Masculine singular

word meanings: bronze

3 total occurrences

language	epigraph	context	translations
Central Minaic	Moussaieff 10	» 'l'lt ḡnḡr((ḡnḡr)) m' q ⁴ b ḡh(b)n w-rḡd 'hl 'qrn «	<ul style="list-style-type: none"> CSAI - English (rows: 3) ³ for 'ttr ḡ-Qbḡ, god of ḡnḡr, and the gods of ḡnḡr, this bronze
Central Minaic	YM 24942+YM 24943	» '(w)hn ms'lm w-mfḡmy ḡhbn «	<ul style="list-style-type: none"> CSAI - English (rows: 2) ² to 'ttr Mtbqbd, god of 'whn, and 'whn an altar and two incense burners in bronze.
Marginal Minaic	Riyād 302F8	» kl 'l'lt M ⁶ n s'qnyt ḡhbn b-gntn ⁷ b-Qryt «	<ul style="list-style-type: none"> Robin in Al-Ghabban 2010 - French (rows: 6) ⁶ in l'offrande de bronze dans l'Oasis



A manual approach. The reasons

❖ Epigraphic nature of the textual sources

- Restricted range of topics, formulaic contexts > limited set of words
- *Hapax legomena* > data sparsity

❖ Morphology of the Semitic languages

- Word formation based on the “root-and-pattern” mechanism + morphemes
- Orthographic unit composed by more than one lexical item (wmlk < w-mlk)
- Morphological and spelling ambiguity

> more than one analysis for a given word form

(*hqnyt* “dedication of” or “she dedicated”)

> more than one word form for one morphological analysis and meaning

(*ḥyl* or *ḥl* “power”; *s¹b¹t* or *s¹by¹* “expeditions”)



- ❖ Semantic and morphological study is still in progress
 - avoid to encode grammatical and lexicographic information on the texts
 - need to modify specific occurrences or groups of them in a reasonable lapse of time
 - need to verify/elaborate rules and correct integrations and interpretations



DASI Lexicon is a preliminary step before an automatic word extractor and a morphological analyzer can be attempted



Is still Arabia at the margins of digital epigraphy?



Thank you for your kind attention!

DASI

dasi.humnet.unipi.it

<http://www.dasiproject.eu/>