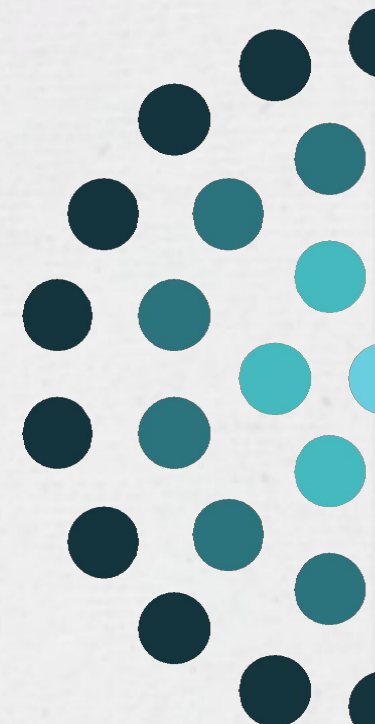


Open data

Creating datasets, documenting, depositing, data papers



FIRST WINTER SCHOOL 2025

Bankya

04-07 February 2025

Dimitar Iliev

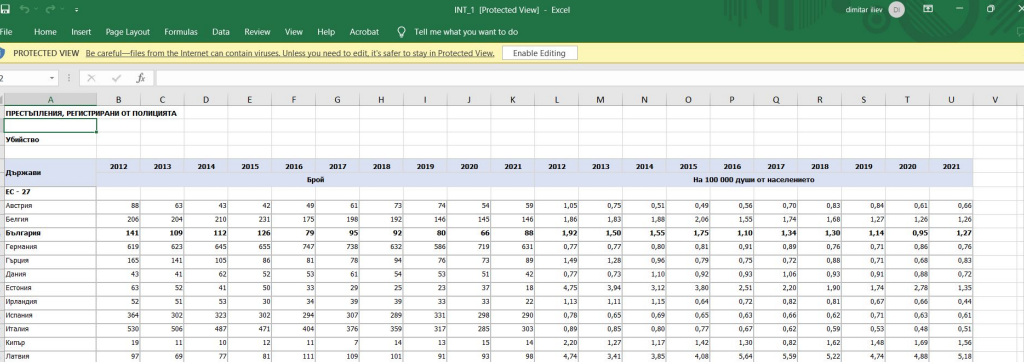
Department of Classics, St. Kliment Ohridski University of Sofia/CLaDA-BG-DARIAH-EU NCC

A



Data comes in all shapes and colours

Възраст в месеци	твърдения	Поставете "X"
РЕЧ И КОМУНИКАЦИЯ		
2	Издаване на гласни звуци, гукане. Липсва издаване на гласни звуци, гукане.	
6	Детето води "монолог" във вид на издаване на гласни звуци или реагира с такъв при появата на родител Необоснован плач Различава се интонация в гукането, включително когато го питат.	
8	Поява на повтарящи се срички : ба-ба-ба, ма-ма-ма. Поява на указателни жестове. Ограничено или необичайно гукане (квичене, крясък). Липсва на имитиране на звуци, жестове, лицева експресия.	
12	Поява на първа дума. Използва лексикална интонация, наподобяваща на изречение. Езикът започва по-често да се използва за коментар на откриващата среда. Води игра с използване на гласни звуци.	
18	Използва жестове и вокализации за привличане на вниманието, показване на обекти, молби. Появила се е първата дума, която обаче не се използва по предназначение. Често произвежда силен вик.	
24	Словесен запас - 3-50 думи. Започва да съставя словосъчетания от две думи. Използва езика за коментар, молба и извършване на действия. Привлича целенасочено вниманието. Възможна ехолалия (повтаря казаното от друг). Започва често да имитира. Появила се е първата дума, която обаче не се използва по предназначение. Често произвежда силен вик.	

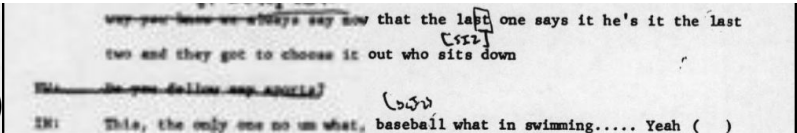


Data for <https://digitalsee.epistone.net/en/index>

```
</list>
</keywords>
</textClass>
</profileDesc>
<revisionDesc>
  <change when="2024-03-01" who="Ivan Valchev">encoder Ivan Valchev</change>
  <change when="2024-09-10" who="KS">Kristiyan Simeonov, encoded XML data</change>
  <change when="2024-09-10" who="DI">Dimitar Iliev, provided full schema revision</change>
</revisionDesc>
</teiHeader>
<text>
  <body>
    <div type="textpart" xml:lang="lat"> <ab></ab></div> <!-- xml:lang тук е за оригиналния език -->
    <div type="edition" xml:lang="deu">
      <p>
        <seg xml:lang="deu">worzu sie noch setzen, daß der Bruder nicht weit von hier ein Schloß gehabt, wovon sie einem noch zur linken Hand, wann man von Nissa kommt, am Ende der Stadt die Mauern weisen.</seg>
      </p>
    </div>
    <div type="commentary" subtype="general">
      <ab> The text is written between <date type="initial">None</date> and <date type="final">None</date>.
      The name of the monument is <name type="source">Sophia </name> or <name type="contemporary">Sofia</name>.
      According to the text, it is It is located <location>
        <geo>42.70725; 23.32167</geo> <geo>None</geo>
      </location>. <!-- произход според текста -->Now it is located here: <location type="original"><geo>None;None</geo>
      <geo>None</geo></location>, in <name type="current" xml:lang="deu">None</name>.
      The internal date in the document is <date>30.07.1718</date>.
      We have evidence that the object was observed in <location type="subsequent1"><geo>None; None</geo> </location>.
      The object has been seen in <location type="subsequent2"><geo>None; None</geo> </location> </ab>
    </div>
    <div type="commentary" subtype="additional">
```

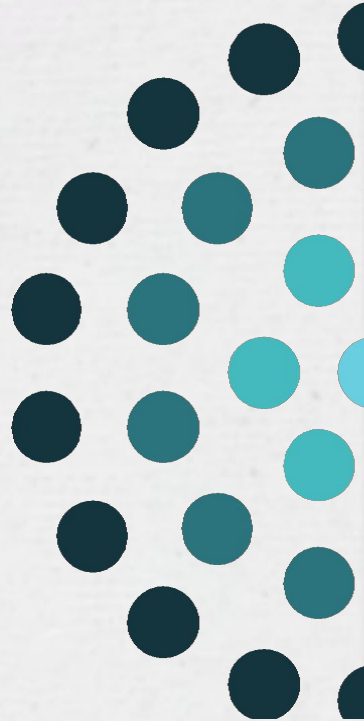
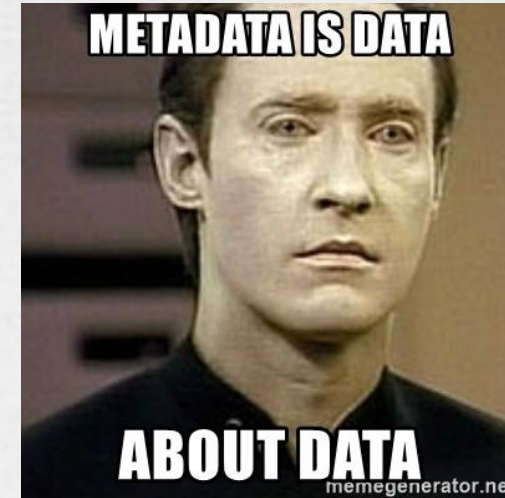
(<https://kunchev.blog.bg/>)

(<http://hdl.loc.gov/loc.afc/afc1986022.ms1505>)



The data cycle

- Data creation
- Data collection
- Data integration/standardization/unification
- Data enrichment
 - metadata
 - paradata
- Data extraction/analysis/modelling



A



Case study 1: the TELAMON collection



Monuments



Printed corpora



Museums

<http://telamon.uni-sofia.bg>

The Telamon template

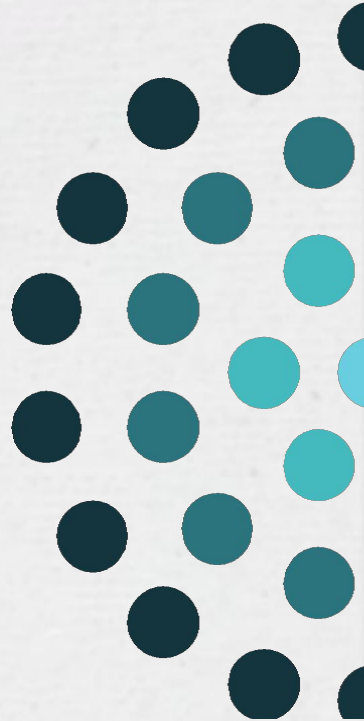
```
24 <sourceDesc>
25 <msDesc>
26 <msIdentifier><country/>
27 <altIdentifier xml:lang="bg">
28 <repository></repository>
29 <idno></idno>
30 </altIdentifier>
31 <altIdentifier xml:lang="en">
32 <repository></repository>
33 <idno></idno>
34 </altIdentifier>
35 </msIdentifier>
36 <msContents>
37 <summary corresp="document-type.xml">
38 <seg xml:lang="bg"></seg>
39 <seg xml:lang="en"></seg>
40 </summary>
41 </msContents>
42 <physDesc>
43 <objectDesc>
44 <summaryDesc>
```



Database



<https://epidoc.stoa.org/>



Case study 2: the DigitalSEE collection

DigitalSEE Digital Repository



Софийския университет „Св. Климент Охридски“
Маркер за иновации и технологичен трансфер SUMMIT



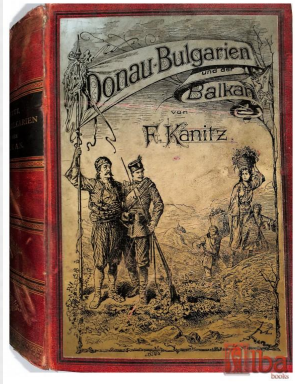
Home Objects Indices ▾ About us Bulgarian

Search...

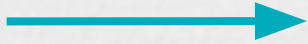


<https://digitalsee.epistone.net/> (demo)

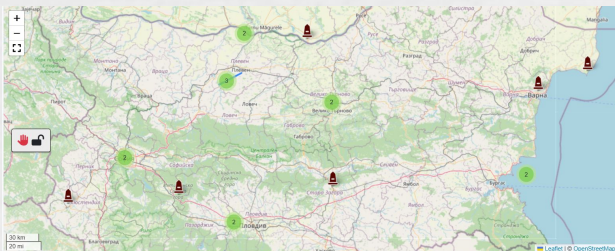
Historical documents



Archaeological sites



In situ visits



TEI-compliant
XML dataset



XML input form

DigitalSEE View Open raw XML Edit Add Map/Image Report bug Logout

General information

Filename:

Author:

Name (according to source):

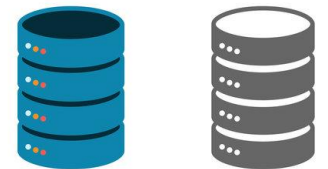
Name (contemporary):

Description:

Provenance Origin:

Latitude:
e.g., 32.12345

Longitude:
e.g., -12.34567



Custom-built
XML
database



Your data

- Added value!
- Consistency

```
2 <?xml-model href="http://www.stoa.org/epidoc/schema/latest/tei-epidoc.rng" schematypens="http://relaxng.org/ns/structure/1.0"?>
3 <?xml-model href="http://www.stoa.org/epidoc/schema/latest/tei-epidoc.rng" schematypens="http://purl.oclc.org/dsdl/schematron"?>
```

- Ascribing and describing authority and responsibility

```
<title xml:lang="bg">База за статуя на император Марк Аврелий, издигната от град Августа Траяна</title>
<title xml:lang="en">Statue base for Emperor Marcus Aurelius, erected by the city of Augusta Traiana</title>
<editor>
  <persName xml:lang="bg">Георги Михайлов</persName>
  <persName xml:lang="en">Georgi Mihailov</persName>
</editor>
<editor>
  <persName xml:lang="bg">Николай Шаранков</persName>
  <persName xml:lang="en">Nicolay Sharankov</persName>
</editor>
</titleStmnt>
```



```
100 <revisionDesc>
101   <change when="2018-06-13" who="EB">encoded</change>
102   <change when="2021-02-16" who="EB">final revision</change> </revisionDesc>
```

- Documenting choices

1. Title

The title of the inscription consists of a description of the document type taken from the Document type authority file and:

- a modern place name if the provenance of the inscription is from a small town or a village (i.e. this element should not be used in the case of places like Plovdiv and Stara Zagora);
- name(s) of attested ancient persons in the monument, especially if it is a honorific or funerary inscription, or, possibly, of the dedicator in the case of dedicatory inscriptions;

- Home
- Apparatus, translations, comm...
- Authority lists
- Encoding personal names
- Encoding the Greek text
- History of the monument
- Identifier and contents of the e...
- Physical description of the mo...

Your dataset is important!

- Creating, curating and managing research data IS research
- Early communication with other researchers
- Methodology

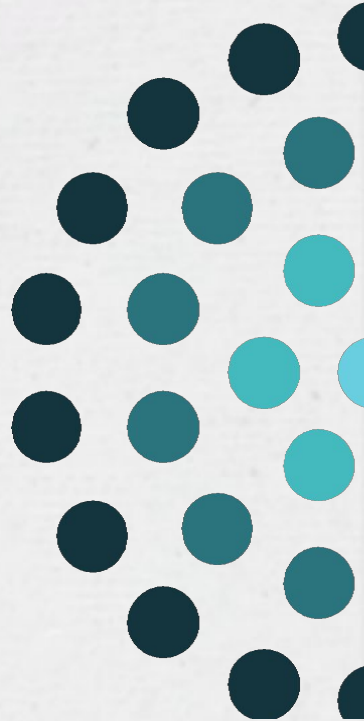
Diversity, inclusiveness and collaboration

- Recognise the diversity of research activities and practices, with a diversity of outputs, and reward early sharing and open collaboration. Consider tasks like peer review, training, mentoring and supervision of Ph.D candidates, leadership roles, and, as appropriate, science communication and interaction with society, entrepreneurship, knowledge valorisation, and industry-academia cooperation. Consider also the full range of research outputs, such as scientific publications, data, software, models, methods, theories, algorithms, protocols, workflows, exhibitions, strategies, policy contributions, etc., and reward research underpinning open science practices such as early knowledge and data sharing a collaboration within science and collaboration with societal actors where Recognise that researchers should not excel in all types of tasks and provide for that allows researchers to contribute to the definition of their research goals an

Tasovac, Romary et al., “The Role of Research Infrastructures in the Research Assessment Reform: A DARIAH Position Paper”: <https://hal.science/hal-04136772v1>

discovery. Activities in and near research infrastructures take different forms of co-creation, such as tool-making, database development, curating datasets and research software, providing training for novel research technologies etc., all of which should be considered in the ongoing attempts to reform research assessment. **Thematic research infrastructures, therefore, provide ideal testbeds for well-balanced, reflexive interventions that can overcome simplistic ideals and models for research evaluation.** While we argue for making invisible work visible and accountable, we are also very much aware that not all that is important can be counted and that not all that can be counted is necessarily important.

CoARA (Coalition for Advancing Research Assessment) Agreement on Reforming Research Assessment:
<https://coara.eu/agreement/the-agreement-full-text/>



In the meanwhile...

- Publish your dataset
- Publish a data paper

scientific **data**

[Explore content](#) ▾

[About the journal](#) ▾

[Publish with us](#) ▾

<https://www.nature.com/sdata/policies/repositories>



<https://www.re3data.org/>



A

<https://github.com/>



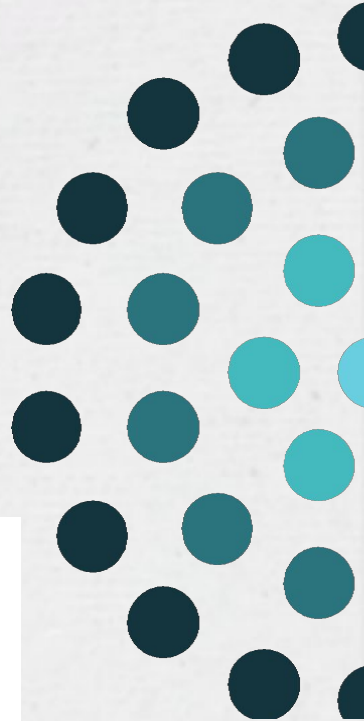
<https://zenodo.org/>



<https://figshare.com/>



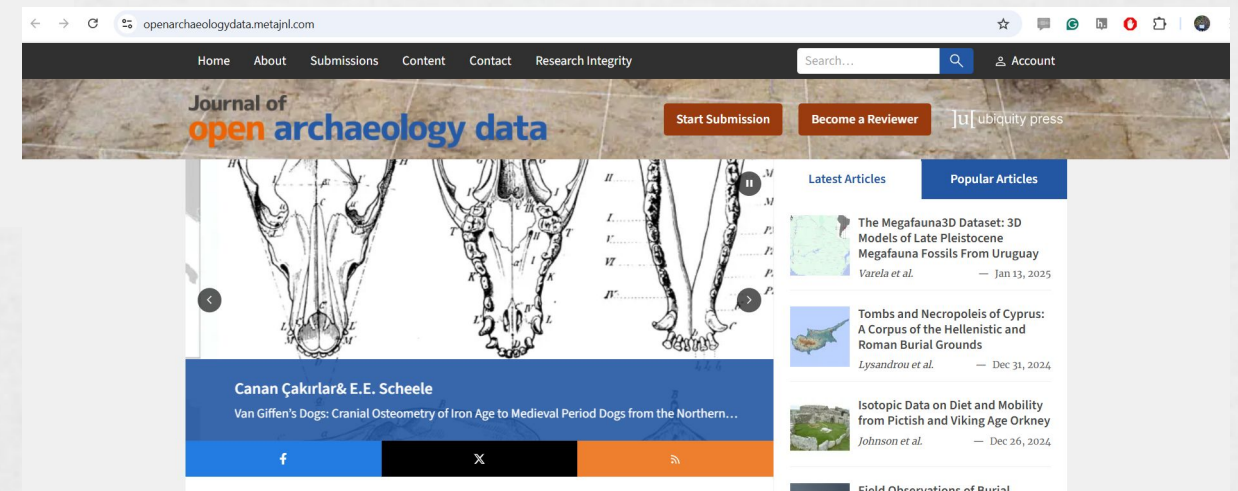
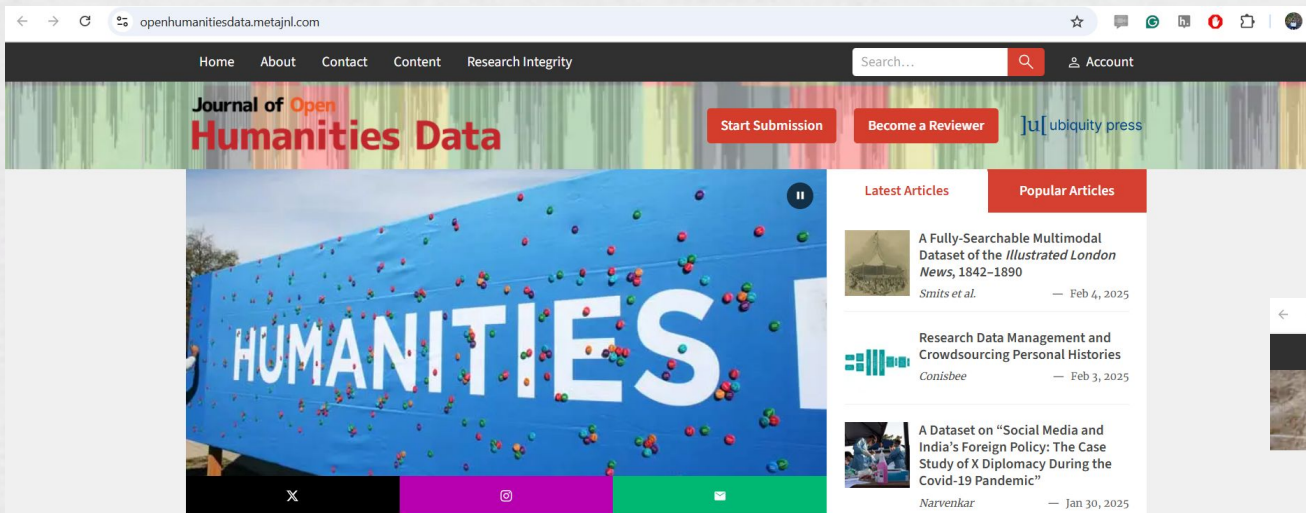
<https://huggingface.co/>



Data papers

Journal of Open Humanities Data
<https://openhumanitiesdata.metajnl.com/>

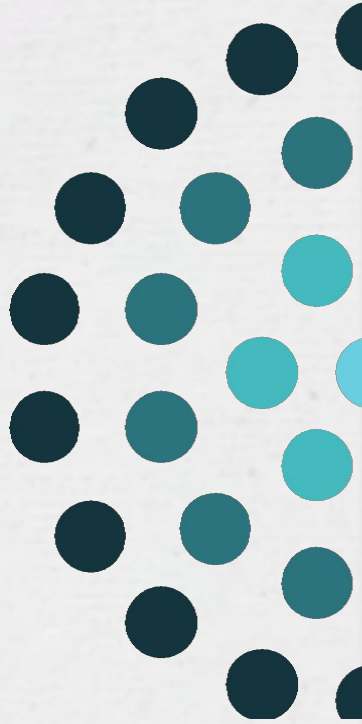
Journal of Open Archaeology Data
<https://openarchaeologydata.metajnl.com/>



Exercise

- Zenodo/Figshare: look for datasets and data papers in your field
- Is there information about:
 - repository location
 - data model, formats, standards
 - data sources and sampling strategy
 - quality control
 - reuse potential
- Describe your own data

A



**БЛАГОДАРЯ ЗА
ВНИМАНИЕТО!**

diilliev@uni-sofia.bg
dimitar.illiev@gmail.com

