

Modelling humanities data with TEI-XML

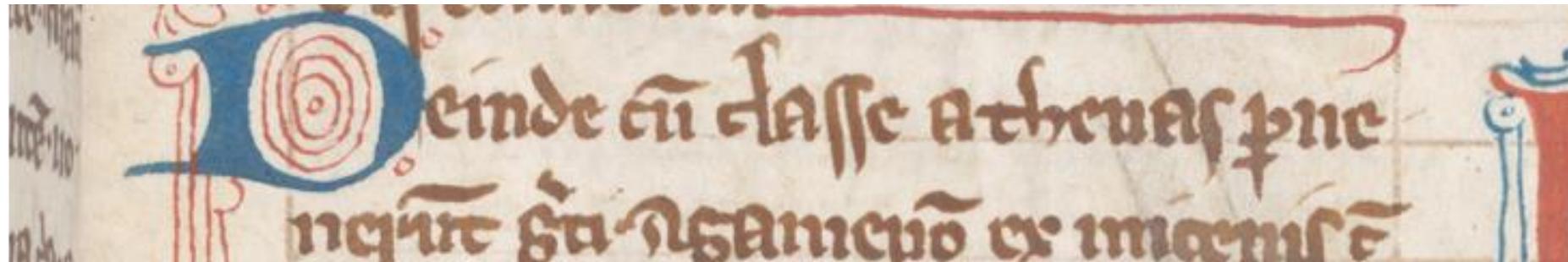
SCHOLARLY EDITING AND MANUSCRIPT CATALOGUING IN THE DIGITAL AGE

Dr Katarzyna Anna Kapitan
12 November 2025

Text Critical Editing

Diplomatic vs. Text-Critical Edition

Then they arrived in Athens with a fleet



TCD 11500: Deinde cu(m) **classe** athenas **p(er)ueneru(n)t**

corpone equali. blandam. affabilem. ueretundam. animo ^{impucem} pia. Deinde
ornati cu classib; athenas conuener². Agam non ex micenis. cum

Cod. Sang. 197: Deinde **ornati** cu(m) **classib(us)** athenas **conuener(unt)**

bonum aduncum oculis rotundis superciliosum. Palamedem gracilem longum sapientem animo magnum blandum. Podalirium crassum valentem superbum tristem. Machaonem fortem magnum certum prudenter patientem misericordem. Merionem rufum mediocri statura corpore rotundo viriosum pertinacem crudelem impatientem. Briseidam formosam non alta statura candidam capillo flavo et molli superciliis iunctis oculis venustis corpore aequali blandam affabilem verecundam animo simplici pian.¹⁰

XIV. Deinde ornati cum classe Graeci Athenas convenerunt: Agamemnon ex Mycenis cum navibus numero C, Menelaus ex Sparta cum navibus numero LX, Arcesilaus et Prothoenor ex Boeotia cum navibus numero L, Ascalaphus et Ialmenus ex Orchomeno cum navibus numero XXX, Epistrophus et Schedius ex Phocide cum navibus numero XL, Ajax Telamonius ex Salamina adduxit secum Teucerum fratrem, ex Buprasione Amphimachum Diorem Thalpium Polyxe-

XIV. Dictys Cretensis I 17.

1 bonum] toruum *O* | fort. aduncum naso 3 Polachrium *G* 7 Bereseidam *G* Breseidam *L* 8 molli et flauo *G* 11 classibus *G* | Graeci add. *A* ante Athenas 12 micenis *LG* | cum et numero fere om. *G*, in eodem numeri litteris scripti sunt 14 archelaus *LG* archesilaus *V* Arcesilaus *Mercerius* | protenor *LG* | boetia *G* 15 Ascolofus *L* Ascalaphus *G* | alimus *L* alumnus *G* almenus *F* corr. in edit. Venet. 1499 | orcomeno *L* orcomeno *G* corr. Mer-

*corporē equali· blandam · affabilē · uerecundā · animo pia · Deinde
ornati cū classib · athenas conuenerunt · Agaménōn ex micenis · cum*

Cod. Sang. 197: Deinde ornati cu(m) **classib(us)** athenas conuener(unt)

Encoding Variants

<listWit> (witness list) lists definitions for all the witnesses referred to by a critical apparatus, optionally grouped hierarchically.

<witness> (witness) contains either a description of a single witness referred to within the critical apparatus, or a list of witnesses which is to be referred to by a single sigil.

@xml:id (identifier) provides a unique identifier for the element bearing the attribute.

Encoding Variants

```
<sourceDesc>
  <listWit>
    <witness xml:id="A">Leipzig, Universitätsbibliothek, Cod. Haen. 3518 </witness>
    <witness xml:id="B">Munich, Bayerische Staatsbibliothek, Clm 305v</witness>
    [...]
  </listWit>
</sourceDesc>
```

msDesc in witness

```
<sourceDesc>
  <listWit>
    <witness xml:id="A">
      <msDesc>
        <msIdentifier>
          <country>Germany</country>
          <settlement>Leipzig</settlement>
          <repository>Universitätsbibliothek</repository>
          <idno>Cod. Haen. 3518</idno>
        </msIdentifier>
        <msContents> [...] </msContents>
      </msDesc>
    </witness>
    [...]
  </listWit>
</sourceDesc>
```

Encoding Variants

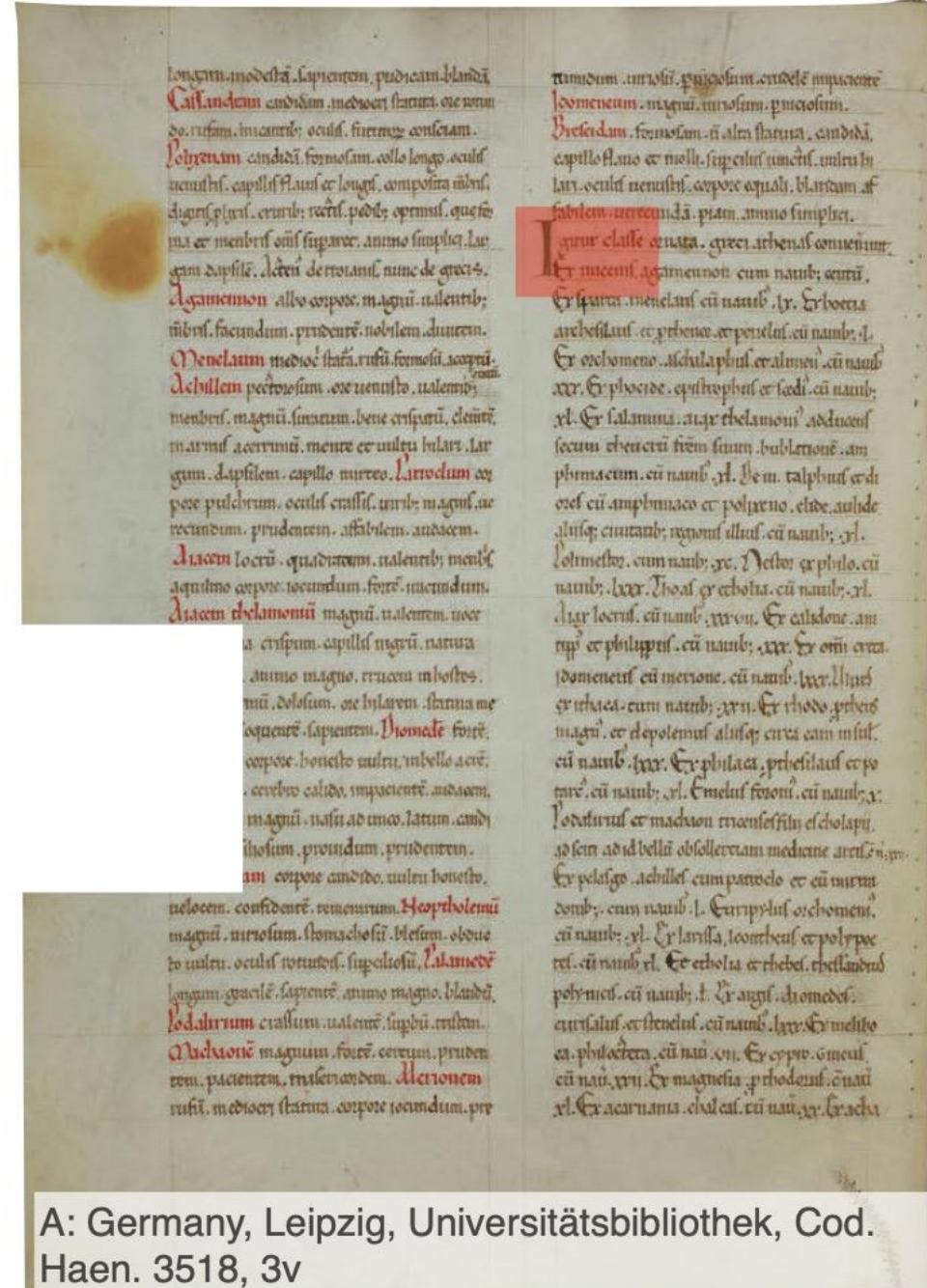
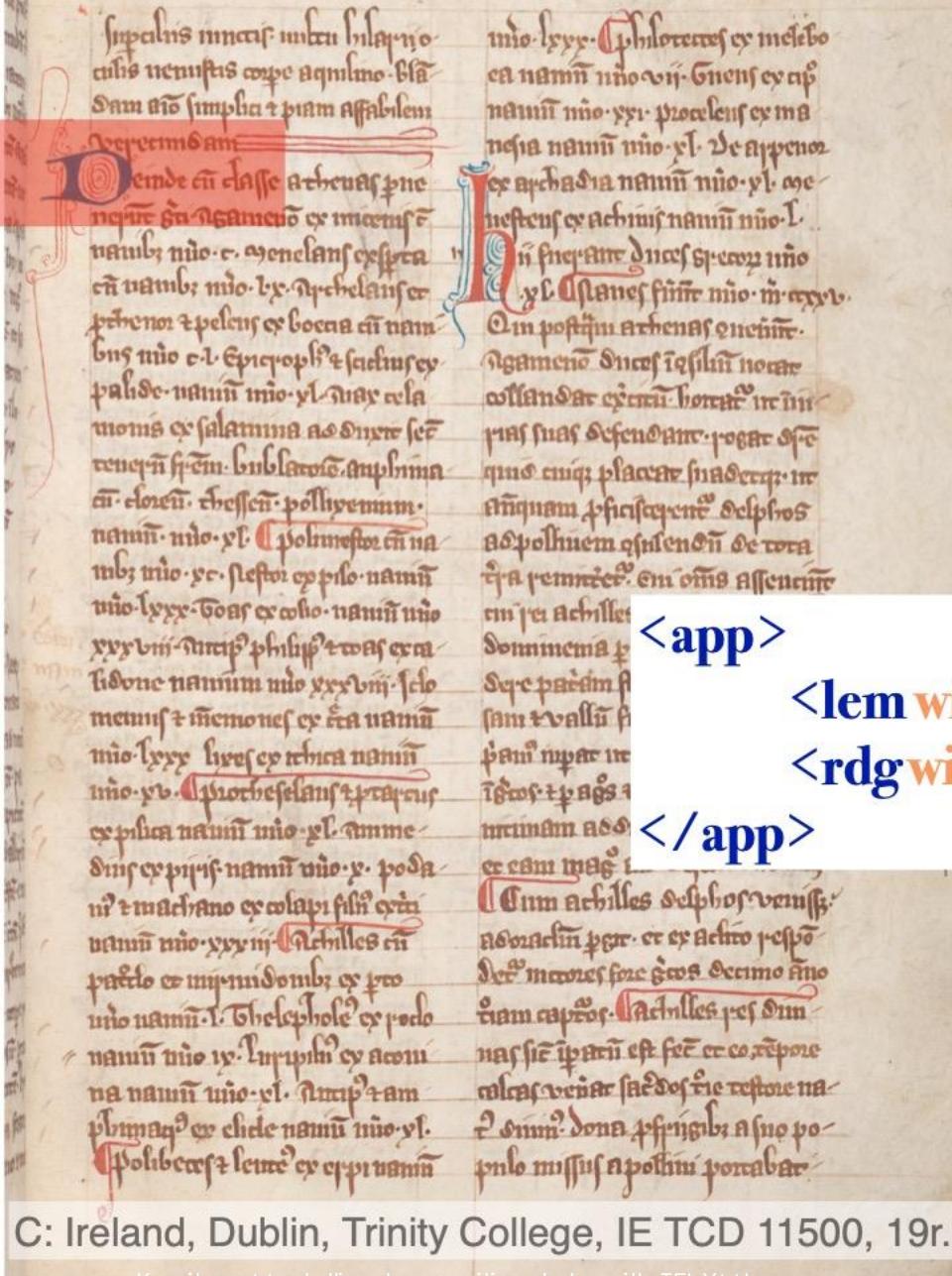
<app> (apparatus entry) contains one entry in a critical apparatus, with an optional lemma and usually one or more readings or notes on the relevant passage.

@type classifies the variation contained in this element according to some convenient typology.

<lem> (lemma) contains the lemma, or base text, of a textual variation

<rdg> (reading) contains a single reading within a textual variation.

@wit (witness or witnesses) contains a space-delimited list of one or more pointers indicating the witnesses which attest to a given reading.



Exercise 1: Editing

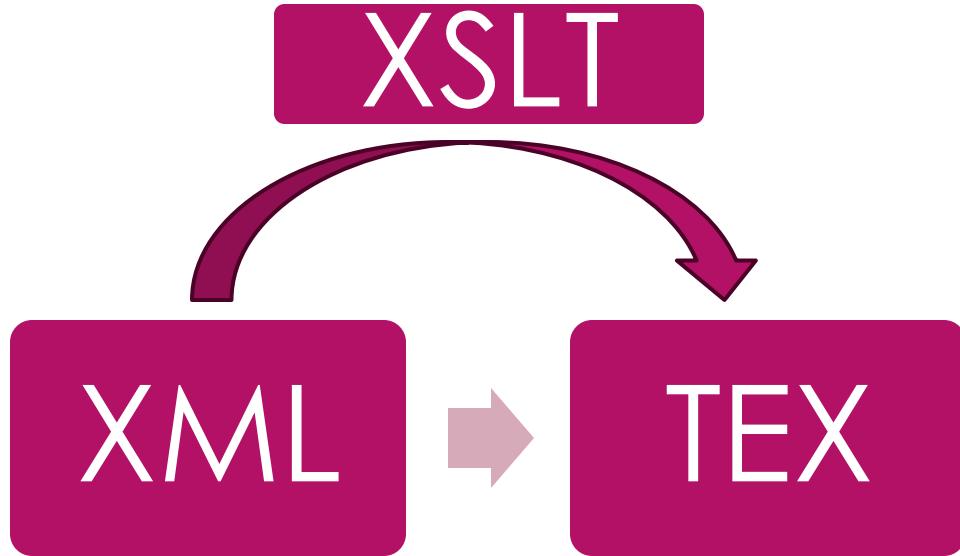
(Group work – 4 people in each group)

- Using Dares_edition_starter.xml prepare an edition of chapter 14 of Dares' *Historia*.
- Your edition will include variants from four manuscripts:
 - 1. Ireland, Dublin, Trinity College, IE TCD 11500
 - 2. Germany, Munich, Bayerische Staatsbibliothek, Clm 305
 - 3. Switzerland, St Gall, Stiftsbibliothek, Cod. Sang. 197
 - 4. Germany, Leipzig, Universitätsbibliothek, Cod. Haen. 3518
- Encode msDesc info about each witness in the externally stored list of witnesses (listWit_starter.xml)
- Entry for each witness should contain msIdentifier and msContents with at least one msItem that gives the locus of Dares' *Historia* (use your Leipzig msDesc for inspiration)

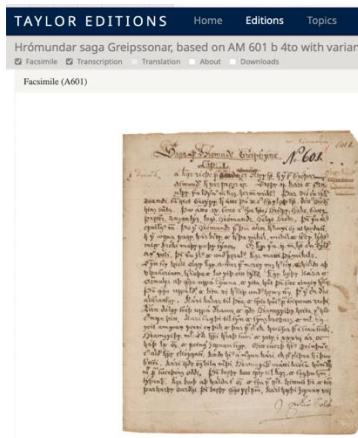
Exercise 1: Transformations

- ▶ Transform your XML encoding into HTML (with default TEI scenarios)
- ▶ Transform your XML encoding into TEX (with a custom XSLT)
 - ▶ Use the XSLT file **transform_ed_with_app.xsl** [save as txt]
 - ▶ This produces a TeX annotation according to the Reledmac guidelines (<https://ctan.org/pkg/reledmac?lang=en>)
 - ▶ How to: **Guidelines_Transformation_Scenario.pdf**

XML to HTML, TeX, etc. (with XSLT)



Kapitan, Modelling humanities data with TEI-XML



Saga af Hrómu(n)de Greipszýne

Cäpituli .I.

be fyrir {Gordom} ¶¶¶ ¶damorc¶¶¶ er Olafr hiet,
Onópar Astmundar, hann var frøgr mabr. Broþr
ulfr, voru landvarnar menn kongz, hermann mikler.
þukur Boandi, sá hiet Greýpr. hann átte þá kono, er
dóttær Hrókz hinz suarta, Þau átto .ix. soño er sva
Hake, Gautr(,) Þrostr, Angantýr, Logi, Hrómundr.
Þeir voru aller efnileger menn. Þo var Hrómundr
m. hann kunni eigi at hrödast, hann var augna fagr,
er þamkíll, mikill oc sterri, lyktiz mið Hróki móþr
Med kongi voru .ij. menn, hiet eirn Byldur, annar
Iller oc underfóruuler. kongr matti Þá mikils. Eytt
r kongr, austur fyrir noreg med her sinn, oc hielldo
i, heriðo, oc lau við eitt Éyland. Kongr býr Kára
ganga uppá Éyuna, oc vita, huort þeir see einginn
engu uppá landib, oc litu .vj. herskip under hölmrum
r eirn dreke allskrautligr. Kári kallar til þeira, oc spir
huorier tyrir skiponom riebi, Eirn dólgr stóð uppá drekanom, oc
quadz Hraungvíþr heita. eþr huort er nafn þitt. Kari sagbi til sýn
oc sýnz broþurz. oc mælti. Eg veit aungvann verri enn þik, oc þar
fyrir skal ek hauggva þic i smá sticki, Hraungvíþr. mælti: Ek hefi

Editions of Hrómundar saga

Source: https://editions.mml.ox.ac.uk/editions/hromundar_A601/ (left)
<https://ora.ox.ac.uk/objects/uuid:403f3a92-b92c-4d07-8945-d74513c95cac> (right)

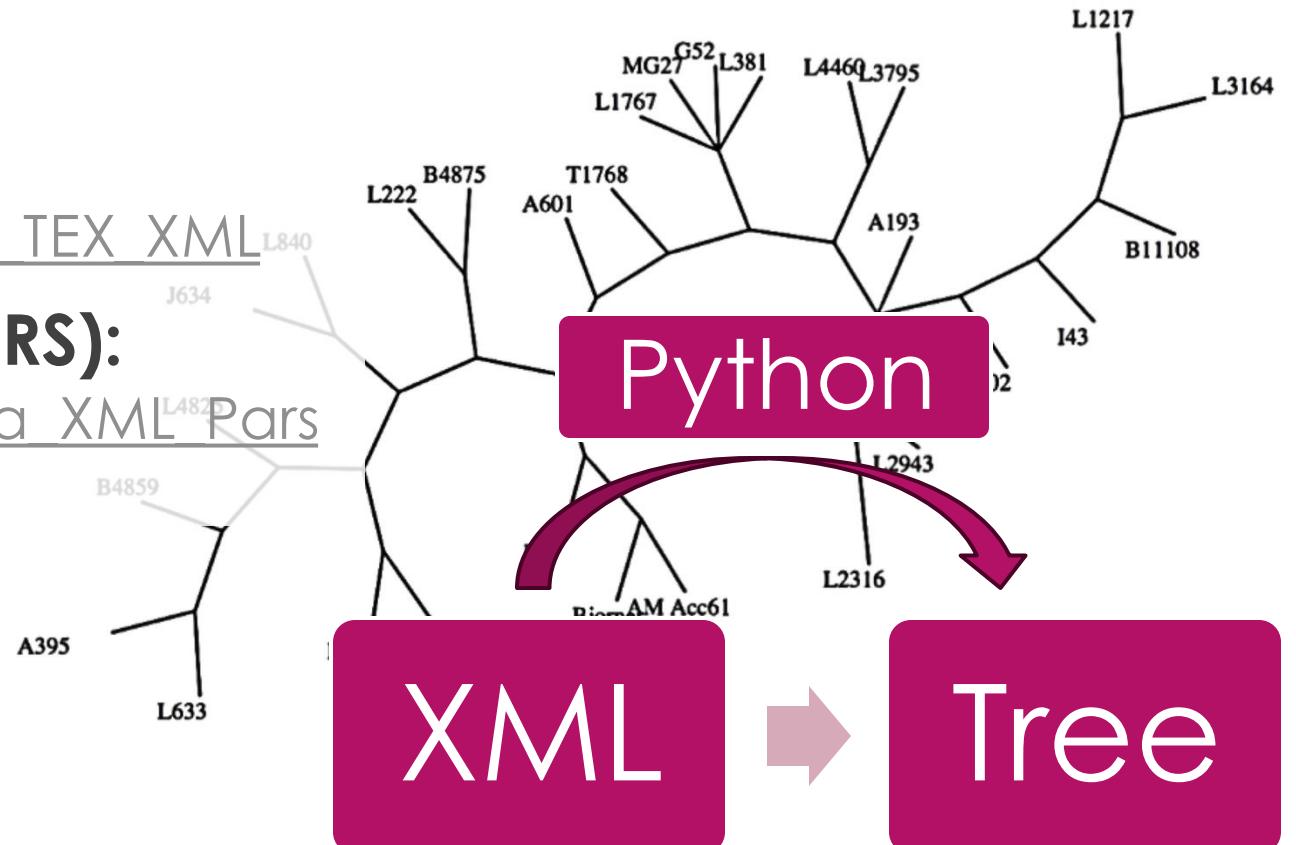
XML to Tree, TeX to XML etc (with Python)

► **TEX to XML (with Python):**

<https://github.com/KAKDH/edition> **TEX XML**

► **XML to Tree (with Python & PARS):**

<https://github.com/KAKDH/stemma> **XML Pars**



Annotations

Kapitan, Modelling humanities data with TEI-XML

For Linguistic Analysis: Part of Speech, Morpho-Syntactic Annotations...

- ▶ **Chapter 18:** Simple Analytic Mechanisms
 - ▶ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html>
- ▶ **<s>** (s-unit) contains a sentence-like division of a text.
- ▶ **<cl>** (clause) represents a grammatical clause.
- ▶ **<phr>** (phrase) represents a grammatical phrase.
- ▶ **<w>** (word) represents a grammatical word.

@lemma

provides a lemma (base form) for the word, typically uninflected and serving both as an identifier (e.g. in dictionary contexts, as a headword), and as a basis for potential inflections.

@lemmaRef

provides a pointer to a definition of the lemma for the word, for example in an online lexicon.

```
<s xml:lang="la">
  <w lemma="timeo">timeo</w>
  <w lemma="danaii">Danaos</w>
  <w lemma="et">et</w>
  <w lemma="donum">dona</w>
  <w lemma="fero">ferentes</w>
</s>
```

```
<s xml:lang="la">
  <w lemmaRef="http://lexicon.org/latin.xml#timeo">timeo</w>
  <w lemmaRef="http://lexicon.org/latin.xml#danaii">Danaos</w>
  <!-- ... -->
</s>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-s.html>

@pos (part of speech) indicates the part of speech assigned to a token (i.e. information on whether it is a noun, adjective, or verb), usually according to some official reference vocabulary.

Examples of vocabularies: for German STTA (Stuttgart-Tübingen Tagset); for English: CLAWS (The Constituent Likelihood Automatic Word-tagging System); For French: French TreeTagger Part-of-Speech Tags.

The victim's friends told police that
Kruger drove into the quarry
and never surfaced

```
<s><w pos="AT0">The</w> <w pos="NN1">victim</w><w pos="POS">'s</w> <w pos="NN2">friends</w>  
<w pos="VVD">told</w> <w pos="NN2">police</w> <w pos="CJT">that</w> <w pos="NPO">Kruger</w>  
<w pos="VVD">drove</w> <w pos="PRP">into</w> <w pos="AT0">the</w> <w pos="NN1">quarry</w>  
<w pos="CJC">and</w> <w pos="AV0">never</w> <w pos="SENT">surfaced</w><pc pos="PUN">. </pc></s>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-s.html>

```
<s type="sentence">
<phr ana="#n">
<phr ana="#gn">
<w ana="#AT0">The</w>
<w ana="#NN1">victim</w>
<m ana="#POS">'s</m>
</phr>
<w ana="#NN2">friends</w>
</phr>
<phr ana="#v">
<w ana="#VVD">told</w>
<phr ana="#n">
<w ana="#NN2">police</w>
</phr>
<cl ana="#fn">
<w ana="#CJT">that</w>
<phr ana="#n">
<w ana="#NP0">Krueger</w>
</phr>
<phr ana="#v">
<phr ana="#v1">
<w ana="#VVD">drove</w>
<phr ana="#pr">
<w ana="#PRP">into</w>
<phr ana="#n">
<w ana="#AT0">the</w>
<w ana="#NN1">quarry</w>
</phr>
</phr>
</phr>
<w ana="#CJC">and</w>
<phr ana="#v2">
<w ana="#AV0">never</w>
<w ana="#VVD">surfaced</w>
</phr>
</phr>
</cl>
</phr>
<c ana="#pun">.</c>
</s>
```

The victim's friends
told police
that Kruger drove into the quarry
and
never surfaced

More linguistic attributes in the TEI Guidelines:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.linguistic.html>

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-s.html>

For Personography:

► Chapter 14: Names, Dates, People, and Places

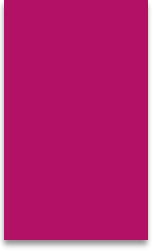
► <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>

<persName> (personal name) contains a proper noun or proper-noun phrase referring to a person, possibly including one or more of the person's forenames, surnames, honorifics, added names, etc.

<surname> (surname) contains a family (inherited) name, as opposed to a given, baptismal, or nick name.

<forename> (forename) contains a forename, given or baptismal name.

```
<persName>
  <forename type="first">Franklin</forename>
  <forename type="middle">Delano</forename>
  <surname>Roosevelt</surname>
</persName>
<persName>
  <forename type="given">Margaret</forename>
  <forename type="unused">Hilda</forename>
  <surname type="birth">Roberts</surname>
  <surname type="married">Thatcher</surname>
</persName>
```



<listPerson> (list of persons) contains a list of descriptions, each of which provides information about an identifiable person or a group of people

<person> (person) provides information about an identifiable individual, for example a participant in a language interaction, or a person referred to in a historical source.

```
<person>
  <persName>
    <forename>Miss Major</forename>
    <surname>Griffin-Gracy</surname>
  </persName>
  <birth when="1940-10-25"/>
  <gender value="transWoman"
    evidence="selfIdentification">trans woman</gender>
  <persPronouns value="she"
    evidence="selfIdentification"/>
  <note>
    <p>Veteran of the Stonewall Riots. Founder of the
      Griffin-Gracy Educational Retreat and Historical
      Center (the House of GG). Activist and advocate for
      transgender and gender-nonconforming people of
      color.</p>
  </note>
</person>
```

@xml:id (identifier) provides a unique identifier for the element bearing the attribute.

@ref (reference) provides an explicit means of locating a full definition or identity for the entity being named by means of one or more URLs.

```
<person xml:id="ArnMag">
  <persName xml:lang="is">Árni Magnússon</persName>
  <persName xml:lang="da">Arne Magnusson</persName>
  <persName xml:lang="la">Arnas Magnæus</persName>
</person>
```

In 1702 <persName ref="#ArnMag">Árni</name> wrote
in one of his letters to his friend...

Dares, ch. 14:

Deinde ornati cum classe Graeci Athenas convenerunt:

Agamemnon ex Mycenis...

```
20
21 <listPerson>
22   <person xml:id="Aga_001">
23     <persName>Agamemnon</persName>
24     <idno type="wikidata">Q128176</idno>
25     <note>King of Mycenae; son of Atreus in Greek mythology.</note>
26   </person>
27 </listPerson>
28 </sourceDesc>
29 </fileDesc>
30 </teiHeader>
31 <text>
32   <body>
33     <div type="chapter" n="14">
34       <p>
35         <s>Deinde ornati cum classe Graeci Athenas convenerunt:</s>
36         <s><persName ref="#Aga_001">Agamemnon</persName> ex Mycenis cum navibus numero
            type="ship" value="100">C</s><num> 100 </s> Menelaus ex Sparta cum navibus numero
```



Item Discussion

Read View history

Search Wikidata

Agamemnon (Q128176)

son of Atreus in Greek mythology

▼ In more languages

Configure

Language	Label	Description	Also known as
English	Agamemnon	son of Atreus in Greek mythology	
French	Agamemnon	personnage de la mythologie grecque	
Spanish	Agamenón	héroe de la Grecia antigua	Agamemnon Agamenon
German	Agamemnon	König der griechischen Mythologie	

All entered languages

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data

Create a new Lexeme
Recent changes
Random Lexeme

Tools

What links here
Related changes
Special pages
Permanent link

Statements

instance of mythological Greek character

edit

View data with TEI-XML

0 references

Search

Select Field:

Personal Names

Select Index:

All VIAF

Search Terms:

Agamemnon

Search

Agamemnon, mythologie grecque 

VIAF ID: 173096818 (Personal)

Permalink: <http://viaf.org/viaf/173096818>

Preferred Forms

 100 0 _ [‡a Agamemnon, ‡c mythologie grecque](#)



For Spatial Analysis:

- ▶ **Chapter 14:** Names, Dates, People, and Places
 - ▶ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>
- <placeName>** contains an absolute or relative place name.

```
<placeName>
  <settlement>Rochester</settlement>
  <region>New York</region>
</placeName>
```



<listPlace> (list of places) contains a list of places, optionally followed by a list of relationships (other than containment) defined amongst them.

<place> (place) contains data about a geographic location

<location> defines the location of a place as a set of geographical coordinates, in terms of other named geo-political entities, or as an address.

<geo> (geographical coordinates) contains any expression of a set of geographic coordinates, representing a point, line, or area on the surface of the earth in some notation.

Dares, ch. 14:

Deinde ornati cum classe Graeci **Athenas** convenerunt:
Agamemnon...

```
27 </listPerson>
28 <listPlace>
29   <place xml:id="Athe_001">
30     <placeName>Athens</placeName>
31     <location>
32       <geo>37.9839 23.7278</geo>
33       <!-- lat long -->
34     </location>
35     <idno type="wikidata">Q1524</idno>
36   </place>
37 </listPlace>
38 </sourceDesc>
39 </fileDesc>
40 </teiHeader>
41 <text>
42   <body>
43     <div type="chapter" n="14">
44       <p>
45         <s>Deinde ornati cum classe Graeci <placeName ref="#Athe_001">Athens</placeName> convenerunt:</s>
```



The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge.

Athens [advanced search]

all countries

enter a location name, ex: "Paris", "Mount Everest", "New York"

Browse the names

- [Countries](#)
- [Postal codes](#)
- [Country statistics](#)
- [Recent modifications](#)

Information

- [About GeoNames](#)
- [Data Sources](#)
- [User manual](#)
- [Ambassadors and Team](#)
- [Forum](#)
- [Blog](#)
- [Mailing list](#)
- [Commercial Support and Consulting](#)

Download

- [Info](#)
- [Free Gazetteer Data](#)
- [Free Postal Code Data](#)
- [Premium Data](#)

Web Services

- [Overview](#)
- [Documentation](#)
- [Client Libraries](#)
- [Premium Web Services](#)

Sponsoring

Kapitan, Modelling humanities data with IEI-XML

Search Tools & Databases

- Library Catalog
- ▶ Getty Research Portal
- Collection Inventories & Finding Aids
- ▶ Photo Archive
- ▶ Research Guides & Bibliographies
- Digital Collections
- ▶ Article & Research Databases
- ▶ Collecting & Provenance Research
- BHA & RILA
- ▼ Getty Vocabularies
 - ▶ Art & Architecture Thesaurus (AAT) ®
 - ▶ Cultural Objects Name Authority (CONA) ®
 - ▼ Getty Thesaurus of Geographic Names (TGN)
®
 - About TGN

Getty Thesaurus of Geographic Names® Online

Search the TGN

[? Help](#)[Search](#)[Clear](#)[Pop-up Search](#)[Browse the TGN hierarchies](#)

Copyright information

Search Tips

For the Find Name field, you may use **AND** and **OR** [e.g., 1) san carlos, 2) carlos OR charles, 3) carl* OR charl*, 4) san AND carlos, 5) carlos AND (san OR saint), 6) (carlos OR charles) AND (san OR saint)] Boolean operators must be in all caps (AND and OR). Wildcard is the asterisk (*); right truncation only. To find an exact match rather than a key word, use quotes

Have a Question?

- ✉ Contact the Vocabulary Program

More Applications



Drama:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DR.html>

<https://dracor.org/doc/odd>



Letters:

<https://encoding-correspondence.bbaw.de/v1/index.html>



Language corpora:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/CC.html>



Dictionaries:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>



Music:

<https://music-encoding.org>

Exercise 2

- ▶ Download the XML file **Dares_Historia_Chapter_14_annotations_starter.xml**
- ▶ Divide the document into sentences **<s> [important for postprocessing]!**
- ▶ Start annotating **names** and **numbers** following the examples provided in the document, using **<persName>**, **<placeName>** and **<num>** for numbers of ships.
- ▶ **Remember to add new people and places to your Source Description and give them unique identifiers (xml:id)!**
- ▶ Transform your XML to HTML and CSV using the two XSLTs provided.

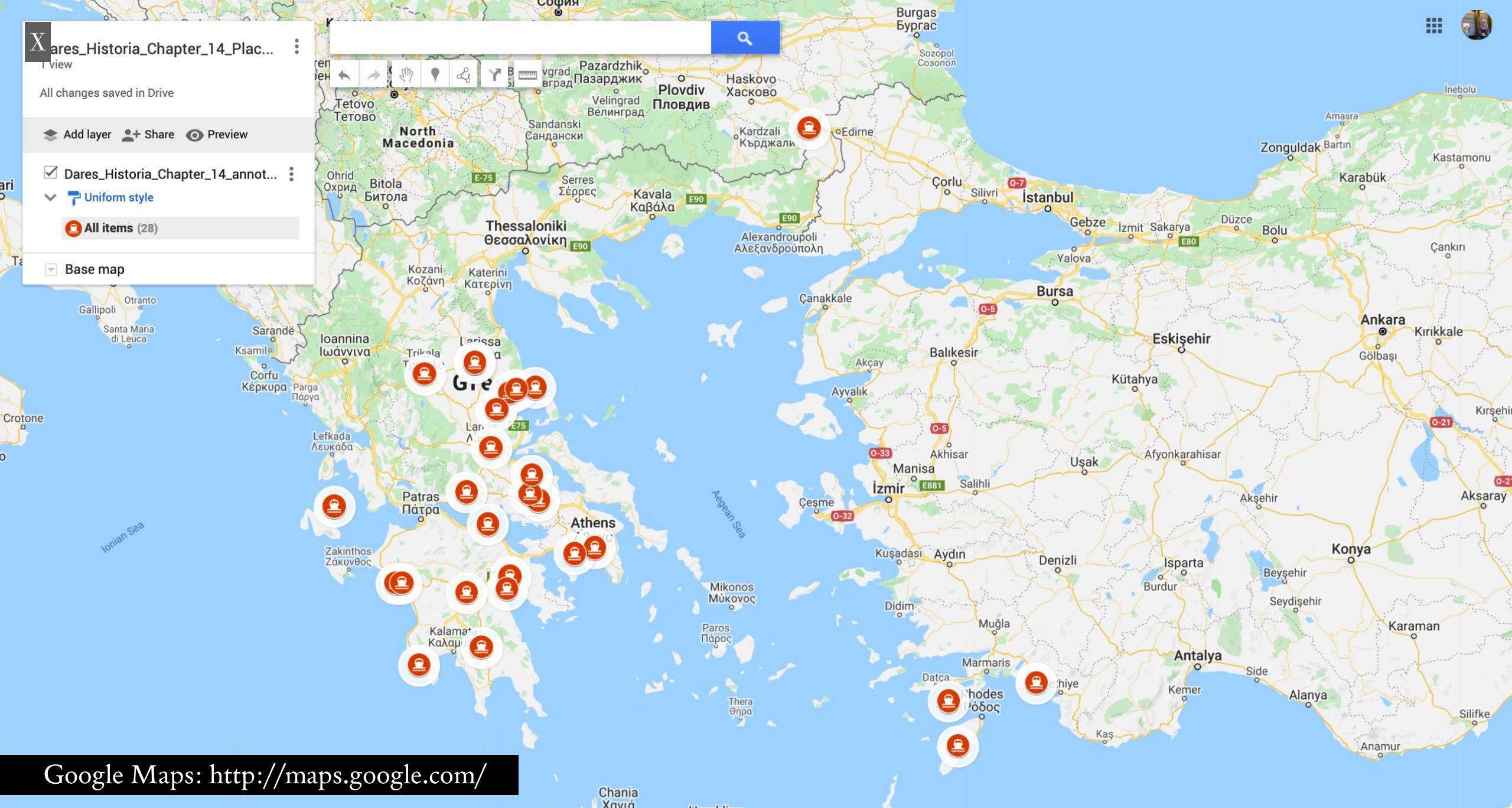
blesum. uultu bonū. aduncū. oculis rotundis. superciliosum; Palamedem.
 gracile. longū. sapientem. animo magnū. blandū. Podaliriu^r. crassū. valentē.
 sup bū. tristem. Machaonē. fortam. magnū. certū. prudentē. pacientē. misericordem.
 Merionē. rufū. mediocrem statuta. corpore retundo. uirosū. prae
 nacem. crudelē. in patientem. Beresidam. formosā. non alta statuta.
 candidā. capillo molli et flauo. superciliosus uulstis. oculis venustis.
 corpore equali. blanda m. affabilē. ueretundā. animo pia. Deinde
 ornati cū classib. athenas conuener^r. Agamīnon ex micenis. cum
 nauib. numero centū. Menelaus ex sparta; nauib. numero sexaginta.
 Archelaus ex pentē. ex boetia nauib. numero L. Ascalaphus et aliū
 nus ex oreomeno nauib. numero xxx. Epistrophus et seadius ex
 phodicenno. nauib. numero xl. Ajax telamonius. ex salamina. ad
 duū secum teverū. frēm bubationē. amphimacū. dorū. thesum.
 polixenū. nauib. numero xl. Nestor ex pilo. nauib. numero octoginta.
 Thoas ex etholia. nauib. numero xl. Uenerius ex imi. nauib. numero
 quinquaginta trib. Ajax oileus ex locris. nauib. numero xxxvii.
 Antiphus. philipp^r. thoas ex caledonie. nauib. numero xxx. Idomeneus.
 et meriones. ex creta. nauib. numero octoginta. Ulixes ex itaca.
 nauib. numero xl. Eumeus ex pirgis. nauib. numero x. Procteus.
 et potareus. ex pilaca. nauib. numero xl. Podalirius et maachon.
 ex colaphisi iuxtraci. nauib. numero xxx iiib. Achilles cū patroclo.
 simul ex myrmidoniib. ex tharkia. Telepolēm

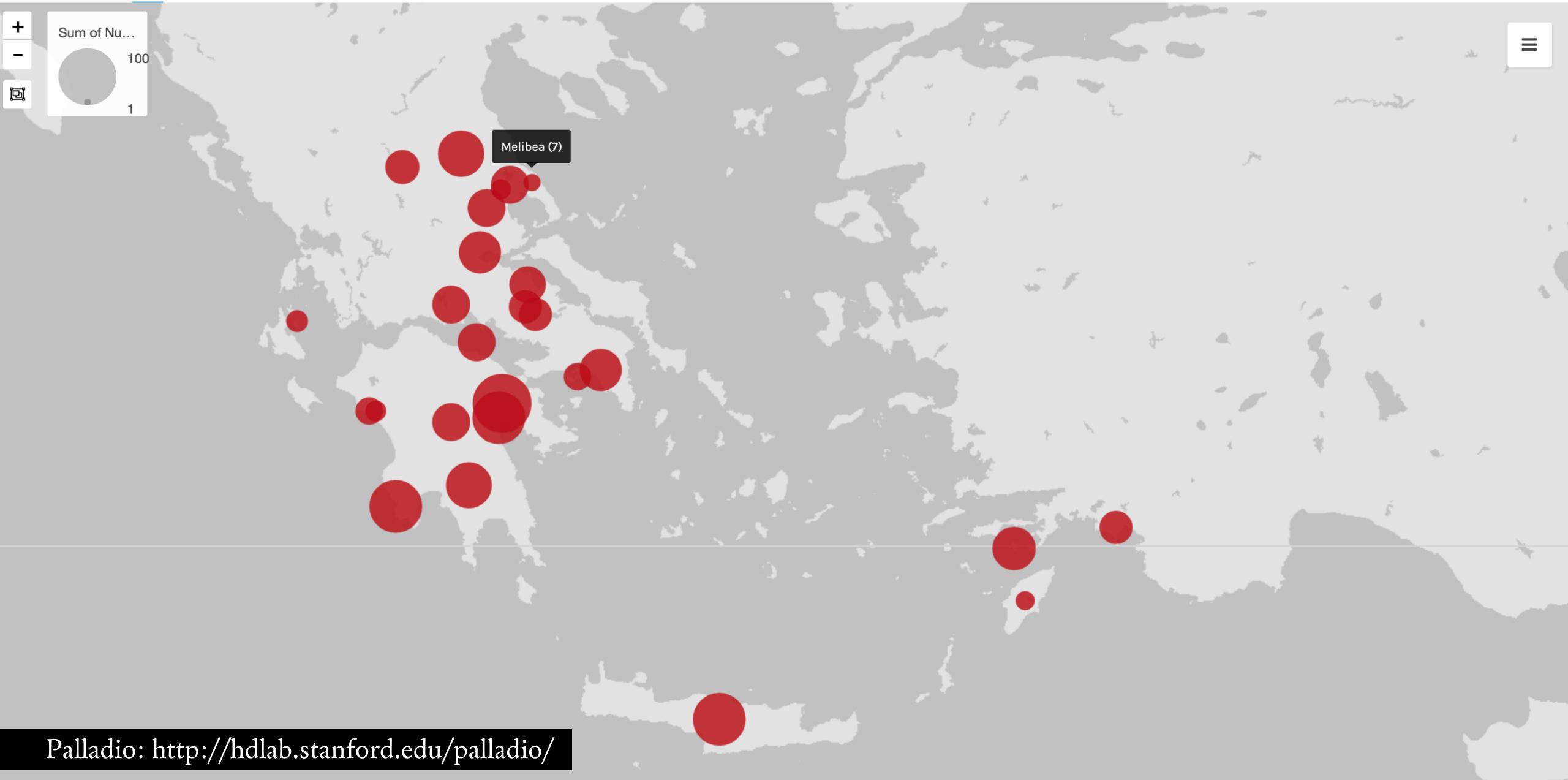
Switzerland, St Gall, Stiftsbibliothek, Cod. Sang. 197, p. 102 (detail).

There were 44 leaders sailing from Greece to Troy. Their names are: Agamemnon Amphimachus Diores Thalpius Polyxenus Nestor Thoas Nierus Aiax Oileus Antip Tlepolemus Eurypylus Antiphus Amphimachus Polypoetes Leonteus Diomedes Ei

Their names in alphabetic order are:

- Achilles
- Agamemnon
- Agapenor
- Aiax Oileus
- Aiax Telamonius
- Amphimachus
- Amphimachus
- Antiphus
- Antiphus
- Arcesilaus
- Ascalaphus
- Diomedes
- Diores
- Epistrophus
- Eumelus
- Eryalus
- Eurypylus
- Guneus
- Ialmenusus
- Idomeneus
- Leonteus
- Machaon
- Menelaus
- Menestheus
- Meriones
- Nestor
- Nierus
- Patroclo
- Phidippus
- Philoctetes
- Podacres





Palladio: <http://hdlab.stanford.edu/palladio/>