# DIGITAL RENAISSANCE

## OLD CULTURES, NEW METHODS

Dr Thomas Köntges / @ThomasKoentges

Digital Humanities, University of Leipzig

# STRUCTURE OF THE TALK

1. Digital Renaissance – What does this mean?
2. Prof Crane and his team
3. Perseus Digital Library and OPP
4. Showcasing some tools and methods
5. Questions

# RENAISSANCE

- Carolingian Renaissance
- Italian Renaissance
- Digital Renaissance

  Access to literature, education and reuse of knowledge

# DIGITAL RENAISSANCE

- Increase proficiency in historical languages among as many people as possible
- Involving as many people as possible in the creation and restoration of knowledge
- Distribute restored/reconstructed and newly generated knowledge as widely as possible

# DIGITAL RENAISSANCE

J. McGann (TLS, 20 November 2009):

---

*“In the next fifty years the entirety of our inherited archive of cultural works will have to be re-edited within a network of digital storage, access, and dissemination. This system, which is already under development, is transnational and transcultural.”*

---

# PROF GREGORY CRANE

- PhD in Classics (Harvard)
- Editor-in-Chief Perseus Digital Library (Tufts)
- Alexander-von-Humboldt Professor (Leipzig)

# HIS TEAM

- Leipzig Team
- Perseus Team

# PERSEUS DIGITAL LIBRARY AND OPP

PERSEUS DIGITAL LIBRARY

# OPEN PHILOLOGY PROJECT (OPP)

# SHOWCASING SOME TOOLS AND METHODS

# TEXTUAL DATA VS. ITS VISUALISATION

```
<tei.2>
<text>

<div1 type="Book" n="1" org="uniform" sample="complete">
<p>
<milestone n="1" unit="section">
Gallia est omnis divisa in partes tres, quarum unam incolunt Belgae, alia
</milestone></p>
</div1>

</text>
</tei.2>
```

# ROBUST PERMANENT CITATIONS

```
urn:cts:greekLit:tlg0012.tlg001:1.10-1.20@οὐλομένην
urn:cite:hmt:msB.1r
urn:cite:hmt:chsimg.VA012RN-0013:0.0513,0.2216,0.12,0.0883
```

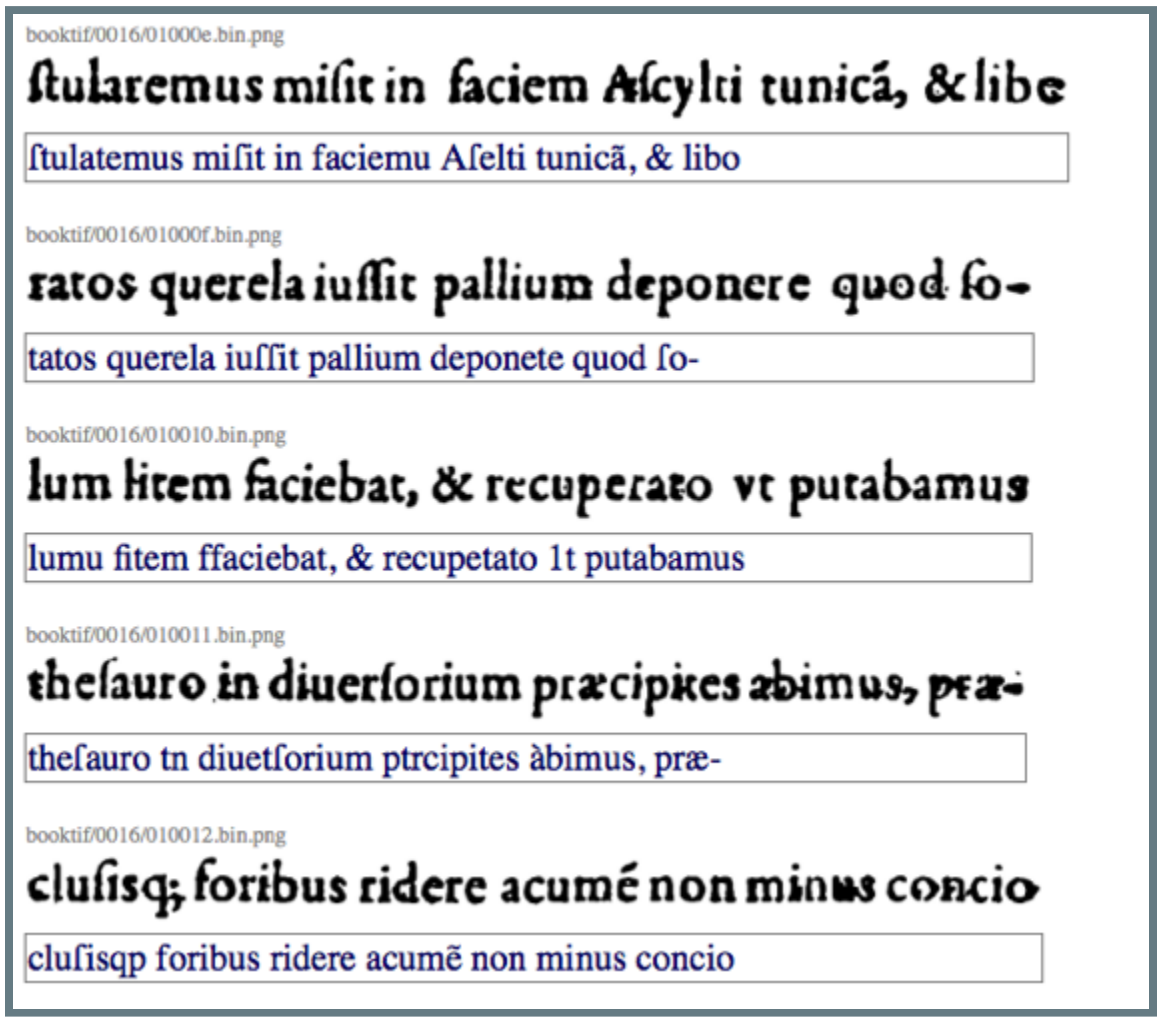
Using CTS/CITE URNs and a Catalogue

- Perseus Catalogue

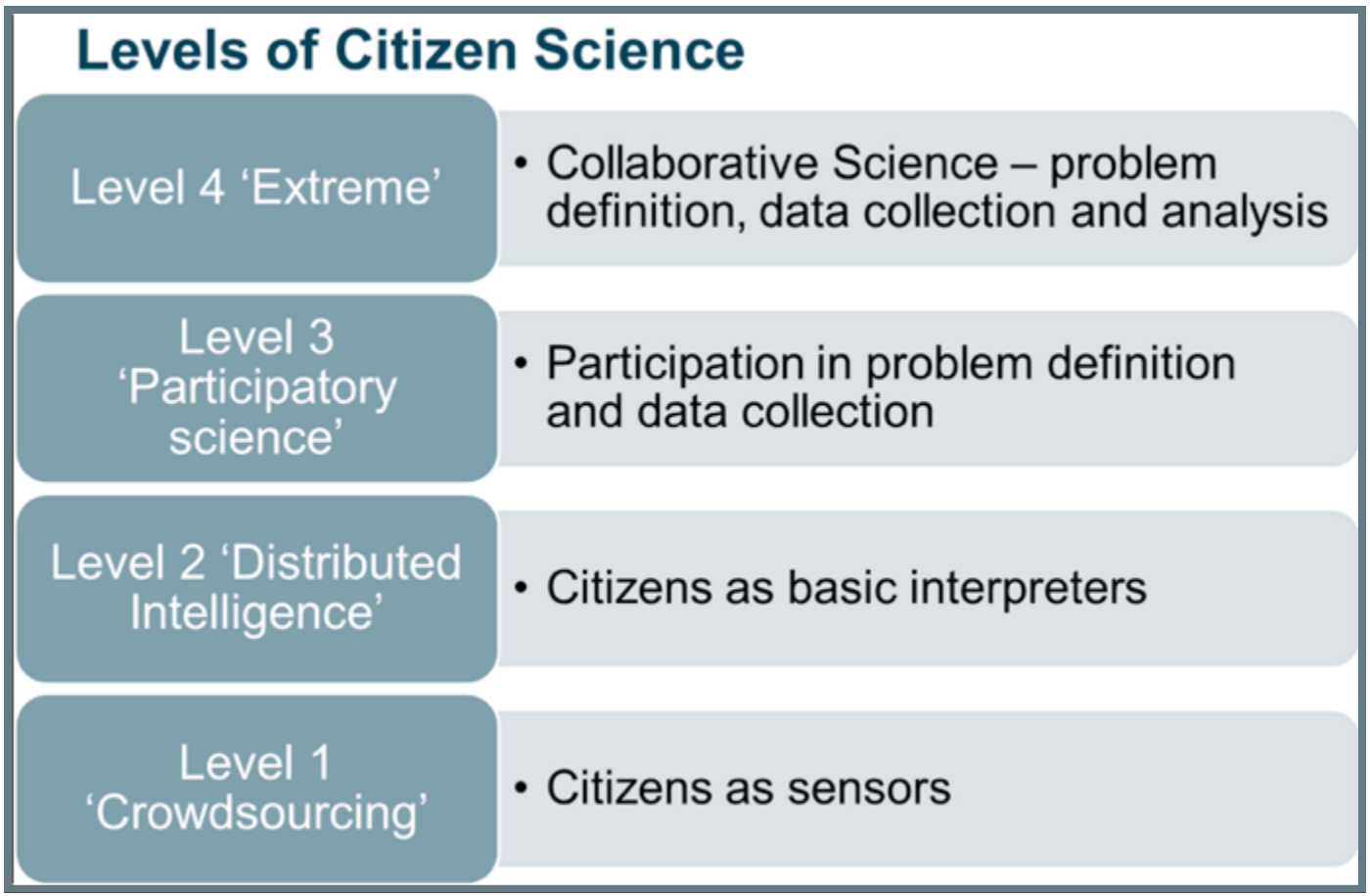
# MAKING TEXTS OPENLY ACCESSIBLE

- Perseus Digital Library
- OPP Github Repositories
- Third Party Contributions

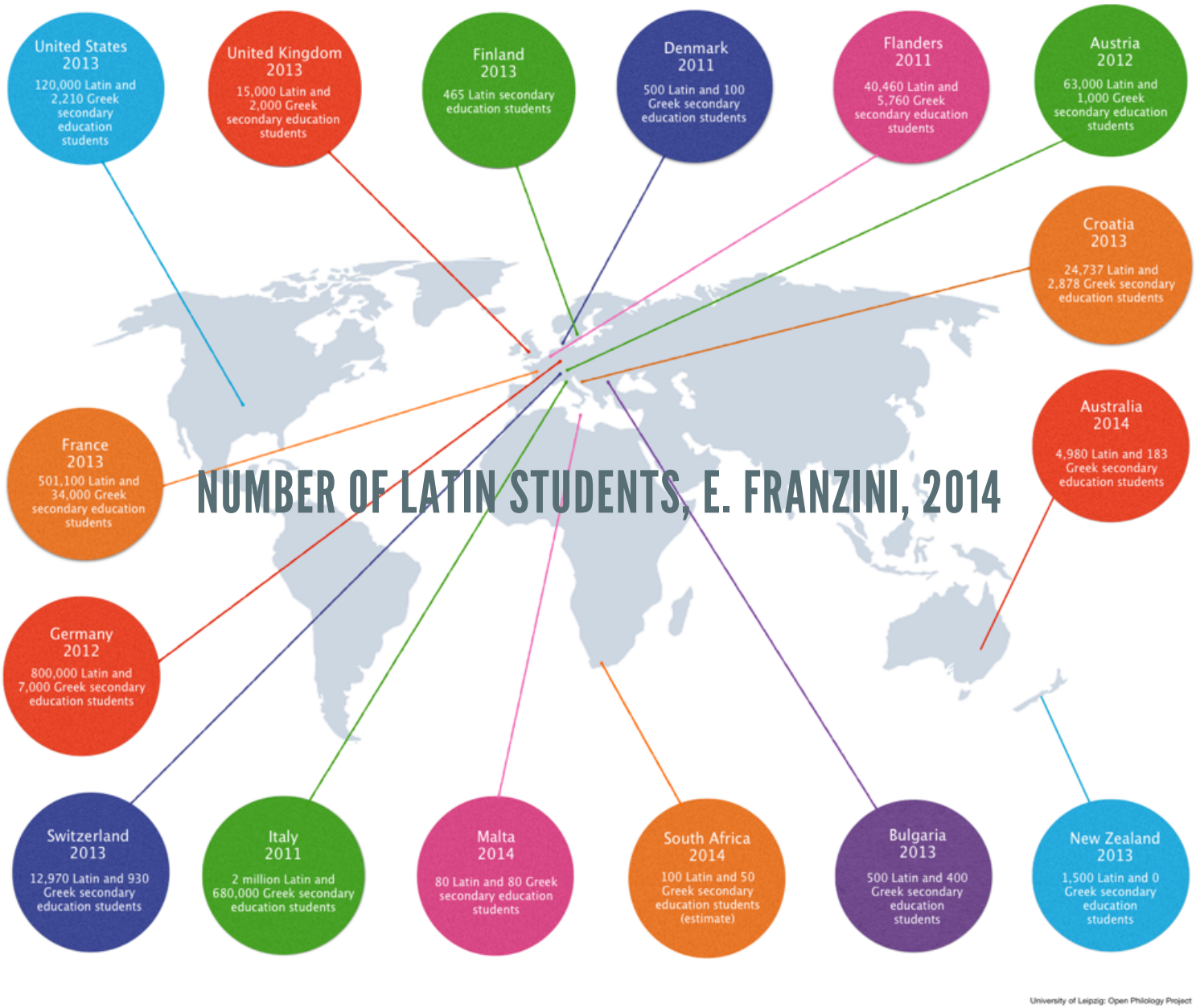
# OCR


booktif/0016/01000e.bin.png
ſtularemus miſit in faciem Aſcylti tunicá, & libe
ſtulatemus miſit in faciemu Aſelti tunicã, & libo
booktif/0016/01000f.bin.png
ratos querela iuſſit pallium deponere quod ſo-
tatos querela iuſſit pallium deponete quod ſo-
booktif/0016/010010.bin.png
lum litem faciebat, & recuperato vt putabamus
lumu fitem ffaciebat, & recupetato 1t putabamus
booktif/0016/010011.bin.png
theſauro in diuerſorium præcipites abimus, præ-
theſauro tn diuetſorium ptrcipites àbimus, præ-
booktif/0016/010012.bin.png
cluſisq; foribus ridere acumé non minus concio
cluſisqp foribus ridere acumẽ non minus concio

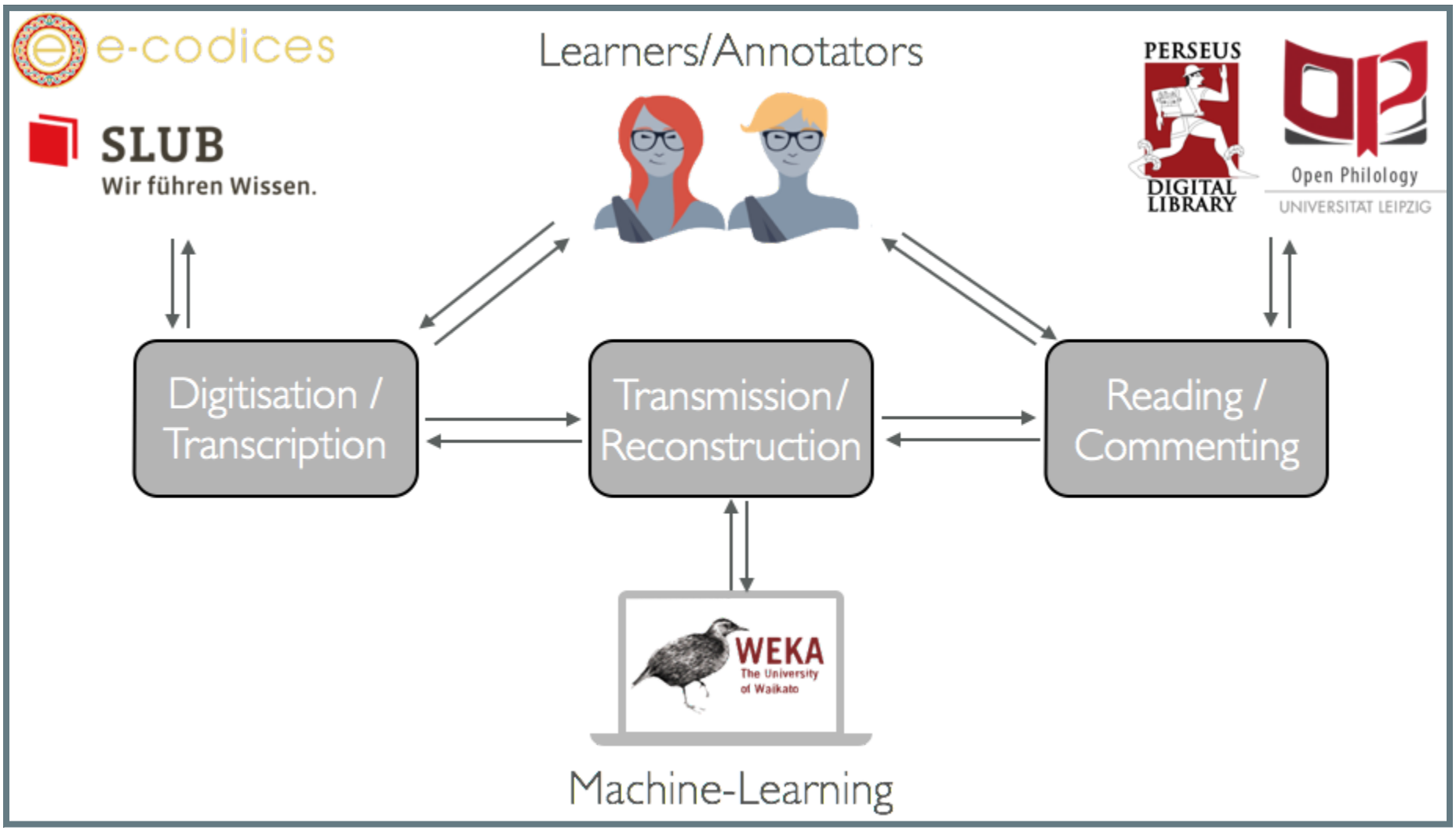
# CITIZEN SCIENCE



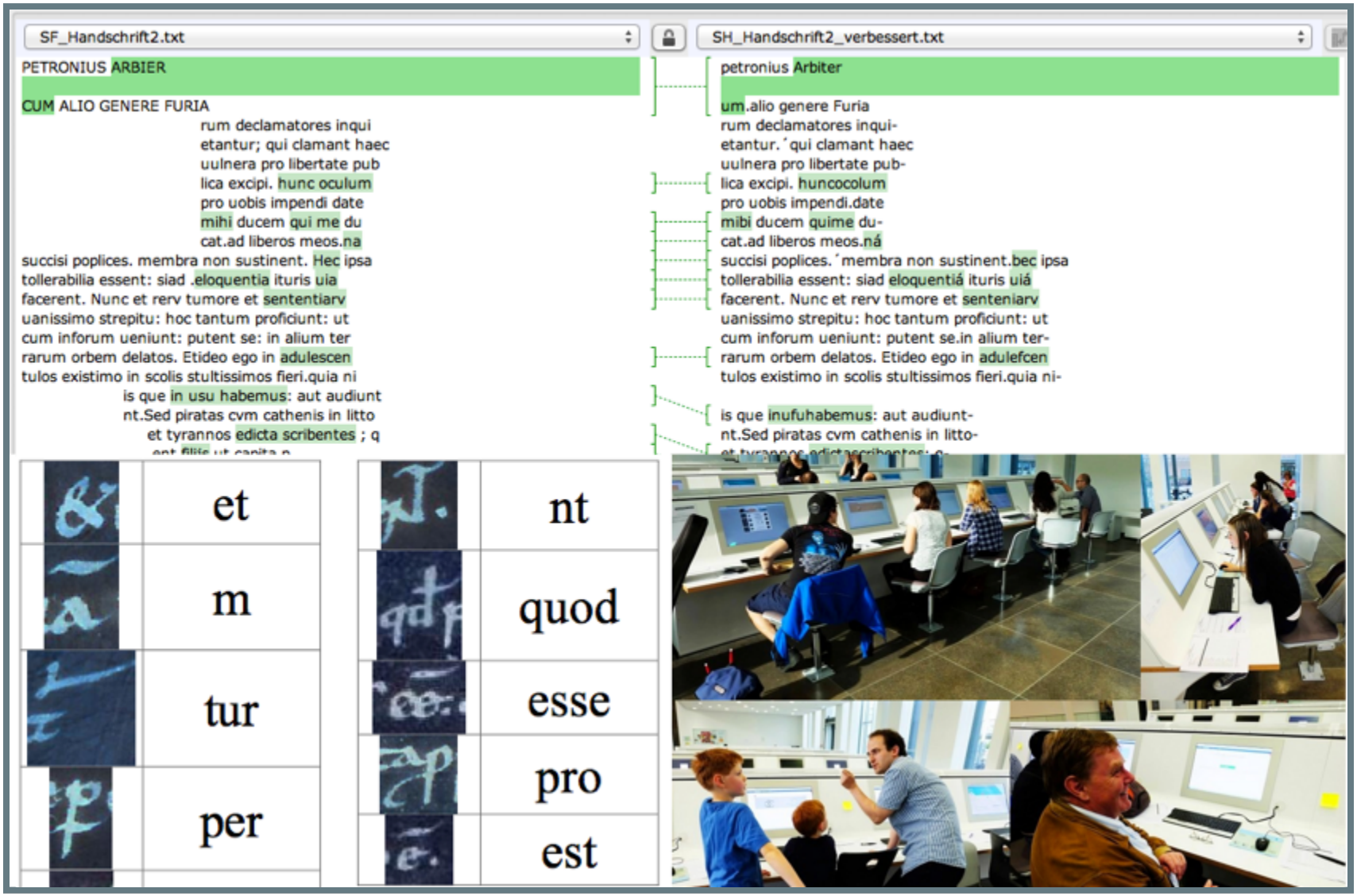

Sui, D.Z., Elwood, S. and M.F. Goodchild (eds.), 2013. Crowdsourcing Geographic Knowledge. Berlin: Springer


United States
2013
120,000 Latin and 2,210 Greek secondary education students
United Kingdom
2013
15,000 Latin and 2,000 Greek secondary education students
Finland
2013
465 Latin secondary education students
Denmark
2011
500 Latin and 100 Greek secondary education students
Flanders
2011
40,460 Latin and 5,760 Greek secondary education students
Austria
2012
63,000 Latin and 1,000 Greek secondary education students
Croatia
2013
24,737 Latin and 2,878 Greek secondary education students
Australia
2014
4,980 Latin and 183 Greek secondary education students
France
2013
501,100 Latin and 34,000 Greek secondary education students
NUMBER OF LATIN STUDENTS, E. FRANZINI, 2014
Germany
2012
800,000 Latin and 7,000 Greek secondary education students
Switzerland
2013
12,970 Latin and 930 Greek secondary education students
Italy
2011
2 million Latin and 680,000 Greek secondary education students
Malta
2014
80 Latin and 80 Greek secondary education students
South Africa
2014
100 Latin and 50 Greek secondary education students (estimate)
Bulgaria
2013
500 Latin and 400 Greek secondary education students
New Zealand
2013
1,500 Latin and 0 Greek secondary education students
University of Leipzig: Open Philology Project

# ARXE.TYPE WORKFLOW



# INVOLVING EVERYBODY



# ELEARNING BASED ON LINKED DATA (PROTOTYPE)

In Progress | Historical Languages eLear

139.18.40.154:8000

Ancient Geek

Login to Ancient Geek!

# Get your Greek on.

Learn Ancient Greek as you read Thucydides' harrowing tale of the war between

οἱ Λακεδαιμόνιοι (the Spartans) and οἱ Ἀθηναῖοι (the Athenians).

Start Learning!

## Game your way through Greek.

Develop your mad Greek skillz through games and choose the path to Greek glory that suits your personal learning style.

## Keep it original.

Explore Classical texts in their original languages, the way they were meant to be read, enhanced by data contributed by people like you.

## Be a scholar!

Contribute micropublications like translations and annotations to help create brand new, digital editions of Classical works!

# ELEARNING BASED ON LINKED DATA

# ELEARNING BASED ON LINKED DATA

Λ λ

**Say it like** the "l" in lithography.

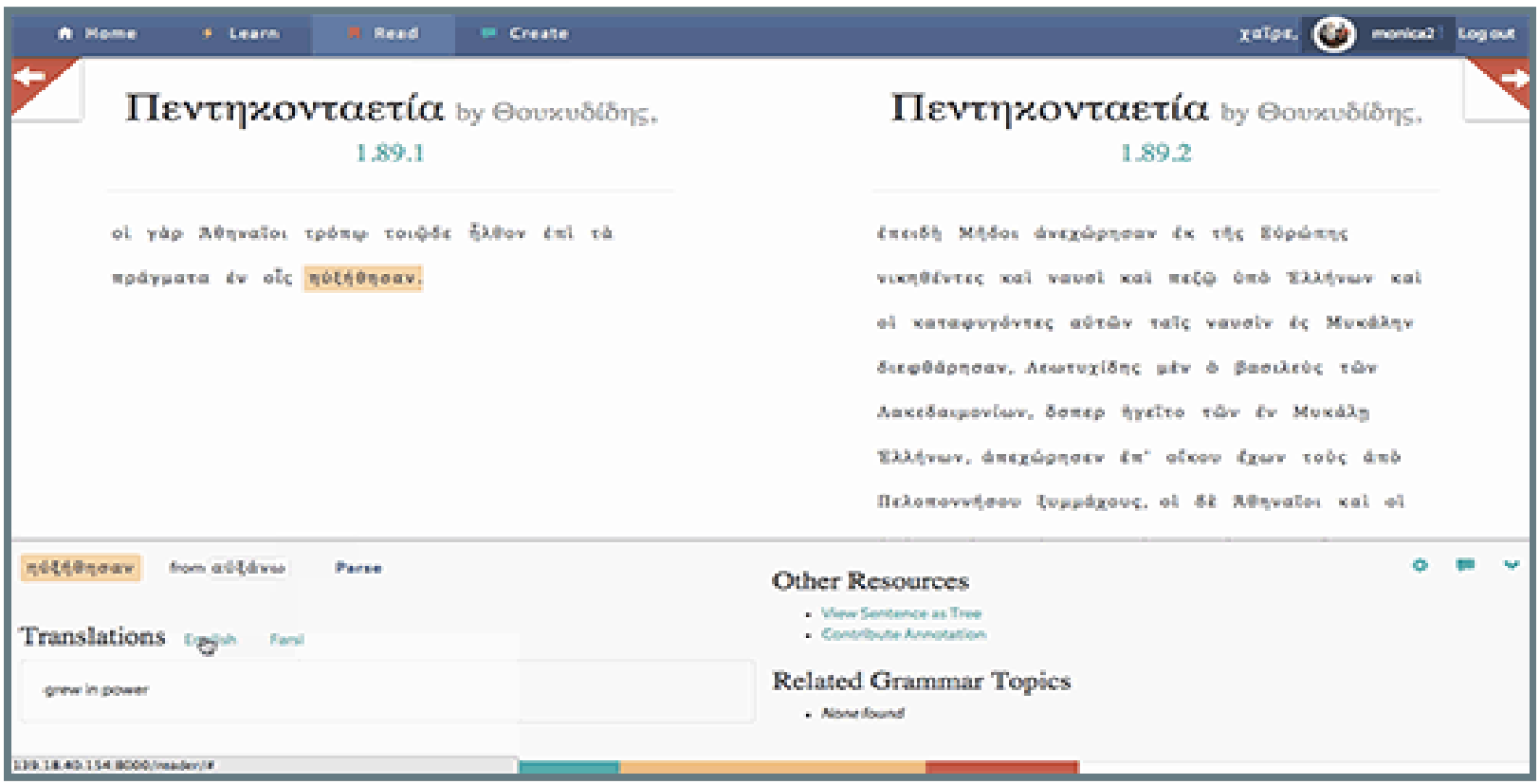
**IPA** [l] Listen

---

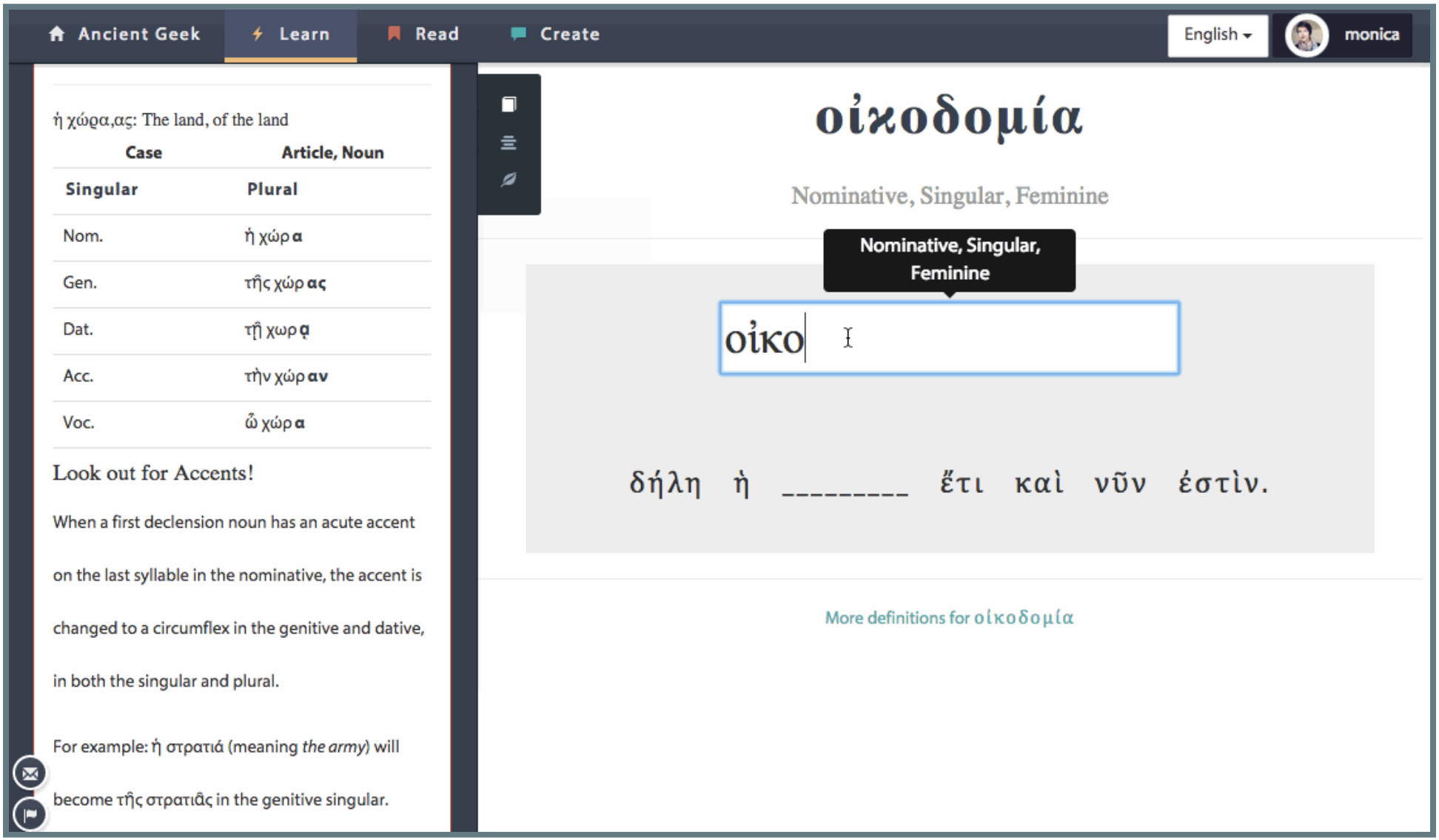
Lambda makes the same sound in λίθος (líthos), meaning "stone".

Think of *neolithic* (literally, new + stones), *lithology* (study of rock formation).

# ELEARNING BASED ON LINKED DATA

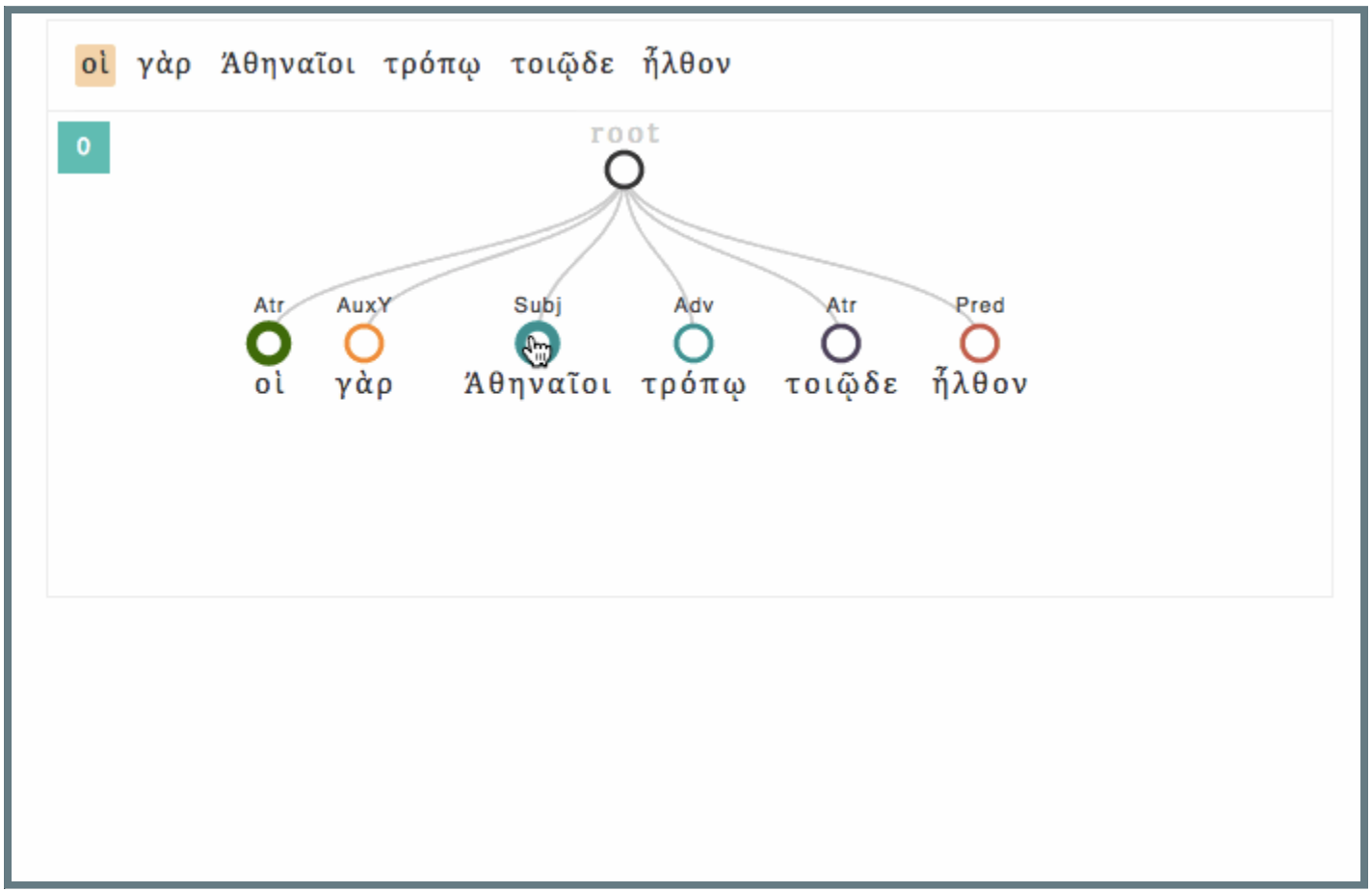


# ELEARNING BASED ON LINKED DATA



# ELEARNING BASED ON LINKED DATA

# ELEARNING BASED ON LINKED DATA



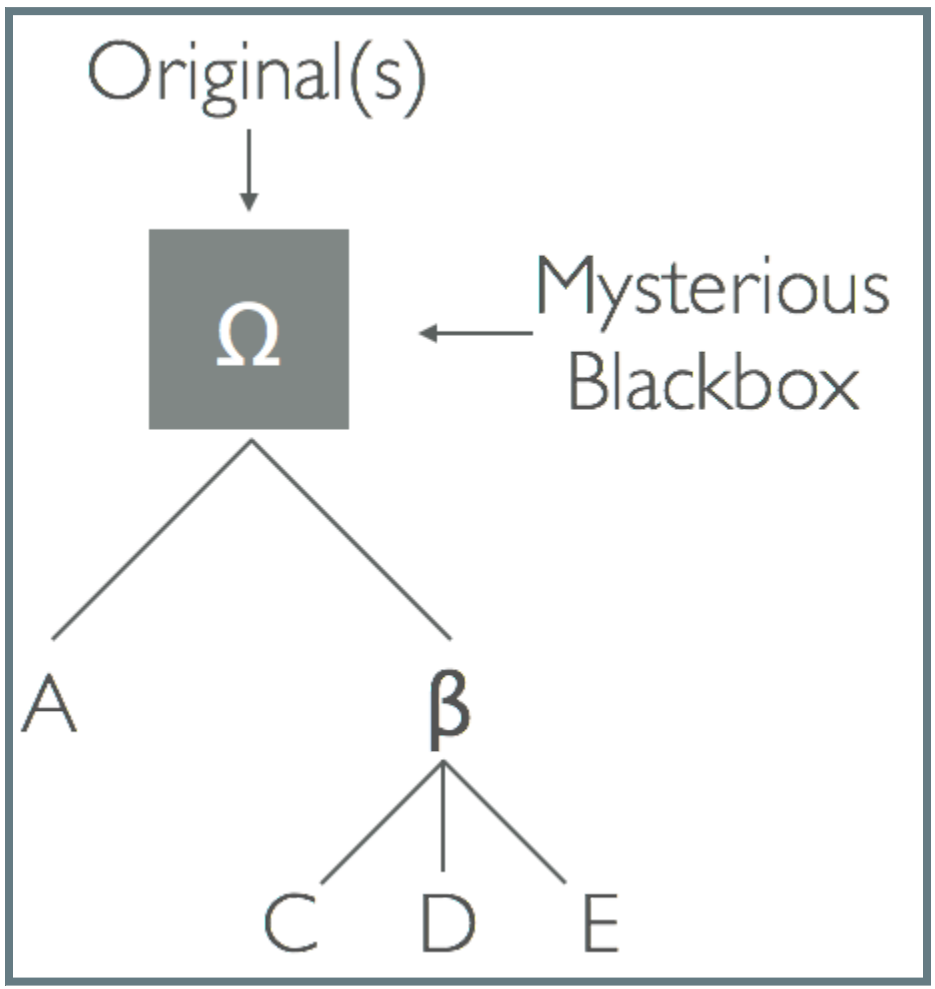
# COLLABORATIVE EDITING (WITH SCHOLARS)

- Fragmentary Texts
- Thematic Annotations
- Treebanks

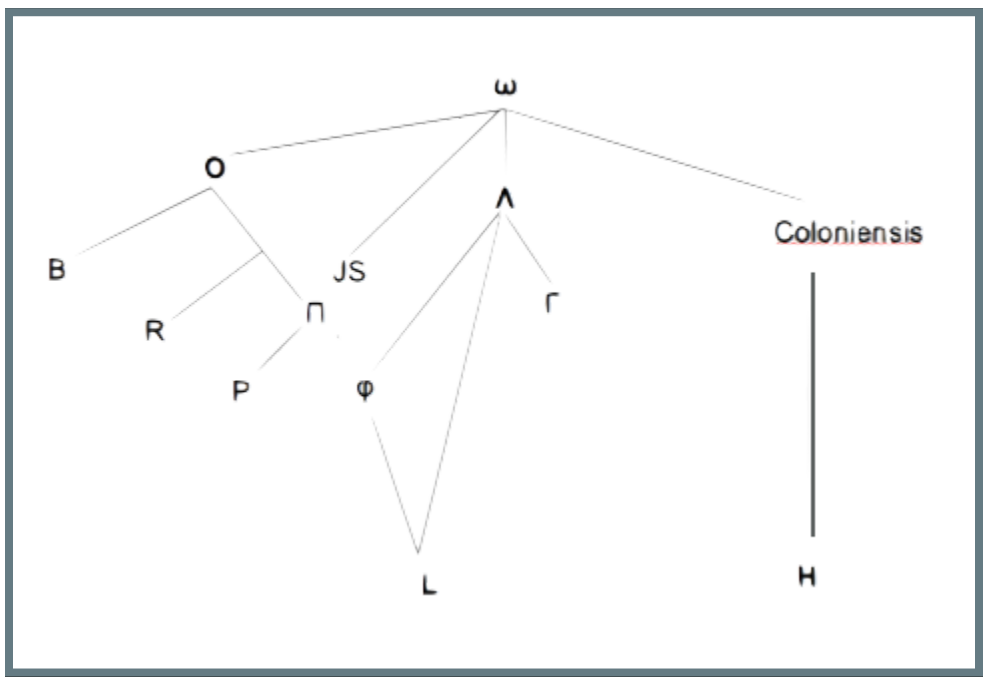
# RESEARCHING TEXTUAL TRANSMISSION

Textual transmission is a mono-directional communication process of text-data, graphics,and meta-data from the past to the present.

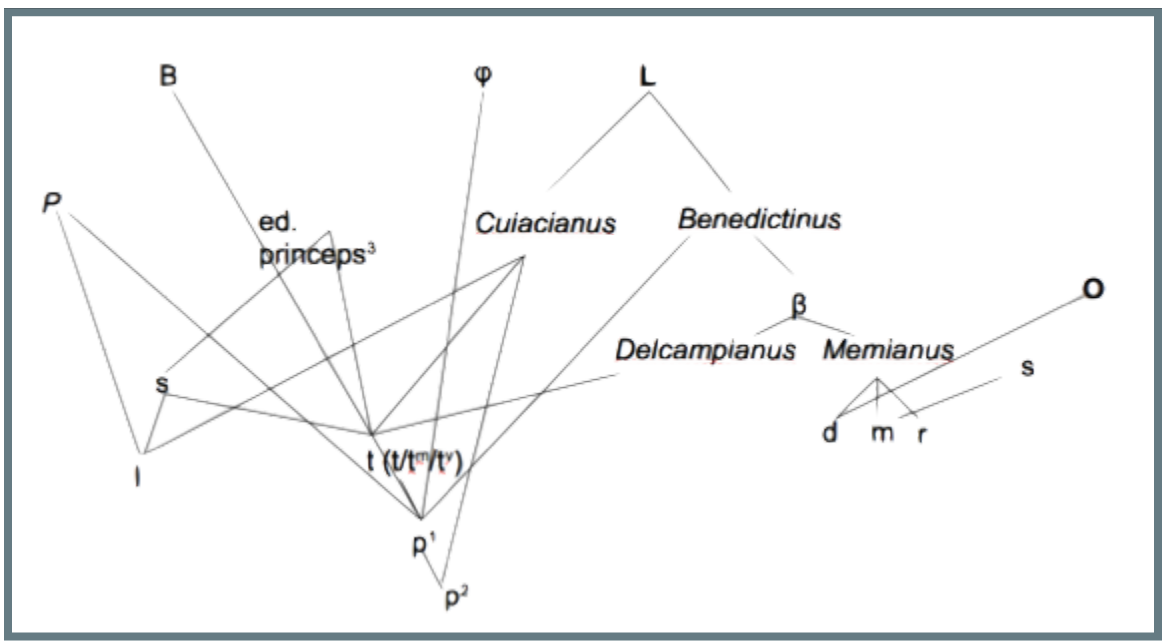
# RESEARCHING TEXTUAL TRANSMISSION



# RESEARCHING TEXTUAL TRANSMISSION



# RESEARCHING TEXTUAL TRANSMISSION



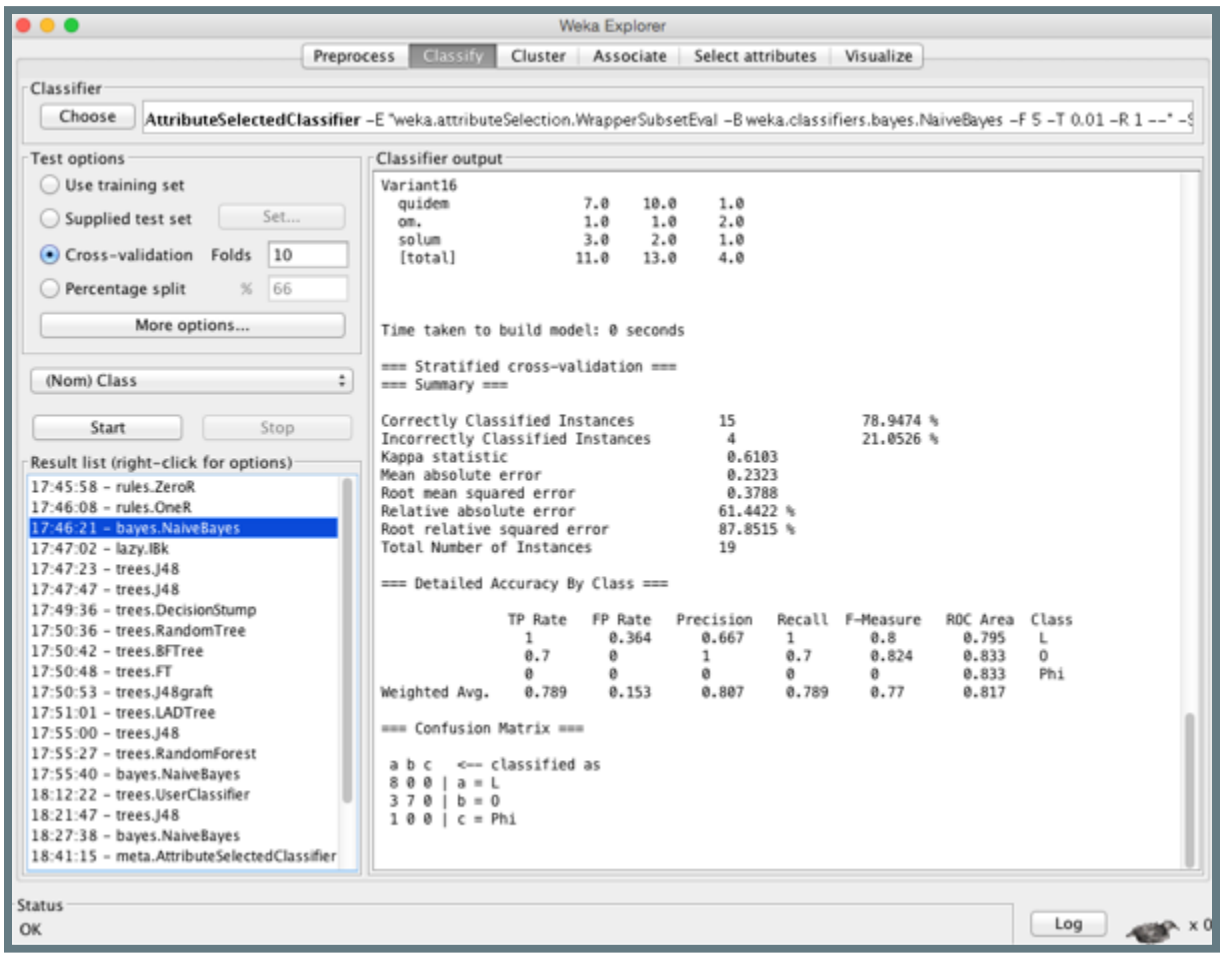

- OPP Manuscript Catalogue

# WEKA

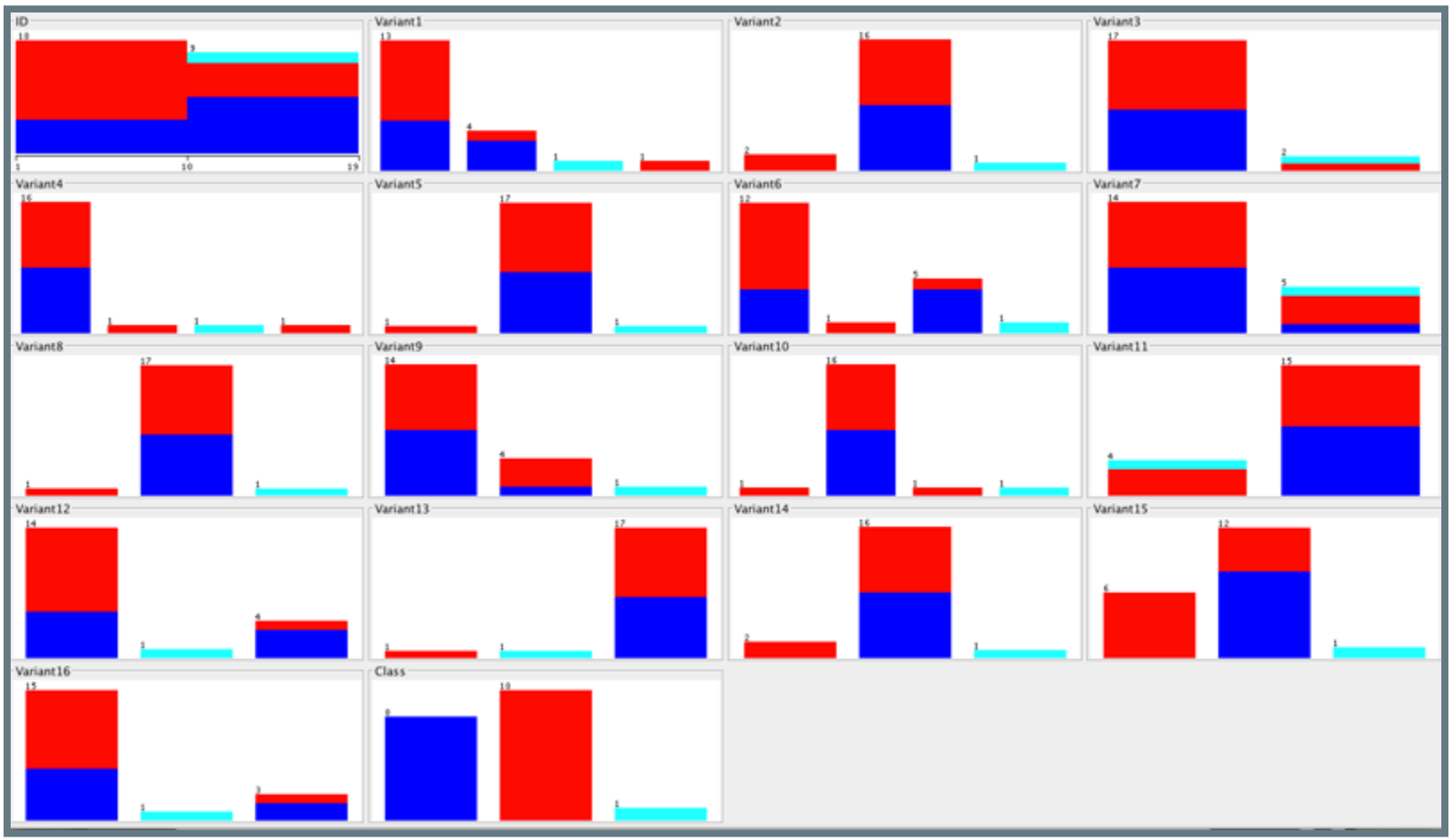
- Weka 3.7.11: Open Source Data Mining Software in Java
- A Collection of machine-learning algorithms for data-mining tasks
- GUI, but also implemented in e.g R, Python
- Although Weka is a powerful tool, classifying the manuscripts will be a challenge

Weka Homepage

# WEKA


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
AttributeSelectedClassifier -E "weka.attributeSelection.WrapperSubsetEval -B weka.classifiers.bayes.NaiveBayes -F 5 -T 0.01 -R 1 --" -S
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Class
Start
Stop
Result list (right-click for options)
17:45:58 - rules.ZeroR
17:46:08 - rules.OneR
17:46:21 - bayes.NaiveBayes
17:47:02 - lazy.IBk
17:47:23 - trees.J48
17:47:47 - trees.J48
17:49:36 - trees.DecisionStump
17:50:36 - trees.RandomTree
17:50:42 - trees.BFTree
17:50:48 - trees.FT
17:50:53 - trees.J48graft
17:51:01 - trees.LADTree
17:55:00 - trees.J48
17:55:27 - trees.RandomForest
17:55:40 - bayes.NaiveBayes
18:12:22 - trees.UserClassifier
18:21:47 - trees.J48
18:27:38 - bayes.NaiveBayes
18:41:15 - meta.AttributeSelectedClassifier
Classifier output
Variant16 7.0 10.0 1.0
quidem 1.0 1.0 2.0
om. 3.0 2.0 1.0
solum [total] 11.0 13.0 4.0
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 15 78.9474 %
Incorrectly Classified Instances 4 21.0526 %
Kappa statistic 0.6103
Mean absolute error 0.2323
Root mean squared error 0.3788
Relative absolute error 61.4422 %
Root relative squared error 87.8515 %
Total Number of Instances 19
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0.364 0.667 1 0.8 0.795 L
0.7 0 1 0.7 0.824 0.833 O
0 0 0 0 0 0.833 Phi
Weighted Avg. 0.789 0.153 0.807 0.789 0.77 0.817
=== Confusion Matrix ===
a b c <-- classified as
8 0 0 | a = L
3 7 0 | b = O
1 0 0 | c = Phi
Status
OK
Log
x 0

# WEKA


ID
Variant1
Variant2
Variant3
Variant4
Variant5
Variant6
Variant7
Variant8
Variant9
Variant10
Variant11
Variant12
Variant13
Variant14
Variant15
Variant16
Class

# WEKA ARFF FILE

```
@relation PetroniusManuscripts

@attribute ID numeric
@attribute Variant1 {Cum,Num,om.,Quum}
@attribute Variant2 {Nec,Non,om.}
@attribute Variant3 {Essent,om.}
@attribute Variant4 {ituris,nuris,om.,turis}
@attribute Variant5 {facent,facerent,om.}
@attribute Variant6 {rerum,runici,verborum,om.}
@attribute Variant7 {et,om.}
@attribute Variant8 {strepita,strepitum,om.}
@attribute Variant9 {terrarum_orbem,orbem_terrarum,om.}
@attribute Variant10 {qui,quia,quorum,om.}
@attribute Variant11 {his,iis}
@attribute Variant12 {et,om.,sed}
@attribute Variant13 {et,om.,sed}
@attribute Variant14 {pestilentia,pestilentiam,om.}
@attribute Variant15 {omnem,omnium,om.}
```

Structure of the ARFF Data file

# DATA CLEANING, VISUALISATIONS, TOPIC MODELLING

- Open Refine (aka GoogleRefine)
- Google Fusion

# BASIC D3 VISUALISATIONS

- Cartoonists' EQ Cartoons
- EQ Cartoons by Publisher
- How many Cartoons published on EQ
- Where were they published? (1)
- Where were they published? (2)
- Articles about EQ vs. Cartoons about EQ

# TOPIC MODELLING

- Done by an algorithm
- Algorithm analyses chunks of texts
- Tries topics to find statistically significant word co-occurrences
- Evaluates the results and reruns
- Topics Index New Zealand (LDAvis)
- Topics Index New Zealand (FusionTable)

# TOPIC MODELLING APP PROTOTYPE

- Cartoon Topic Explorer

# QUESTIONS?