

# Αξιοποίηση Ηλεκτρονικών Πόρων στην Ελληνική Φιλολογία-Digital Humanities (1): Μια ματιά στον χώρο των Digital Humanities

Στέλλα Μαρκαντωνάτου

Υπολογιστική Γλωσσολόγος

Διευθύντρια Ερευνών

Ινστιτούτο Επεξεργασίας του Λόγου/ΕΚ ΑΘΗΝΑ

stilianimarkantonatou@gmail.com

# 1. Εισαγωγή

Περιήγηση στον χώρο των Ψηφιακών  
Ανθρωπιστικών Επιστημών

# Ψηφιακές Ανθρωπιστικές Σπουδές

## Digital Humanities

- Σχετικά νεόκοπος όρος με αρκετά πλατειά δήλωση (denotation)
- Δηλώνει μεγάλο εύρος δραστηριοτήτων και ειδικοτήσεων: από την αποθήκευση κειμένων μέχρι την ανάπτυξη μοντέλων 3D!
- Χονδρικά ο όρος δηλώνει **ό,τι κάνουμε σήμερα με τους υπολογιστές και το Διαδίκτυο για να εξυπηρετήσουμε τις ανθρωπιστικές σπουδές**

# Ψηφιακές Ανθρωπιστικές Σπουδές

## Το μάθημά μας

- Θα περιηγηθούμε στον χώρο, όχι εξαντλητικά, ώστε να αποκτηθεί μία ευρύτερη, αναγκαστικά εγκυκλοπαιδικού χαρακτήρα, εικόνα αφού πρόκειται για παλίμψηστο απαιτητικών επιστημών.
- Θα σταθούμε στο κείμενο αλλά και πάλι η εικόνα θα είναι σε μεγάλο βαθμό εγκυκλοπαιδικού χαρακτήρα όσον αφορά τους **ΠΟΡΟΥΣ (resources)**:
  - το πώς αποθηκεύεται το (λογοτεχνικό ) κείμενο για το σήμερα και το αύριο
  - κάποια από τα εργαλεία (tools) και τους πόρους (resources) που είναι απαραίτητα για την χρήση του από ανθρώπους και μηχανές.

# Ψηφιακές Ανθρωπιστικές Σπουδές

## Το μάθημά μας

- Θα σταθούμε στην αναπαράσταση της δομής και της σημασίας του κειμένου
- Θα δούμε με μεγαλύτερη λεπτομέρεια και θα κάνουμε χειρωνακτική (hands-on) δουλειά
  - στους λεξικούς πόρους
  - στην ανάπτυξη μεταδεδομένων για την τεκμηρίωση των ανθρωπιστικών δεδομένων

Μια γενική εισαγωγή στις φιλολογικές σπουδές στην εποχή του διαδικτύου:

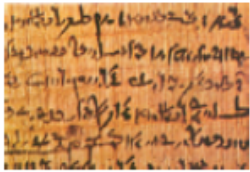
<http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-8&toc.depth=1&toc.id=ss1-2-8&brand=9781405103213> brand

## Ξεκινάμε από την περιήγηση

# Κείμενο: ψηφιοποίηση-επεξεργασία-συμπεράσματα

## General process

Raw data (analog)



Converting analog to digital  
(e.g., Scanning + OCR,  
Recording + Voice-to-text)



Digital data



Analysis (e.g.,  
machine learning,  
NLP, Computer vision)



**insights**

Visualization



**Graphic interface**

**Digital Humanities**

TAU 2016-2017

Nachum Dershowitz, Kfir Bar



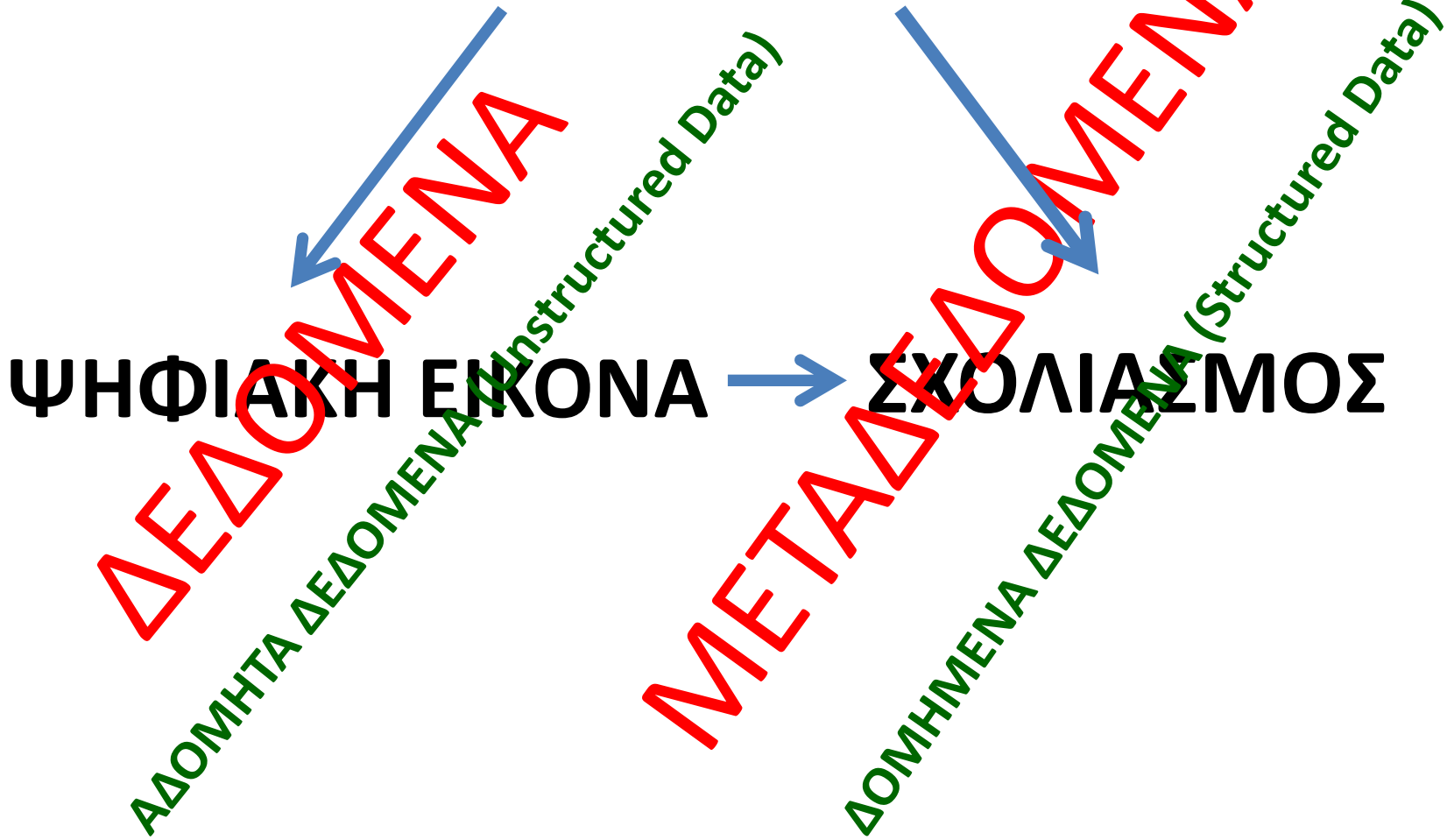
**ΤΑ ΘΕΜΑΤΑ  
ΠΟΥ ΘΑ ΔΟΥΜΕ**

Δεδομένα, μεταδεδομένα και  
διασυνδεδεμένα δεδομένα

**Data, Metadata and Linked Data**



# ΑΝΤΙΚΕΙΜΕΝΟ



# ΔΗΛΑΔΗ, ...

...θα μιλήσουμε για την **ΤΑΞΙΝΟΜΗΣΗ** των δεδομένων με **ΣΥΜΒΟΛΙΚΕΣ** προσεγγίσεις κυρίως

...θα δούμε ποιες λογικές σχέσεις κρύβονται πίσω από μια ταξινόμια

...θα μιλήσουμε για **ΟΝΤΟΛΟΓΙΕΣ**

# Zotero

The screenshot displays the Zotero application window. On the left, a sidebar shows a hierarchical library structure under 'My Library', including folders like 'academia book', 'kalamata course', 'bib', and 'my material'. The main pane shows a list of items with columns for 'Title' and 'Creator'. The selected item is 'When Metadata Becomes Outreach: Indexing, Describing, and Encodin... Wilson and Alexander'. The right pane provides a detailed view of this item, including its type ('Web Page'), title, authors ('Wilson, Emma Annette' and 'Alexander, Mary'), abstract, website title, URL, and dates.

**Library Structure:**

- My Library
  - academia book
  - kalamata course
    - bib
    - my material
  - My Publications
  - Duplicate Items
  - Unfiled Items
  - Trash

**Item List:**

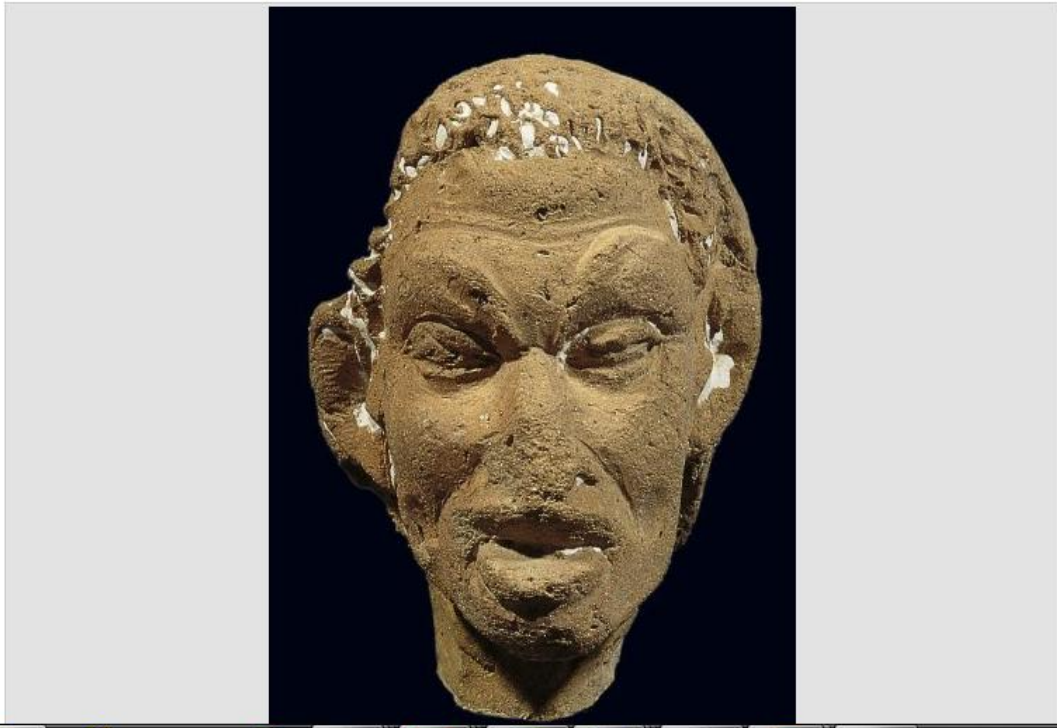
Title	Creator
Developing Research Tools via Voices from the Field ← dh+lib	
When Metadata Becomes Outreach: Indexing, Describing, and Encodin... Wilson and Alexander	Wilson and Alexander
Introduction to Metadata: Intro to Metadata	
Introduction to XML for Text	

**Item Details:**

- Item Type: Web Page
- Title: When Metadata Becomes Outreach: Indexing, Describing, and Encoding For DH
- Author: Wilson, Emma Annette
- Author: Alexander, Mary
- (...) Abstract: How can metadata become the most cuttin...
- Website Title: dh+lib where the digital humanities and librarianship meet
- Website Type: [Empty field]
- Date: [Empty field]
- Short Title: [Empty field]
- URL: <https://acrl.ala.org/dh/2016/07/29/when-m...>
- Accessed: [Empty field]
- Language: [Empty field]
- Rights: [Empty field]
- Extra: [Empty field]
- Date Added: 7/8/2018, 10:18:26 π.μ.
- Modified: 7/8/2018, 10:28:34 π.μ.

lecture\_1 my\_material

Home > Collections & Archives > Prehistoric, Ancient Greek and Roman Art >

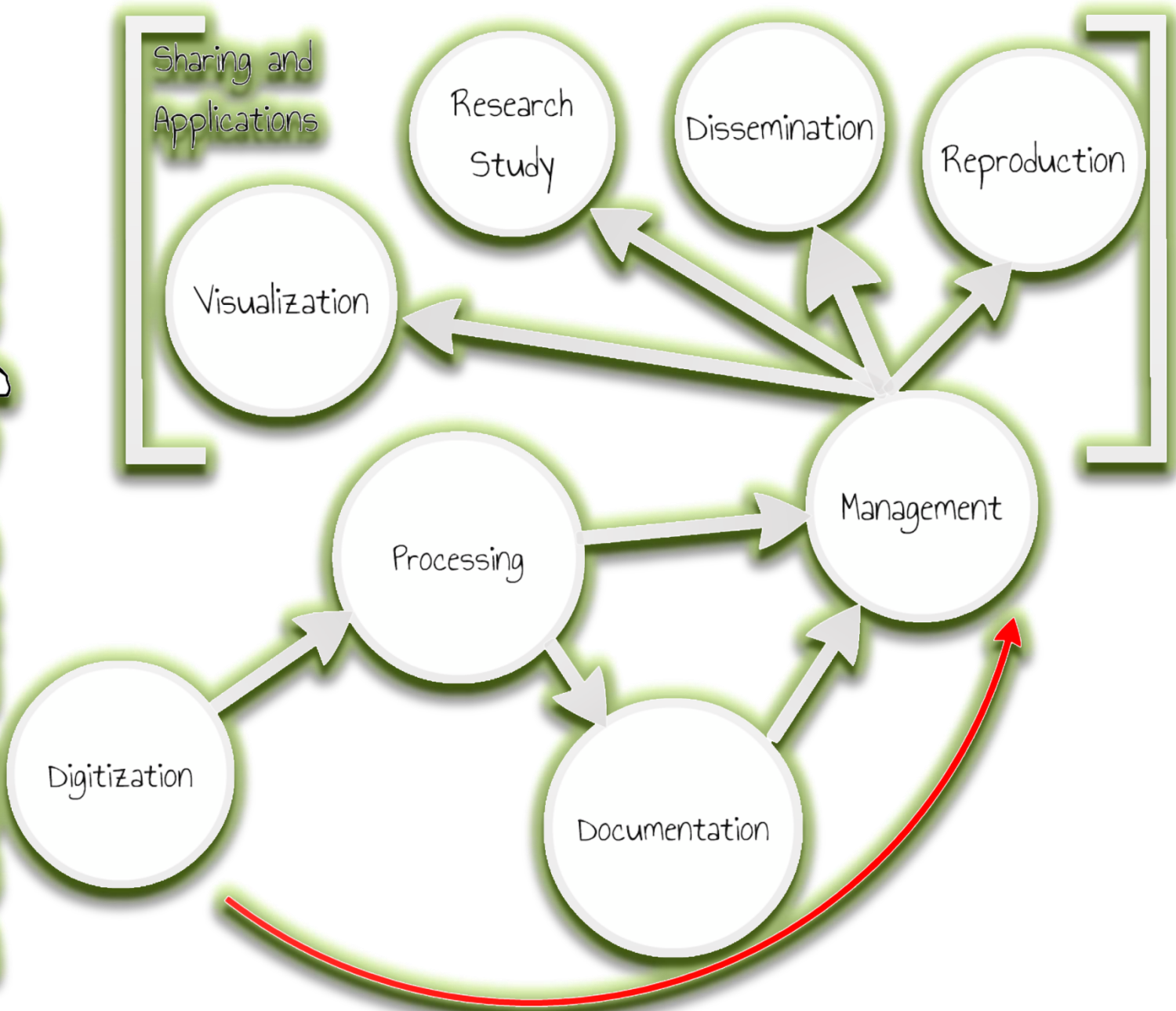


Head of terracotta figurine (caricature) of a man. The facial features are overly enlarged and certainly exaggerated. There are traces of white slip and red pigment. From Egypt, Hellenistic or Roman period. H. 0.048 m. (ΓΕ 8203)

MUSEUM OF GREEK CULTURE

Ψηφιοποιούμε κείμενα, εικόνες και αντικείμενα αλλά αυτό δεν αρκεί για να τα βρίσκουμε εύκολα. Δεν αρκεί ούτε για να τα καταλαβαίνουμε κάπως καλά. Χρειάζονται και **μεταδεδομένα**, κατάλληλα για κάθε είδος αντικειμένου. Και για να μιλούν οι μηχανές μεταξύ τους και άρα να υπάρχει πρόσβαση σε κατανεμημένη πληροφορία, θα πρέπει να ταιριάζουν μεταξύ τους τα διάφορα μεταδεδομένα ώστε να έχουμε **διασυνδεδεμένα δεδομένα**.

# Data Life Cycle



# ΕΝΑ ΠΑΡΑΔΕΙΓΜΑ

Emma Annette Wilson and Mary Alexander . 2016. When Metadata Becomes Outreach: Indexing, Describing, and Encoding For DH

*American Fashion and Fabrics* is a trade publication for the clothing industry that was published under a variety of titles from the 1940s until the 1980s. Various partial indexes of its issues and articles exist, but when Professor Amanda Thompson from the Department of Clothing, Textiles, and Interior Design, approached the ADHC about this project, **she did so because much of the most valuable material within the publications is not documented or searchable.** Contents not indexed include advertisements indicating trends in fabric design and purchasing; trade reports; fabric swatches retaining their original colors thanks to living in library stacks; and some editorials, articles, poems, and advertorials. **Knowing that a digital index would enhance the research and teaching use of this publication, Thompson approached metadata librarian Mary Alexander directly about the problem, and from there the project transitioned into a DH initiative at the ADHC.**

# Συνέχεια...

Alexander, with the assistance of metadata librarian Vanessa Unkeless-Perez, has worked at length with sample issues of *American Fashion and Fabrics* to create a custom schema capable of documenting this wide variety of materials. Significant challenges have arisen in terms of describing the **fabric swatches**, a centerpiece of the publication and its importance in the field. For instance, **how much detail is required to make these usable and findable?** Can we, and should we, **describe the type of weave used in a fabric?** Its **texture? Pattern? Weight? Color?** In a project of this scale, multiple different people will need to create the metadata, and the more complex the information documented, the more specialized their knowledge needs to be, and the more prescriptive the input guidelines.



# ...τέλος!

One of the most challenging aspects of this project involved the very human problem of perception. Everyone perceives colors slightly differently and they might describe patterns in fabric slightly differently. A controlled vocabulary for all of these different elements of the fabric swatches would be enormous to the point of unwieldiness not only in terms of finding the apposite terms but also the time required to describe every single swatch. Our current approach to this problem is under discussion and involves a two-pronged approach. **The controlled vocabulary for colors will be primary and secondary hues, thus limiting perception differences by limiting the number of color terms available.** The second prong will further develop critical thinking skills of upper-class students majoring in textiles, our proposed data-input workforce, **to analyse characteristics of fabrics and to apply appropriate terms from controlled vocabularies for weaves, patterns, and other areas, or to pause and seek clarification when the degree of certainty for applying a term is not 100%.** Thompson is responsible for answering student questions and proposing new terms to be added to the controlled vocabulary lists, so the lists can continue to expand.

Πολλαπλή κληρονομικότητα

# Συνδεδεμένα δεδομένα (linked data)

The technological universe of crosswalks, mapping, federated searching over heterogeneous databases, and aggregating full-text resources and metadata sets into single repositories is rapidly changing. Crosswalks are still at the heart of supporting conversion projects and enabling the semantic interoperability that makes it possible to search across heterogeneous distributed databases. Inherently, there will always be limitations to crosswalks; there is rarely a one-to-one correspondence between metadata standards, even when one standard is a subset of another. **Mapping the elements or fields of metadata systems is only one part of the picture. Crosswalks of controlled terms or thesauri will further enhance searchers' ability to retrieve the most precise search results. As the number and size of online resources increases, the ability to refine searches and to use controlled vocabularies and thesauri both at the metadata creation stage and at the moment of searching will become increasingly important.**

Mary S. Woodley. 2016. Metadata Matters: Connecting People and Information. In (Ed.) Murtha Baca. *Introduction to Metadata*. 3<sup>rd</sup> Edition. Getty Foundation

**BIBΛΙΟΓΡΑΦΙΑ**

# Τι θα συζητήσουμε

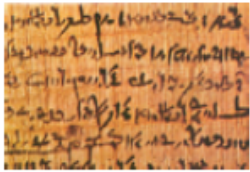
- Τεκμηρίωση μουσείων (ανάγκες, πρότυπα, σχέση με την υφιστάμενη πολιτιστική πραγματικότητα)
- Θησαυροί για την τεκμηρίωση μουσείων (AAT, EKT, Muse...)
- EUROPEANA και τα λεξικά της
- Συνδεδεμένα δεδομένα (linked data)
- Τι κάνουμε στην Ελλάδα

Ηλεκτρονικά κείμενα,  
επισημείωση και markup  
**E-Texts, Annotation & Markup**

# Κείμενο: ψηφιοποίηση-επεξεργασία-συμπεράσματα

## General process

Raw data (analog)



Converting analog to digital  
(e.g., Scanning + OCR,  
Recording + Voice-to-text)



Digital data



Analysis (e.g.,  
machine learning,  
NLP, Computer vision)



**insights**

Visualization



**Graphic interface**

**Digital Humanities**

TAU 2016-2017

Nachum Dershowitz, Kfir Bar



# Ηλεκτρονικά κείμενα, σχολιασμός και markup

## E-Texts, Annotation & Markup

### markup:

-Παραδοσιακά, δηλώνει τον σχολιασμό και γενικά τα σημάδια πάνω σε ένα κείμενο που καθοδηγούν τον τυπογράφο στην μορφοποίηση

-Στην ηλεκτρονική εποχή, ο όρος καλύπτει και τους ειδικούς χαρακτήρες που εισάγονται στα ηλεκτρονικά κείμενα για να καθοδηγούν την μορφοποίηση, το τύπωμα και, σε κάποιες περιπτώσεις, την επεξεργασία των κειμένων

- Οι σύγχρονοι κειμενογράφοι (editors) όπως το Word ή τα προγράμματα μορφοποίησης (π.χ. LATEX) στην ουσία «επισημειώνουν» το κείμενο με ειδικούς χαρακτήρες που ερμηνεύονται από τα προγράμματα που δημιουργούν την τελική μορφή του (αυτήν που διαβάζουμε ή τυπώνουμε).
- Τα κείμενα ---χειρόγραφα ή τυπωμένα--- που προέρχονται από ψηφιοποίηση ή από κανονική αντιγραφή προφανώς δεν έχουν τέτοιους χαρακτήρες.

```
Source Rich Text Edit Find $ s B I H A L E X C  
37 many NLP (Natural Language Processing) researchers among others. This  
38 may indeed be a good way of accounting for basic \textit{propositional}  
39 \textit{logic}, but it accounts at best for only a  
40 very limited subset of natural language use. An alternative view is  
41 that logics are by-products of natural language. At the very least, we  
42 may say that the relationship between language and logic is not well  
43 understood. If the "Lego set" theory of meaning in language were  
44 tenable, it would not have been necessary for NLP and AI (Artificial  
45 Intelligence) researchers such as \cite{ide2006}, after many  
46 years of intensive (and expensive) effort, to declare that projects in  
47 \textit{Word Sense Disambiguation (WSD)} have failed to achieve even their most  
48 basic goals.  
49 \begin{quotation}  
50 At present, WSD work is at a crossroads: systems have hit a reported  
51 ceiling of 70\%+ accuracy \citep[kilgarriff2004], the source and  
52 kinds of sense inventories that should be used in WSD work is an issue  
53 of continued debate, and the usefulness of stand-alone WSD systems for  
54 current NLP applications is questionable. (\citealt[15]{ide2006}).  
55 \end{quotation}  
56  
57  
58 The alternative view mentioned here is supported by lexicographers such
```

Patrick Hanks, Ismail El Maarouf & Michael Oakes

### 1 Introduction: Phraseology and Multi-Word Expressions

Traditionally, people have long believed that each word has one or more meanings and that these meanings can be selected and put together as if in a child's Lego set, to construct propositions, questions, etc. This belief is still widely (and unquestioningly, unthinkingly) held by many NLP (Natural Language Processing) researchers among others. This may indeed be a good way of accounting for basic propositional logic, but it accounts at best for only a very limited subset of natural language use. An alternative view is that logics are by-products of natural language. At the very least, we may say that the relationship between language and logic is not well understood. If the "Lego set" theory of meaning in language were tenable, it would not have been necessary for NLP and AI (Artificial Intelligence) researchers such as Ide & Wilks (2007), after many years of intensive (and expensive) effort, to declare that projects in Word Sense Disambiguation (WSD) have failed to achieve even their most basic goals.

At present, WSD work is at a crossroads: systems have hit a reported ceiling of 70%+ accuracy (Kilgarriff et al. 2004), the source and kinds of sense inventories that should be used in WSD work is an issue of continued debate, and the usefulness of stand-alone WSD systems for current NLP applications is questionable. (Ide & Wilks 2007: 15).

The alternative view mentioned here is supported by lexicographers such as Atkins and Kilgarriff and Hanks (2000). These lexicographers argue that much of the meaning of an utterance is carried by underlying patterns of co-selection of the words actually used, rather than by simple concatenations. These conclusions overlap to some extent with the tenets of construction grammar, though the methodologies are very different. In corpus linguistics, Sinclair declared, after a lifetime's empirical research into texts, corpora, and meaning, "Many if not most meanings require the presence of more than one word for their normal realisation" (Sinclair 1998: 4).

If these lexicographers and corpus linguists are right, it might appear that MWEs play a central role in the meaningful use of language. They are not merely an irritating set of exceptions, as used to be thought. According to this, MWEs



Γιατί

**επισημειώνουμε τα κείμενα που  
αποθηκεύουμε ως προς την μορφή τους;**

Διότι

**η μορφή των κειμένων φέρει**

**ΣΗΜΑΣΙΑ**

# ΜΟΡΦΕΣ ΜΕ ΣΗΜΑΣΙΑ

- Παραθέσεις:** Πόσους και ποιους παραθέτει αυτούσιους στα έργα του ο Παναγής Λεκατσάς;
  - Σημειώσεις στο περιθώριο:** Πόσες και ποιες σημειώσεις βλέπουμε στο περιθώριο στο χειρόγραφο του Γιούγκερμαν του Καραγάτση;
  - Γραφικοί χαρακτήρες (για χειρόγραφα):** Πόσοι μοναχοί τελικά δούλεψαν πάνω στο χειρόγραφο της «Πολιτείας» του Πλάτωνα που φυλάσσεται στο Βατικανό;
  - Διορθώσεις (για δακτυλογραφημένα έντυπα):** Πόσες και ποιες διορθώσεις περιέχει το δακτυλογραφημένο κείμενο του Τρίτου Στεφανιού του Ταχτσή;
- Χρησιμοποιούμε μια ειδική γλώσσα, την XML, και αναπτύσσουμε ένα πρότυπο για τον σκοπό αυτό.**

Ηλεκτρονικά κείμενα, επισημείωση & markup  
E-Texts, Annotation & Markup

**ΕΠΙΣΗΜΕΙΩΣΗ ΩΣ ΠΡΟΣ ΤΑ  
ΓΛΩΣΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ  
ΤΟΥ ΚΕΙΜΕΝΟΥ**

**Κύρια Ονόματα (named entities):** Ποια κύρια ονόματα και με ποια συχνότητα χρησιμοποιεί στην Ασκητική ο Καζαντζάκης;

**Σημειώσεις στο περιθώριο:** Ποια ρήματα χρησιμοποιεί ο Καζαντζάκης στις σημειώσεις του στο περιθώριο στο χειρόγραφο του «Ο Χριστός ξανασταυρώνεται»;

<https://www.historical-museum.gr/webapps/kazantzakis-pages/gr/work/mutations-keimeno01-synexes01.php>

**Πολυλεκτικά:** Πόσες πολυλεκτικές/ιδιωματικές εκφράσεις απαντώνται στα κείμενα του Παπαδιαμάντη;

**Συντακτικές δομές:** Τελικά, ο Ταχτσής χρησιμοποιεί το ασύνδετο σχήμα στο Τρίτο Στεφάνι;

**Χρησιμοποιούμε την XML και για αυτόν τον σκοπό.**

# Επισημειωτής κειμένου για επεξεργασία φυσικής γλώσσας

## Perspective

Sentence ▾


page: 3 ▾

## Selector


Automatic (deepest) ▾

## Legend • Entity

(Hide)

 LVC.full

 VID

 LVC.cause

228 Το αποτέλεσμα της πολιτικής δημοπρασίας, όμως, για την αδειοδότηση, προκαλεί ακόμη περισσότερο με την επιδότηση που παρέχει τώρα, στους ιδιοκτήτες αφού:

229 Παραγράφει τις ευθύνες τους από το 2008 μέχρι σήμερα για τις χιλιάδες απολύσεις, μειώσεις μισθών, ατομικές συμβάσεις εργασίας, υπονόμευση ασφαλιστικών δικαιωμάτων, συνδικαλιστικές διώξεις.

230 (LVC.full) (LVC.full)  
Δίνει ακόμη στους ιδιοκτήτες άλλοθι να μην πληρώσουν δεδουλευμένα, αποζημιώσεις, οφειλές στα Ταμεία, υποχρεώσεις σε Τράπεζες κ.λπ.

231 Προσφέρει « μπόνους » στους υπερθεματιστές να αναπτύξουν τις δραστηριότητές τους, καθιερώνοντας για αυτούς καθεστώς προστασίας και παρέχοντάς τους « φθηνά χέρια » με τσακισμένα δικαιώματα από τις τάξεις των παλαιών και των νέων ανέργων και εργαζομένων για το σταθμό τους.

232 Τους εξασφαλίζει, ταυτόχρονα, διαφημιστική δαπάνη, η οποία προβλέπεται ιδιαίτερα αυξημένη, επειδή συγκεντρώνεται πλέον μόνο σε συγκεκριμένους παραλήπτες.

233 (VID - -)  
Η επιλογή του Πρωθυπουργού στη ΔΕΘ να κουνήσει το δάχτυλο σε συναδέλφους, οι οποίοι προσπαθούν να σώσουν την δουλειά τους, δεν απαντάει ούτε στα δημοσιογραφικά ερωτήματα, ούτε στο αμείλικτο ερώτημα της ανεργίας και, βεβαίως, δεν είναι επιχείρημα η κυβερνητική θέση: « όπως οι άλλοι κι εσείς ».

234 Εκτός και αν η στάση αυτή υπηρετεί διαδικασία κοινωνικού αυτοματισμού και εν προκειμένω δαιμονοποίησης του δημοσιογραφικού κόσμου στο σύνολό του, στον οποίο επιχειρεί να φορτώσει όλα τα δεινά που πληρώνει η κοινωνία.

235 Στο ίδιο κλίμα κυβερνητικά στελέχη, τόσο στο πρόσφατο παρελθόν, όσο και τώρα, στοχοποιούν δημοσιογράφους με σκοπό να τους απαξιώσουν, όπως έγινε και με τη συνάδελφο κ. Ευαγγελία Τσικρικά, με χυδαία επίθεση που υπέστη από τον αναπληρωτή του Υπουργού Υγείας Ανδρέα Ξανθού, Πάυλο Πολάκη.

236 Δεν είναι, όμως, άμοιροι ευθυνών παλιοί και νέοι ιδιοκτήτες των ΜΜΕ, όχι μόνο επειδή με τις τρεις ολονυκτίες τους στη Γενική Γραμματεία Ενημέρωσης νομιμοποίησαν τη διαδικασία, την οποία ταυτόχρονα καταγγέλλουν, αλλά διότι τώρα δεν διστάζουν να χρησιμοποιήσουν εργαζόμενους στις όποιες επιχειρηματικές επιλογές και επιδιώξεις τους.

237 (LVC.full) (LVC.full)  
Η αγωνία για τη διατήρηση της θέσης εργασίας δεν πρέπει να δίνει την εντύπωση ότι η απόσταση ανάμεσα στα δικαιώματα του ενοαζόμενου και στο διευθυντικό δικαίωμα του ενοδοτή υποσεί να υπενδιστεί.

# Τί θα συζητήσουμε

- Λεξικοί πόροι
  - WordNet, FrameNet, BabelNet
  - Πύλη, άλλα ελληνικά ηλεκτρονικά λεξικά
  - Πολυλεκτικά-ΙΔΙΟΝ
- Γλωσσικά εργαλεία
- Αποθετήρια
- Κατάλληλα επισημειωμένα σώματα κειμένων

**ΑΛΛΑ ΘΕΜΑΤΑ**

# Ψηφιακές δημοσιεύσεις

## Digital Publishing

- Digital publishing (online publishing or e-publishing): δημοσίευση στο διαδίκτυο ή σε άλλους φορείς ηλεκτρονικής πληροφορίας περιοδικών, ηλεκτρονικών βιβλίων (e-books), αναφορών αλλά και μουσικής και εικόνων, γενκώς κάθε μορφής περιεχομένου.
- Η διάθεση των ηλεκτρονικών δημοσιεύσεων μπορεί να είναι επί χρήμασι ή δωρεάν (ελεύθερη).

**Η ελεύθερη διάθεση περιεχομένου κατακτά έδαφος στην ακαδημαϊκή κοινότητα.**



# Οι ψηφιακές δημοσιεύσεις είναι **πολυτροπικές**

Amazon.com's Ereader, Kindle με την  
πλατφόρμα KDP

Παιχνίδια, apps στο Google Play Store

Αρχεία MP3 στο iTunes που δημοσιεύει  
ψηφιακά μουσική και την διαθέτει στα iPods,  
iPhones κ.λπ.

Ψηφιακές εικόνες που δημιουργήθηκαν με  
Adobe Illustrator ή Adobe Photoshop

# Ακαδημαϊκές ψηφιακές δημοσιεύσεις

Κάλλιπος <https://repository.kallipos.gr/>

<https://repository.kallipos.gr/handle/11419/2209>

- Τίτλος: Σημασιολογικές προσεγγίσεις κειμενικών δεδομένων
- Συγγραφείς: Τάντος, Αλέξανδρος  
Μαρκαντώνη, Στέλλα
- Κριτικός Αναγνώστης: Γκοτσούλια, Παρασκευή

# LangSci

<http://langsci-press.org/opendata>

- Language Science Press publishes high quality Open Access monographs and edited volumes in the domain of linguistics.
- **The project was born from the insight that community-based publishing of scientific literature will be more cost-efficient than publishing by commercial publishers.** New technologies and the collaborative work methods enabled by the Internet mean that many of the traditional tasks provided by publishers are no longer needed or can be provided at a fraction of the cost. Furthermore, the inherent interest of maximising profit margins is often at odds with the interest of science. This is first and foremost the case of business models which rely on monetising access to information via paywalls. These models have to make sure that the information they provide remains a scarce resource. A publishing platform run by scientists themselves, which states the furthering of access to knowledge as its goal, rather than the restriction of access to knowledge, represents a valuable asset for society at large.

# Principles of Language Science Press

- **1. Open Access: all books are available without a fee, worldwide.** Printed copies can be ordered for low prices via print-on-demand service providers. Additionally, all other aspects of the publication process should also be as open as possible (Open Source Software, Open Business Data, optional Open Peer Review)
- **2. Quality: Peer review is mandatory.** Language Science Press positions itself as a premium publisher with the corresponding requirements for manuscripts. There is a fair chance that a manuscript will be rejected, and authors should know this.

# Χωρικές Ανθρωπιστικές Επιστήμες

## Maps, Apps & the GeoWeb: Spatial Humanities

- Οργάνωση της γεωγραφικής πληροφορίας με ψηφιακά gazetteers με βάση τις ιστορικές, αρχαιολογικές και γενικά πολιτιστικές μαρτυρίες
- Συνάρτηση της γεωγραφικής πληροφορίας με τις χρονικές περιόδους (time periods).
- Η γεωγραφία στην ψηφιακή φιλολογική και κειμενική ανάλυση.

Δύο βίντεο:

- [YouTube/ymI5Uv5cGU4](https://www.youtube.com/watch?v=ymI5Uv5cGU4)
- [youtu.be/bMi281yzsTA](https://youtu.be/bMi281yzsTA)

# ...και ένα πρόβλημα: ο χάρτης της Ελλάδας και η μελιτζανοσαλάτα



Τυπικά η μελιτζανοσαλάτα περιέχει ψιλοκομμένες ψημένες, ίσως και καμένες, μελιτζάνες και κάτι άλλο ... **εξαρτάται από τον τόπο.**

Υπάρχουν και τα **διαλεκτικά** ονόματα που μπορεί να τα δει κανείς σε έναν κατάλογο εστιατορίου.

Αν θέλαμε να φτιάξουμε ένα εργαλείο για τον τουρίστα που

-εξηγεί τί περιέχει η μελιτζανοσαλάτα με βάση την γεωγραφική θέση του χρήστη

-μεταφράζει καταλόγους ελληνικών εστιατορίων

θα πρέπει να ενημερώσουμε το σύστημα με gazetteers.

# ...και ένα πρόβλημα: ο χάρτης της Ελλάδας και η μελιτζανοσαλάτα

Πολύ πρόχειρο “Gazetteer” για τη μελιτζανοσαλάτα

- Standard name: μελιτζανοσαλάτα
- Country: Greece
- Dialectical name: (Pontus): χαβίτς
- Ingredients: Varieties:
  - Southern Greece: aubergines, lemon/vinegar, garlic, olive oil
  - Northern Greece: aubergines, peppers, vinegar, garlic, olive oil
  - Pontus (now Northern Greece): aubergines, yogurt, lemon, ....
- Technique: roasted--nearly burnt -aubergines, finely chopped, mixed with...
- Related with: Serbia:ajvar, Bulgaria: kyopolou, Turkey: patlican salatasi

**ΤΡΙΣΔΙΑΣΤΑΤΕΣ ΑΝΑΚΑΤΑΣΚΕΥΕΣ,  
ΑΡΧΑΙΟΛΟΓΙΑ**

**3D RECONSTRUCTION, ARCHAEOLOGY**



# Ροτόντα Θεσσαλονίκη



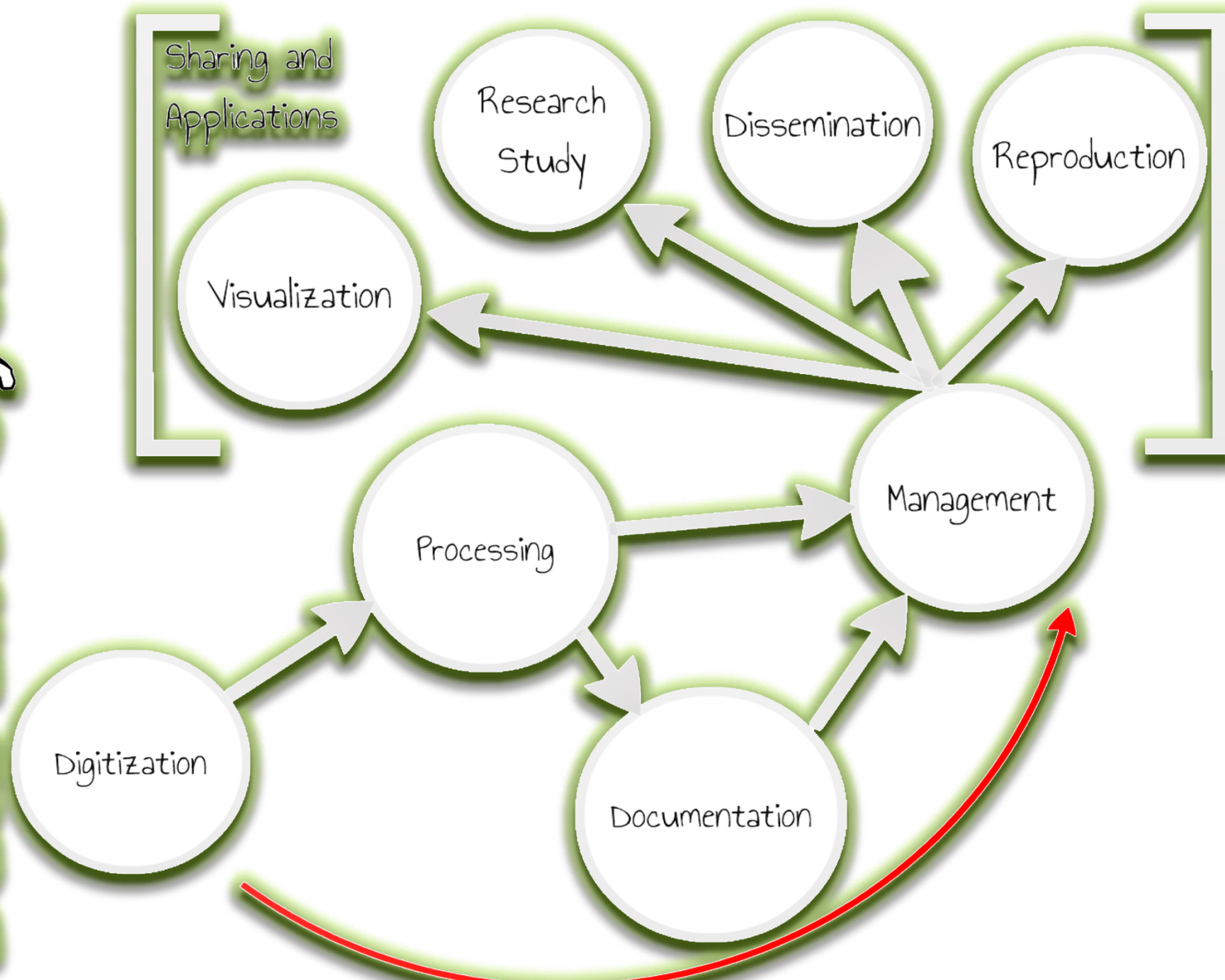
• gg =

# Αργοστόλι, Κεφαλονιά



<https://www.youtube.com/watch?v=wTGJkxv0w>

# Data Life Cycle



# ΠΑΙΧΝΙΔΙΑ-ΕΜΒΥΘΙΣΗ GAMING



<https://www.youtube.com/watch?v=L6AmoG9K5tA>

# STORY TELLING



# ΔΟΜΗ ΜΑΘΗΜΑΤΟΣ

- Δ1: Εισαγωγή
- Δ2: Text Encoding
- Δ3-Δ6: Λεξικοί πόροι
- Δ7-Δ8/Δ9: Γλωσσικά εργαλεία –Ειδικές βάσεις κειμενικών δεδομένων-Αποθετήρια
- Δ9/Δ10-Δ12: Μεταδεδομένα και δεδομένα για βιβλιοθήκες και μουσεία, ελεγχόμενα λεξιλόγια-θησαυροί, Αποθετήρια
- Δ13: Παρουσίαση της εξέλιξης των εργασιών, συζήτηση (μπορεί να μετατεθεί λίγο νωρίτερα)

# ΕΞΕΤΑΣΗ ΜΑΘΗΜΑΤΟΣ 1

65% ΕΡΓΑΣΙΑ

Κωδικοποίηση 15 ΠΛΕ στο ΙΔΙΟΝ

Η εργασία θα δοθεί στην 4<sup>η</sup> διάλεξη και θα παραδοθεί στο τέλος Δεκεμβρίου 2018.



# ΕΞΕΤΑΣΗ ΜΑΘΗΜΑΤΟΣ 2

**30% ΠΑΡΟΥΣΙΑΣΕΙΣ** κατά την διάρκεια των διαλέξεων αναφερόμενες σε άρθρα σχετικά με το μάθημα. Οι παρουσιάσεις θα γίνονται από ομάδες φοιτητών (2 άτομα).

1. PowerPoint
2. Φύλλο αξιολόγησης από το ακροατήριο (ανώνυμη αξιολόγηση)
3. Όλοι οι φοιτητές: μισή σελίδα κείμενο «Τι άκουσα σήμερα»





# Για το επόμενο μάθημα: Ανεμόσκαλα

- <http://www.greek-language.gr/Resources/literature/tools/concordance/index.html>

Κριτική θεώρηση των δυνατοτήτων αναζήτησης:

-ποιες είναι αυτές;

-τι επισημείωση προϋποθέτουν;

-επαρκείς;

-τι άλλο;

# Ευχαριστίες

Στον Γιώργο Παυλίδη, Δ/ντή Ερευνών ΙΕΛ/ΕΚ  
Αθηνά για την ευγενική χορηγία υλικού για  
αυτήν την παρουσίαση.

