

ΣΕΙΡΑ ΔΙΑΔΙΚΤΥΑΚΩΝ
ΜΑΘΗΜΑΤΩΝ

Επιμέλεια μεταδεδομένων ψηφιακών συλλογών

ΣΤΙΣ ΑΝΘΡΩΠΙΣΤΙΚΕΣ
ΕΠΙΣΤΗΜΕΣ ΚΑΙ ΤΕΧΝΕΣ

30 ΜΑΡΤΙΟΥ 2021
06 ΑΠΡΙΛΙΟΥ 2021
13 ΑΠΡΙΛΙΟΥ 2021

5.00-8.00 μ.μ.

απο ωνίς



ΑΘΗΝΑ Έρευνα & Καινοτομία
Τεχνολογίες Πληροφορίας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ
ΟΙΚΟΝΟΜΙΑΣ & ΑΝΑΠΤΥΞΗΣ
ΕΛΛΗΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΣΤΑΤΙΣΤΙΚΗΣ
& ΔΕΛΤΙΑ ΣΤΑΤΙΣΤΙΚΗΣ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΠΑΝΕΚ 2014-2020
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ
ΕΠΙΧΕΙΡΗΣΙΑΚΟΤΗΤΑ
ΚΑΙΝΟΤΟΜΙΑ

ΕΣΠΑ
2014-2020
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Οδηγίες - Κανονισμοί

- Ενημερώνουμε ότι η παρούσα εκδήλωση θα βιντεοσκοπηθεί και ενδέχεται να δημοσιευτεί με την άδεια των εισηγητών.
- Παρακαλούμε να μην πάρετε ή δημοσιεύσετε screenshots των παρευρισκομένων στο μάθημα χωρίς την προηγούμενη συγκατάθεσή τους
- Παρακαλούμε να έχετε το μικρόφωνό σας κλειστό κατά τη διάρκεια του μαθήματος
- Για τυχόν απορίες και ερωτήσεις, σας παρακαλούμε να χρησιμοποιήσετε το chat, αναγράφοντας και σε ποιον ομιλητή απευθύνεται η ερώτηση. Οι ερωτήσεις θα _ και απαντηθούν στο τέλος του μαθήματος

ΑΠΟΛΛΩΝΙΣ

Εθνική Υποδομή για τις Ψηφιακές Ανθρωπιστικές
Επιστήμες και Τέχνες και για τη Γλωσσική Έρευνα
και Καινοτομία

Εταίροι

Ε.Κ. Αθηνά
Ακαδημία Αθηνών
ΕΚΕΦΕ «Δημόκριτος»
Εθνικό Δίκτυο Υποδομών Τεχνολογίας και Έρευνας
ΕΚΠΑ
ΕΠΙΣΕΥ - ΕΜΠ
ΑΣΚΤ
ΙΤΕ
ΑΠΘ
Πανεπιστήμιο Αιγαίου
Ιόνιο Πανεπιστήμιο



> Σδυάς >

απολλωνίς <

clarin:el <



Μάθημα II: Επιμέλεια των μεταδεδομένων

Κάποιες διευκρινίσεις

Πύλη υπηρεσιών ΑΠΟΛΛΩΝΙΣ

- Συγκεντρώνει σε ένα σημείο ψηφιακές υπηρεσίες, εργαλεία και πόρους για τη διευκόλυνση της έρευνας
- Παρέχει τη δυνατότητα δημιουργίας συλλογών και ευρετηρίων κατόπιν συμφωνίας με την υποδομή
- Πιλοτική εφαρμογή για Δεκαετία 1940

Υπηρεσίες

Ανακαλύψτε διαδικτυακές υπηρεσίες, ψηφιακά εργαλεία και πόρους των δύο συνιστωσών της Υποδομής ΑΠΟΛΛΩΝΙΣ, CLARIN:EL και DARIAH-GR/ΔΥΑΣ, και υποστηρίξτε την έρευνά σας σε Ανθρωπιστικές Επιστήμες, Τέχνες και Γλωσσική Τεχνολογία.

Υπηρεσίες & Εργαλεία

Εδώ θα βρείτε διαδικτυακές υπηρεσίες που μπορείτε να χρησιμοποιήσετε εντός της Υποδομής και ψηφιακά εργαλεία που μπορείτε να κατεβάσετε τοπικά στον υπολογιστή σας, για να εκτελέσετε εργασίες κατά τη διάρκεια όλου του κύκλου ζωής της έρευνάς σας.

Μάθημα II: Επιμέλεια των μεταδεδομένων

Κάποιες διευκρινίσεις

Δεκαετία 1940 - Πάροχοι

92.286

Τόποι | Χρόνοι | Πρόσωπα | Θέματα | Γεγονότα

Μάθημα II: Επιμέλεια των μεταδεδομένων

Διδάσκουσες

Μαρία Ηλβανίδου, Επιστημονική Συνεργάτιδα,
Μονάδα Ψηφιακής Επιμέλειας, ΙΠΣΥ, Ε.Κ. Αθηνά &
Τμήμα Πληροφορικής ΟΠΑ

Βίκη Ντρίτσου, Επιστημονική Συνεργάτιδα, Μονάδα
Ψηφιακής Επιμέλειας, ΙΠΣΥ, Ε.Κ. Αθηνά &
Μεταδιδακτορική Ερευνήτρια, Τμήμα Πληροφορικής
ΟΠΑ

Βίκυ Λιακοπούλου, Επιστημονική Συνεργάτιδα,
Μονάδα Ψηφιακής Επιμέλειας, ΙΠΣΥ, Ε.Κ. Αθηνά

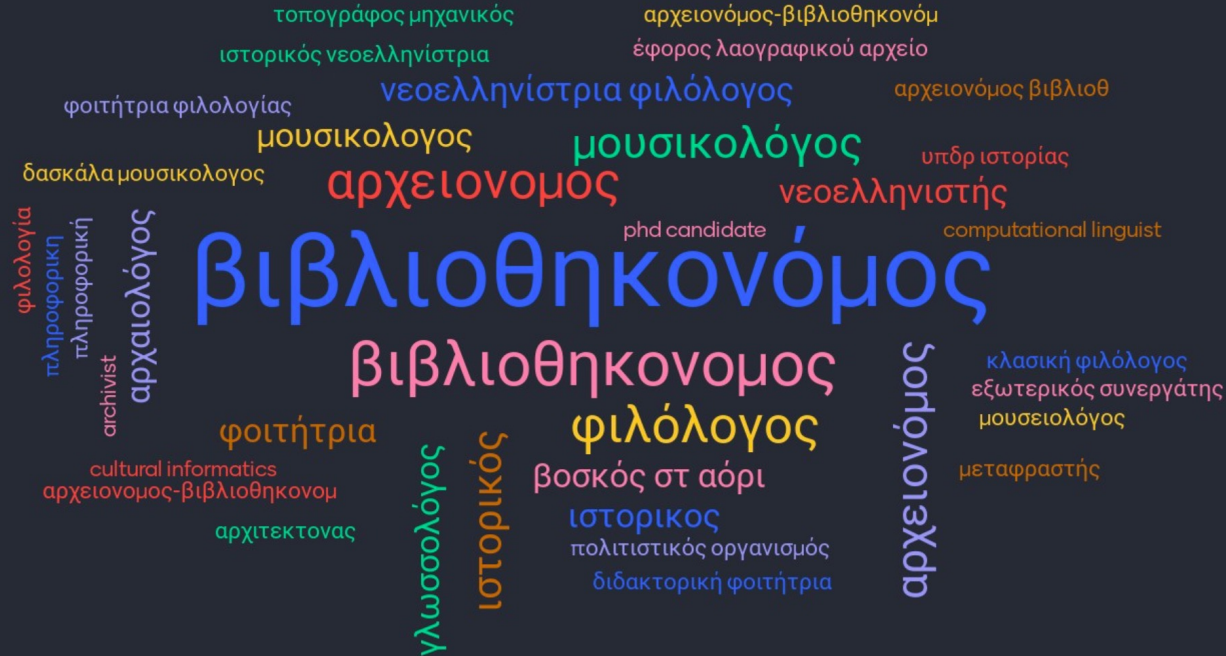
Επιμέλεια των μεταδεδομένων - Καθαρισμός

Μαρία Ηλβανίδου

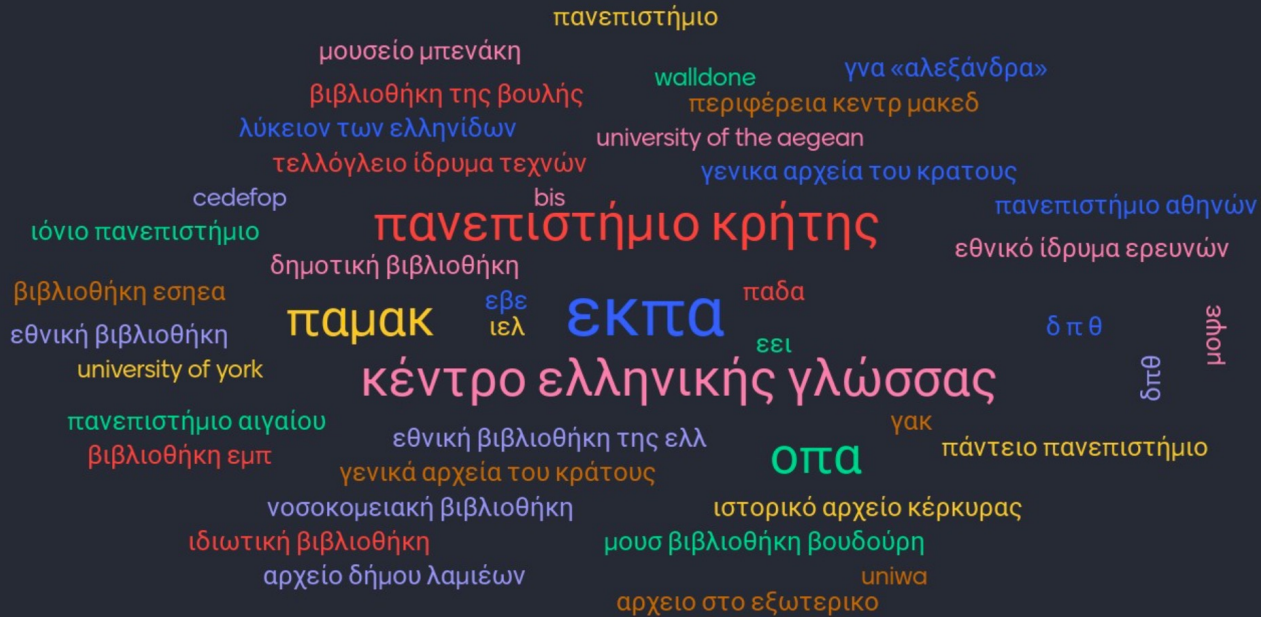
Επιστημονική συνεργάτιδα, Μονάδα Ψηφιακής Επιμέλειας, ΙΠΣΥ, Ε.Κ. Αθηνά
& Τμήμα Πληροφορικής, ΟΠΑ



Ποια είναι η ειδικότητά σας;



Με ποιον οργανισμό (τυχόν) συνδέεστε;



Οι προκλήσεις του περιεχομένου 1/2

- Διαφορετικοί **μορφότυποι**
- Διαφορετικά **σχήματα τεκμηρίωσης**
- Διαφορετικά λεξιλόγια / **ορολογία**
- **Έλλειψη τυποποίησης** στην τεκμηρίωση / ανακολουθίες
- Τυπογραφικά και ορθογραφικά **λάθη**, εναλλαγή ελληνικού και λατινικού αλφαβήτου, **συντομογραφίες** και αρκτικόλεξα, **κεφαλαιογράμματα** γραφή, **καθαρεύουσα**, κ.λπ.

Οι προκλήσεις του περιεχομένου 2/2

Η δημιουργία ενός συναθροιστή μεταδεδομένων που να επιτρέπει την κοινή **ευρετηρίαση** βάσει των εξής κριτηρίων:

πρόσωπα | τόπος | χρόνος | θέματα | γεγονότα

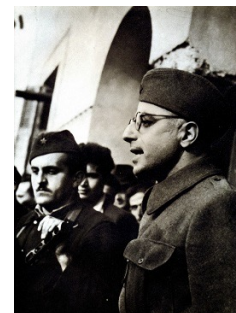


Οι προκλήσεις του περιεχομένου 2/2

Η δημιουργία ενός συναθροιστή μεταδεδομένων που να επιτρέπει την κοινή **ευρετηρίαση** βάσει των εξής κριτηρίων:

πρόσωπα

1. Α. Τζήμας (Β. Σαμαρινιώτης ή Σαμαριναίος) Μέλος Στρατιωτικής Διοίκησης Γενικού Στρατηγείου Ε.Λ.Α.Σ.
2. Α. Τζήμας (Β. Σαμαρινιώτης ή Σαμαριναίος) Μέλος Στρατιωτικής Διοίκησης Γενικού Στρατηγείου Ε.Λ.Α.Σ. και Μόνιμος Αντιπρόσωπος Γενικού Στρατηγείου Ε.Λ.Α.Σ. στην Ανώτατη Διοίκηση του Λαϊκού Απελευθερωτικού Στρατού Γιουγκοσλαβίας
3. Τζήμας
4. Τζήμας Ανδρέας (Σαμαρινιώτης)
5. Σαμαριναίος
6. ΤΖΗΜΑΣ



Οι προκλήσεις του περιεχομένου 2/2

Η δημιουργία ενός συναθροιστή μεταδεδομένων που να επιτρέπει την κοινή **ευρετηρίαση** βάσει των εξής κριτηρίων:

| τόπος

Τόπος	GeoNames
<i>Αθήνα</i>	
<i>ΑΘΗΝΑ</i>	
<i>Αθήναι</i>	id: 264371
<i>Αθηνών</i>	
<i>Αθηναίων</i>	



Οι προκλήσεις του περιεχομένου 2/2

Η δημιουργία ενός συναθροιστή μεταδεδομένων που να επιτρέπει την κοινή **ευρετηρίαση** βάσει των εξής κριτηρίων:

| χρόνος

Χρονικές εκφράσεις	Ημ/νία
<i>06/25/44</i>	
<i>25-6-1944</i>	
<i>25 Ιουνίου 1944</i>	<i>ηη/μμ/xxxx</i>
<i>Καλοκαίρι 1944</i>	
<i>1944</i>	



Οι προκλήσεις του περιεχομένου 2/2

Η δημιουργία ενός συναθροιστή μεταδεδομένων που να επιτρέπει την κοινή **ευρετηρίαση** βάσει των εξής κριτηρίων:

| Θέματα

Θέματα

Fronti Nacional Clirimtar (Αλβανικό Ε.Α.Μ.)

Αλβανικό Ε.Α.Μ.

Αλβανικό Εθνικό Απελευθερωτικό Μέτωπο

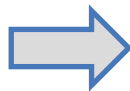
Αλβανοί αντάρτες



Γιατί καθαρισμός των μεταδεδομένων;

Το πρόβλημα

1. ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ
2. ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ
3. ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ_ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ
4. ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ
5. ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ
6. ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ
7. ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ
8. ΜΟΝΑΔΑ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ
9. ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΑ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ



ΜΟΝΑΔΕΣ ΚΑΙ ΣΧΗΜΑΤΙΣΜΟΙ ΤΟΥ ΕΛΛΗΝΙΚΟΥ ΣΤΡΑΤΟΥ

Προ καθαρισμού ανακτώνται

14.724 εγγραφές

Μετά τον καθαρισμό ανακτώνται

17.168 εγγραφές



Διαδικασία καθαρισμού 1/5

Βήμα 1

Επιλέξτε το κατάλληλο εργαλείο καθαρισμού των μεταδεδομένων.



AutoSave OFF AA_data.xls - Compatibility Mode

Home Insert Draw Page Layout Formulas **Data** Review View Acrobat Tell me

Get External Data Refresh All Connections Properties Edit Links Stocks Currencies Geography

Sort Filter Advanced Flash Fill Text to Columns Remove Duplicates Data Validation Consolidate What-If Analysis Group Ungroup Subtotal Analysis Tools

L1 x ✓ fx klidia

	F	G	H	I	J	K	L	M	N
1	ar	pigi	ap2	par2	ar2	pigi2	klidia	sh	
2	τηλ. 994, most immediate, 28 Οκτ. 1940	F.O.371/2 4919/R80 55					I. Μεταξάς Πρωθυπουργός Ελλάδας, Sir C. M. Palairot Πρεσβευτής Βρετανίας στην Αθήνα, Βρετανία - Υπουργείο Εξωτερικών (Foreign Office) (F.O.), E. Grazzi Πρεσβευτής Ιταλίας στην Αθήνα, Ιταλία, Ιταλία - Ενοπλες δυνάμεις - Επιχειρήσεις στην Ελλάδα, Βρετανία - Ενοπλες δυνάμεις - Επιχειρήσεις στην Ελλάδα, Κέρκυρα, Αθήνα, Ελλάδα - Οικονομία		

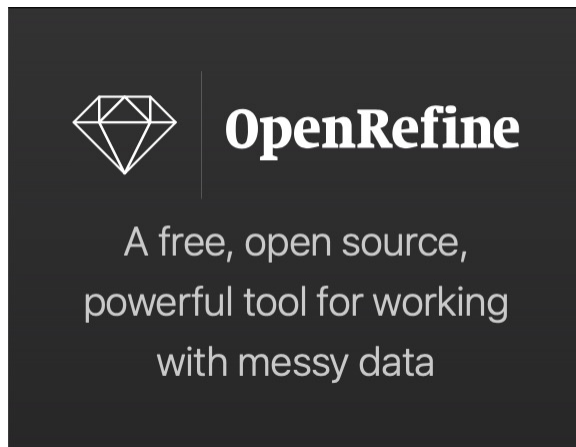
AA_data +

Ready Count: 28556 209%

Διαδικασία καθαρισμού 1/5

Βήμα 1

Επιλέξτε το κατάλληλο εργαλείο καθαρισμού των μεταδεδομένων.



Διαδικασία καθαρισμού 2/5

Βήμα 2

Εξετάστε τους υποστηριζόμενους μορφότυπους και εισαγάγετε τα δεδομένα

Βήμα 3

Εξετάστε εάν υπάρχουν σφάλματα κωδικοποίησης ή μετατροπής κατά την εισαγωγή



Διαδικασία καθαρισμού 3/5

Βήμα 4

Διατυπώστε τους κανόνες καθαρισμού με βάση το υλικό και τους στόχους σας. Π.χ.:

- θα διορθώσετε μόνο ορθογραφικά και τυπογραφικά σφάλματα;
- θα επιχειρήσετε και τυποποίηση των τιμών;
- ποια είναι η προτιμητέα ορθογραφία (σε περίπτωση εναλλακτικών τρόπων γραφής των ίδιων όρων);
- με ποια μορφή θέλετε να αναγράφονται οι ημερομηνίες;
- επιλέγετε τη χρήση ενός όρου αντί άλλου;



Διαδικασία καθαρισμού 4/5

Βήμα 5

Εφαρμόστε τον καθαρισμό με βάση τους κανόνες που διατυπώσατε.

α. Οριζόντια

- Κενές εγγραφές ή διπλοεγγραφές
- Κενά διαστήματα στην αρχή και το τέλος των τιμών
- Κωδικοποίηση ειδικών χαρακτήρων
- Χρήση πεζών και κεφαλαίων γραμμάτων

β. Ανά πεδίο

- Τύποι πεδίων
- Δομή πεδίων
- Ομαδοποίηση και συγχώνευση τιμών με τη χρήση αλγορίθμων
- Χειροκίνητη διόρθωση σφαλμάτων.



Διαδικασία καθαρισμού 5/5

Βήμα 6

Επαναλάβετε όποιο τυχόν από τα βήματα της εκτέλεσης του καθαρισμού κρίνετε απαραίτητο μέχρις ότου πετύχετε το μέγιστο δυνατό βαθμό τυποποίησης

Βήμα 7

Εξαγάγετε τα καθαρισμένα δεδομένα σας στον επιθυμητό μορφότυπο

