Περιγραφή Συλλογών Δεκαετίας ’40

Βίκη Ντρίτσου
Μονάδα Ψηφιακής Επιμέλειας, ΙΠΣΥ / Ερευνητικό Κέντρο Αθηνά
&
Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών

Μονάδα Ψηφιακής Επιμέλειας
ΙΠΣΥ - Ερευνητικό Κέντρο Αθηνά

απολλωνίς

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΕΠΕΝΔΥΣΕΩΝ
ΕΙΔΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΠΡΟΓΡΑΜΜΑΤΩΝ ΕΤΠΑ & ΤΣ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΑνΕΚ
ΕΠΑνΕΚ 2014-2020
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ ΚΑΙΝΟΤΟΜΙΑ
ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Χάρτης εργασιών Ψηφιακής Επιμέλειας

Σκοπός

- Ενοποίηση / ενιαία πρόσβαση
- Κοινή Ευρετηρίαση
- Εμπλουτισμός
- Ανάδειξη διασυνδέσεων πληροφοριακών πόρων, ακόμα και αν ανήκουν σε διαφορετικές αρχειακές συλλογές

Δομική & Σημασιολογική Ανομοιογένεια

- Απόκτηση: 6 ψηφιοποιημένες αρχειακές συλλογές
  - Διαφορετικά σχήματα τεκμηρίωσης για κάθε φορέα
  - Διαφορετικά σχήματα τεκμηρίωσης ακόμα και σε επιμέρους αρχεία του **ίδιου** φορέα
- Δομική ανομοιογένεια: διαφορετικά πεδία σε κάθε σχήμα
  - min 8 πεδία, max 53 πεδία
- Σημασιολογική ανομοιογένεια: πεδία με κοινούς τίτλους, που περιέχουν διαφορετικές σημασιολογικά πληροφορίες

ΑΣΚΙ

τίτλος αναγνωριστικό δημιουργός ημ/νία δημιουργίας
καλλιτέχνης φωτογράφος ημ/νία μετατροπής
τόπος προέλευση κατηγορία γλώσσα θέμα
μεταφραστής εκδότης συντάκτης επιμελητής
τόπος δημιουργίας τόπος έκδοσης ημ/νία έκδοσης
αποδέκτης αναφερόμενος εικονογράφος
διαστάσεις περιοδικό εφημερίδα υπότιτλος
περίληψη σημείωση αυθεντικότητα
φάκελος υποφάκελος αντίγραφα

Εβραϊκό Μουσείο

αναγνωριστικό
τίτλος
ημ/νία έκδοσης
προέλευση
τύπος
περιγραφή
θέμα
τόπος

Ενοποίηση συλλογών

- Ανάγκη υιοθέτησης **κοινού** σχήματος (εννοιολογικού μοντέλου)
  - Ομογένεια
  - Απαλοιφή αμφισημιών
- Προϋποθέτει λεπτομερή μελέτη των υφισταμένων σχημάτων ΚΑΙ των τιμών που τα πεδία περιέχουν
  - Ανάδειξη αμφισημιών, λανθασμένης ονοματολογίας πεδίων
- Βέλτιστη πρακτική: υιοθέτηση προτύπου / κοινώς αποδεκτού σχήματος
  - Διαλειτουργικότητα

Ενοποίηση συλλογών Δεκαετίας ‘40

- Μελέτη προτύπων/μοντέλων αναφοράς
- Υιοθέτηση του Europeana Data Model (EDM)[1]
  - Χρήση πρόσθετων πεδίων ως επεκτάσεις του
- Αντιστοίχιση πεδίων συλλογών σε πεδία EDM
  - Νοητική διεργασία
- Παραγωγή κώδικα για διατύπωση δηλώσεων επιμέρους συλλογών σύμφωνα με το νέο σχήμα
  - Ένα σενάριο για κάθε συλλογή
- Εκτέλεση κώδικα (έξοδος: πολλαπλά αρχεία XML)

[1]http://pro.europeana.eu/page/edm-documentation

EDM

http://pro.europeana.eu/page/edm-documentation

```xml
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="
http://purl.org/dc/terms/" xmlns:edm="http://www.europeana.eu/schemas/edm/" xmlns:ore="http://www.openarchives.org/ore/terms/"
xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:skos="
http://www.w3.org/2004/02/skos/core#" xmlns:apl="https://apollonis-infrastructure.gr/apl/">
  <edm:ProvidedCHO rdf:about="https://apollonis-infrastructure.gr/apl/ASKI_10083">
    <dc:identifier>ASKI_10083</dc:identifier>
    <dc:title>Γυναίκες ηθοποιοί στην Ελλάδα</dc:title>
    <dc:language xml:lang="el">Ελληνική</dc:language>
    <dc:language>el</dc:language>
    <rdfs:comment xml:lang="el">Αριθμός αντιτύπου: 1</rdfs:comment>
    <rdfs:comment xml:lang="el">Αριθμός τεκμηρίου: 4.877.103</rdfs:comment>
    <dcterms:issued>26/6/1947</dcterms:issued>
    <apl:archive xml:lang="el">Βιβλιοθήκη</apl:archive>
    <dc:type xml:lang="el">Κείμενα εφημερίδας</dc:type>
    <rdfs:comment xml:lang="el">Χωρίς διαβάθμιση</rdfs:comment>
    <dcterms:extent>2</dcterms:extent>
    <dc:source>ΠΟΛΙΤΙΣΜΟΣ ΚΑΙ ΑΡΙΣΤΕΡΑ</dc:source>
    <dc:source>Ελεύθερη Ελλάδα, Απογευματινή Εφημερίδα, Όργανον του Πολιτικού Συνασπισμού των κομμάτων του ΕΑΜ</dc:source>
    <dc:description xml:lang="el">Τεύχος: 851</dc:description>
    <apl:author>Στέφη Νίκη</apl:author>
    <dc:subject xml:lang="el">Παξινού Κατίνα</dc:subject>
    <apl:subtitle xml:lang="el">Επαγγελματικός μόχθος και καλλιτεχνικά όνειρα</apl:subtitle>
  </edm:ProvidedCHO>
  <ore:Aggregation rdf:about="https://apollonis-infrastructure.gr/apl/ASKI_10083">
    <edm:aggregatedCHO rdf:resource="https://apollonis-infrastructure.gr/apl/ASKI_10083"/>
    <edm:provider xml:lang="el">Αρχεία Σύγχρονης Κοινωνικής Ιστορίας</edm:provider>
    <edm:dataProvider>ΑΣΚΙ</edm:dataProvider>
    <edm:dataset xml:lang="el">Δεκαετία 1940</edm:dataset>
  </ore:Aggregation>
</rdf:RDF>
```


Date

Agent

Subject

Χάρτης εργασιών Ψηφιακής Επιμέλειας

Βάση Επαυξημένης Γνώσης

- Βάση γνώσης (knowledge base): Αναπαράσταση οντοτήτων συλλογών με χρήση διασυνδέσεων και σημασιολογικών μεταδεδομένων
  - Σκοπός η ολοκλήρωση της πληροφορίας, η διασύνδεση αντικειμένων διαφορετικών συλλογών και ο διαμοιρασμός τους
- Διαχωρισμός της πρωτογενούς πληροφορίας από την εξαγόμενη (κατόπιν εκτέλεσης εργασιών ψηφιακής επιμέλειας στα πρωτογενή δεδομένα)
  - Ευρετήρια υποδομής: πρωτογενής πληροφορία
  - Βάση Γνώσης: πρωτογενής + εξαγόμενη (πρόσθετη) = επαυξημένη πληροφορία
- Αφορμή για τις εργασίες αυτές ήταν οι πλούσιες τιμές σε συγκεκριμένα πεδία συλλογών (πεδία που εφαρμόστηκε ΕΦΓ)

Παράδειγμα πλούσιας τιμής: Συλλογή ΔΙΣ

ΤΗΛΕΓΡΑΦΗΜΑ ΤΟΥ ΑΡΗ ΒΕΛΟΥΧΙΩΤΗ ΤΗΣ 18/04/1944 ΠΡΟΣ ΤΗΝ ΠΟΛΙΤΙΚΗ ΕΠΙΤΡΟΠΗ ΕΘΝΙΚΗΣ ΑΠΕΛΕΥΘΕΡΩΣΗΣ ΚΑΙ ΤΟ ΓΕΝΙΚΟ ΣΤΡΑΤΗΓΕΙΟ, ΣΧΕΤΙΚΑ ΜΕ ΤΗ ΔΙΑΛΥΣΗ ΤΟΥ 5/42 ΣΥΝΤΑΓΜΑΤΟΣ ΕΥΖΩΝΩΝ ΚΑΙ ΤΟ ΘΑΝΑΤΟ ΤΟΥ ΣΥΝΤΑΓΜΑΤΑΡΧΗ ΨΑΡΡΟΥ ΚΑΙ ΤΟΥ ΤΑΓΜΑΤΑΡΧΗ ΚΑΠΕΤΣΩΝΗ (ΑΡΧΕΙΟ ΔΟΥΚΑ)

Παραλήπτης (Ομάδα) : ΠΟΛΙΤΙΚΗ ΕΠΙΤΡΟΠΗ ΕΘΝΙΚΗΣ ΑΠΕΛΕΥΘΕΡΩΣΗΣ

Παραλήπτης (Ομάδα) : ΓΕΝΙΚΟ ΣΤΡΑΤΗΓΕΙΟ

Ομάδα: 5/42 ΣΥΝΤΑΓΜΑ ΕΥΖΩΝΩΝ

Πρόσωπο: ΣΥΝΤΑΓΜΑΤΑΡΧΗΣ ΨΑΡΡΟΣ

Πρόσωπο: ΤΑΓΜΑΤΑΡΧΗΣ ΚΑΠΕΤΣΩΝΗΣ

Θέμα: ΣΧΕΤΙΚΑ ΜΕ ΤΗ ΔΙΑΛΥΣΗ ΤΟΥ 5/42 ΣΥΝΤΑΓΜΑΤΟΣ ΕΥΖΩΝΩΝ ΚΑΙ ΤΟ ΘΑΝΑΤΟ ΤΟΥ ΣΥΝΤΑΓΜΑΤΑΡΧΗ ΨΑΡΡΟΥ ΚΑΙ ΤΟΥ ΤΑΓΜΑΤΑΡΧΗ ΚΑΠΕΤΣΩΝΗ

Γεγονός: ΔΙΑΛΥΣΗ ΤΟΥ 5/42 ΣΥΝΤΑΓΜΑΤΟΣ ΕΥΖΩΝΩΝ

Γεγονός: ΘΑΝΑΤΟΣ ΤΟΥ ΣΥΝΤΑΓΜΑΤΑΡΧΗ ΨΑΡΡΟΥ

Γεγονός: ΘΑΝΑΤΟΣ ΤΟΥ ΤΑΓΜΑΤΑΡΧΗ ΚΑΠΕΤΣΩΝΗ

Ανάπτυξη Σχήματος Βάσης Γνώσης

- Εννοιολογική μοντελοποίηση
  - πρωτογενούς γνώσης
  - πρόσθετης γνώσης
- Σκοπός: δημιουργία ενιαίου σχήματος για την αναπαράσταση των πληροφοριακών πόρων
- Μελέτη μοντέλων αναφοράς
- Αναπτύχθηκε μοντέλο συμβατό με το πρότυπο CIDOC CRM[2]
  - Πρότυπο ολοκλήρωσης πληροφορίας στο πεδίο της πολιτιστικής κληρονομιάς
  - Γιατί όχι υιοθέτηση του CIDOC CRM;
    - Ανάγκη αναπαράστασης ενδιαφέρουσας πληροφορίας (από τη σκοπιά του ιστορικού) αρχειακών τεκμηρίων

[2]http://www.cidoc-crm.org/

E55 Type
P127 has broader (or narrower) term
Genre Form
Access Restrict Type
E22 Man-Made Object
E31 Document
E33 Linguistic Object
Subject
has subject
P46 is composed of
E57 Material
has_genre_form_type
has_access_restrict_type
Storage Medium
has_storage_medium_type
Authentic Record
P67 refers to/ P70 documents
E1 CRM ENTITY
P2 has type
E55 Type
P1 is identified by
E41 Apellation
P3 has note
E62 String
P89 falls within
P122 borders with
E53 Place
P121 overlaps with
P2 has type
E55 Type
P87 Is identified by
E41 Appellation
P53 has former or current location
P7 took place at
P72 has language
E56 Language
P102 has title
E35 Title
P1 is identified by
E42 Identifier
has id physical record
Id Physical Record
P46 is composed of
E78 Collection
Analogue
Born Digital
Digitized
P108 has produced
E53 Place
P94 has created
Creation
P7 took place at
E12 Production
P109 has current or former curator
P49 has former or current keeper, P51 has former or current owner
E41 Appellation
P1 is identified by
E52 Time Span
E55 Type
in_time-span
Has Role In Time-Span
in_role
Role
has_role in_time-span
E41 Apellation
P76 has contact point
P4 has time span
P14 carried out by
E52 TimeSpan
carried out by as author/ has participant as recipient
E39 Actor
P11 had participant/ P12 occurred in the presence of
E5 Event
P9 consists of
E4 Period
AP28 occurs before
P4 has time span
E52 TimeSpan
P89 falls within (contains)
AP25 occurs during
P2 has type
E55 Type
P74 has current or former residence
P152 has parent
Person
has date of death
E41 Apellation
has date of birth
E41 Apellation
has sex
E62 String
P107 has current or former member
E74 Group
P2 has type
E55 Type
P143 joined
P145 separated
E52 Time Span
P4 has time-span
E 85 Joining
P144 joined with
P146 seperated from
E86 Leaving
P4 has time-span
E52 Time Span

Οφέλη Χρήσης Σχήματος Βάσης Γνώσης

- Μοντελοποίηση δεδομένων βάσει σχήματος
- Πλούσια (σημασιολογικά) αναπαράσταση διαθέσιμης πληροφορίας
- Διασύνδεση εγγραφών διαφορετικών συλλογών
- Δυνατότητα διατύπωσης επερωτήσεων μεγάλης πολυπλοκότητας

ΤΗΛΕΓΡΑΦΗΜΑ ΤΟΥ ΑΡΗ ΒΕΛΟΥΧΙΩΤΗ ΤΗΣ 18/04/1944 ΠΡΟΣ ΤΗΝ ΠΟΛΙΤΙΚΗ ΕΠΙΤΡΟΠΗ ΕΘΝΙΚΗΣ ΑΠΕΛΕΥΘΕΡΩΣΗΣ ΚΑΙ ΤΟ ΓΕΝΙΚΟ ΣΤΡΑΤΗΓΕΙΟ, ΣΧΕΤΙΚΑ ΜΕ ΤΗ ΔΙΑΛΥΣΗ ΤΟΥ 5/42 ΣΥΝΤΑΓΜΑΤΟΣ ΕΥΖΩΝΩΝ ΚΑΙ ΤΟ ΘΑΝΑΤΟ ΤΟΥ ΣΥΝΤΑΓΜΑΤΑΡΧΗ ΨΑΡΡΟΥ ΚΑΙ ΤΟΥ ΤΑΓΜΑΤΑΡΧΗ ΚΑΠΕΤΣΩΝΗ (ΑΡΧΕΙΟ ΔΟΥΚΑ)

Event
refers to
is a
Subject
Authentic Record
Creation
Date
has subject
has created
has
time-span
author /
recipient
Group
refers to
Actor
Person
Δημιουργία
τηλ. 8/929/Γ/6β
Θάνατος
Συνταγματάρχη Ψαρρού
Τηλεγράφημα
8/929/Γ/6β
Άρης
Βελουχιώτης
18/04/1944
has subject
Διάλυση του 5/42
Συντάγματος Ευζώνων
has subject
Γενικό
Στρατηγείο
Πολιτική Επιτροπή
Εθνικής Απελευθέρωσης

Event
refers to
Subject
has subject
Authentic Record
refers to
Group
Actor
has
member
Person
has role in
time-span
Role in
time-span
in time-span
in role
Role
Time-span
Άρης
Βελουχιώτης
Θάνατος
Ψαρρός
Συνταγματάρχης
04/1944
5/42 Σύνταγμα
Ευζώνων

Χωρίς τίτλο

dc:description	Ο Myers τηλεγραφεί ότι δυνάμεις του Ε.Λ.Α.Σ. υπό τον Αρη Βελουχιώτη αφόπλισαν τα τμήματα ανταρτών του Ψαρρού. [Ο Βελουχιώτης αναλαμβάνει Καπετάνιος του Γενικού Στρατηγείου του Ε.Λ.Α.Σ. που συγκροτείται στις 19 Μαΐου]. τηλ. 204, 205, άκρως απόρρ., 13 Μαΐου 1943
dc:language	el
dc:subject	Ε.Λ.Α.Σ. Α. Βελουχιώτης (Άρης) Καπετάνιος Γενικού Στρατηγείου Ε.Λ.Α.Σ. Myers 5/42 Σύνταγμα Ευζώνων Ε.Κ.Κ.Α.
edm:dataProvider	Ακαδημία Αθηνών
dcterms:created	13/5/1943
apl:receiver	S.O.E. Κάιρο
edm:type	TEXT
edm:provider	Ακαδημία Αθηνών
edm:dataset	Δεκαετία 1940

Παραγωγή RDF δεδομένων

- Για τη διατύπωση της πληροφορίας χρησιμοποιήθηκε το W3C πρότυπο RDF (Resource Description Framework)[2]
  - Δηλώσεις με δομή τριπλέτας (RDF triples): subject – predicate – object
- Διατύπωση σχήματος σε RDF
- Προγραμματιστική μετατροπή της πρωτογενούς και εξαγόμενης πληροφορίας σε RDF τριπλέτες
- Δημιουργία βάσης επαυξημένης γνώσης

[1] https://www.w3.org/RDF/

Βάση Επαυξημένης Γνώσης Δεκαετίας 1940

Ενσωμάτωση παραγόμενης γνώσης

Συσχέτιση πόρων που προέρχονται από διαφορετικές συλλογές

Ελευθερία διατύπωσης πολύπλοκων επερωτήσεων

Δυσκολία διατύπωσης επερωτήσεων από μη εξοικειωμένους χρήστες

Alpha version βάσης γνώσης (blazegraph)

Υποβοήθηση μη εξοικειωμένου χρήστη (Ι)

- Με σκοπό την υποβοήθηση του μη εξοικειωμένου χρήστη, καθορίστηκαν 14 συχνές/ενδιαφέρουσες επερωτήσεις
  - Από την οπτική του ιστορικού
- Διατύπωση αυτών στη γλώσσα επερωτήσεων SPARQL
- Διαμοιρασμός του συνόλου των 14 επερωτήσεων με τους χρήστες της υποδομής

```
PREFIX apl: <https://services.apollonis-infrastructure.gr/decade1940#>
PREFIX cidoc: <http://www.cidoc-crm.org/cidoc-crm/>
SELECT ?identifier ?number ?location ?title
WHERE {
    ?ar cidoc:P1_is_identified_by ?identifier.
    ?ar cidoc:P55_has_current_location ?number.
    ?number cidoc:P89_falls_within ?location.
    ?ar cidoc:P102_has_title ?title.
    ?ar cidoc:P128_carries ?text.
    ?text cidoc:P67_refers_to ?activity.
    ?activity cidoc:P117_occurs_during    apl:Περίοδος_Β_Παγκοσμίου_Πολέμου.
}
```

```
PREFIX apl: <https://services.apollonis-infrastructure.gr/decade1940#>
PREFIX cidoc: <http://www.cidoc-crm.org/cidoc-crm/>
SELECT ?identifier ?number ?location ?period ?title
WHERE {
    ?ar cidoc:P1_is_identified_by ?identifier.
    ?ar cidoc:P55_has_current_location ?number.
    ?number cidoc:P89_falls_within ?location.
    ?ar cidoc:P102_has_title ?title.
    ?ar cidoc:P128_carries ?text.
    ?text cidoc:P67_refers_to ?period.
    ?period cidoc:P120_occurs_before apl:Περίοδος_Εμφυλίου_Πολέμου
}
```

Συχνές Επερωτήσεις (σε φυσική γλώσσα)

Αναζήτησε τεκμήρια:

1. συγκεκριμένου τύπου αρχείου (π.χ. φωτογραφίες)
2. ψηφιοποιημένα ή αμιγώς ψηφιακά ή αναλογικά
3. που ανήκουν σε συγκεκριμένο φορέα
4. συγκεκριμένου είδους (π.χ. στρατιωτικά έγγραφα, διατριβές)
5. παραγόμενα από συγκεκριμένο παραγωγό ή από συγκεκριμένο τύπο φορέα
6. που έχουν δημιουργηθεί σε συγκεκριμένη χρονική περίοδο
7. βάσει τίτλου
8. που περιέχουν κείμενο συγκεκριμένου δράστη
9. βάσει θέματος
10. διατυπωμένα σε συγκεκριμένη φυσική γλώσσα
11. που φέρουν πληροφορία αναφορικά με συγκεκριμένο δράστη
12. που φέρουν πληροφορία για συγκεκριμένη τοποθεσία
13. που φέρουν πληροφορία για ενέργειες συγκεκριμένου τύπου (π.χ. στρατιωτικές επιχειρήσεις)
14. που φέρουν πληροφορία για συγκεκριμένη χρονική περίοδο

Υποβοήθηση μη εξοικειωμένου χρήστη (ΙΙ)

- Λογισμικά/διεπαφές ανοιχτού κώδικα, που επιτρέπουν την εκτέλεση επερωτήσεων μέσα από γραφικό μενού
  - Ο χρήστης επιλέγει από λίστες τιμών την επερώτηση που επιθυμεί να εκτελέσει
  - Η SPARQL επερώτηση εκτελείται στο υπόβαθρο, χωρίς να είναι εμφανής στον χρήστη
  - Τα αποτελέσματα επιστρέφονται σε φιλική μορφή (γραφική απεικόνιση, πίνακες κλπ.)
- Αξιοποίηση λογισμικού γραφικής αναζήτησης στη βάσης γνώσης για την περίπτωση της Δεκαετίας ’40
  - Εν εξελίξει
  - Αναμενόμενη δημοσιοποίηση: Ιούνιος 2021

Σας ευχαριστώ!

Co-financed by Greece and the European Union