

ΔΙΑΧΕΙΡΙΣΗ, ΕΠΙΜΕΛΕΙΑ ΚΑΙ ΕΚΔΟΣΗ ΨΗΦΙΑΚΩΝ ΠΟΡΩΝ

Βίκη Ντρίτσου

ΠΜΣ Ψηφιακές Μέθοδοι στις Ανθρωπιστικές Επιστήμες

Οικονομικό Πανεπιστήμιο Αθηνών

2025-2026

Ροές εργασίας

Μέσω μελέτης περίπτωσης Δεκαετίας '40

”

Workflow: the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules.”

P. Lawrence, editor. Workflow Handbook 1997,
Workflow Management Coalition. John Wiley and Sons,
New York, 1997.

Ροή εργασίας

- Επαναλαμβανόμενο και επαναχρησιμοποιήσιμο μοτίβο εκτέλεσης μιας εργασίας
- Συστηματική οργάνωση βημάτων που περιέχουν
 - διεργασίες
 - εισόδους
 - εξόδους
 - σχέσεις ανάμεσα σε διεργασίες
- Μπορεί να θεωρηθεί ως μια αφαιρετική αναπαράσταση μιας πραγματικής εργασίας
- Αντικείμενό μας εδώ: Ροές εργασιών ψηφιακής επιμέλειας

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.

ΑΠΟΛΛΩΝΙΣ

Εθνική υποδομή που υποστηρίζει και προωθεί τις ψηφιακές ανθρωπιστικές επιστήμες και τέχνες, και τη γλωσσική τεχνολογία και καινοτομία στην Ελλάδα.

ΠΥΛΗ ΥΠΗΡΕΣΙΩΝ ΑΠΟΛΛΩΝΙΣ

- Πύλη πρόσβασης σε υπηρεσίες, εργαλεία και πόρους
- Δυνατότητα δημιουργίας συλλογών
- Υποστήριξη ευρετηρίων
- Πιλοτική δράση για τη Δεκαετία 1940



Σχετικά με τη δράση

Αναζητήστε αρχειακά τεκμήρια της δεκαετίας του 1940 από τις συλλογές έξι διαφορετικών οργανισμών και ανακαλύψτε πρόσωπα, τόπους, χρόνους και γεγονότα της ταραγμένης αυτής εποχής.

Πάροχοι περιεχομένου

Φορέας	Αριθμός τεκμηρίων
Ακαδημία Αθηνών	4.868
Αρχεία Σύγχρονης Κοινωνικής Ιστορίας	9.766
Ανωτάτη Σχολή Καλών Τεχνών	287
Διεύθυνση Ιστορίας Στρατού	76.979
Εβραϊκό Μουσείο Ελλάδας	13
Ιστορικό Αρχείο ΕΚΠΑ	373

Χαρακτηριστικά περιεχομένου

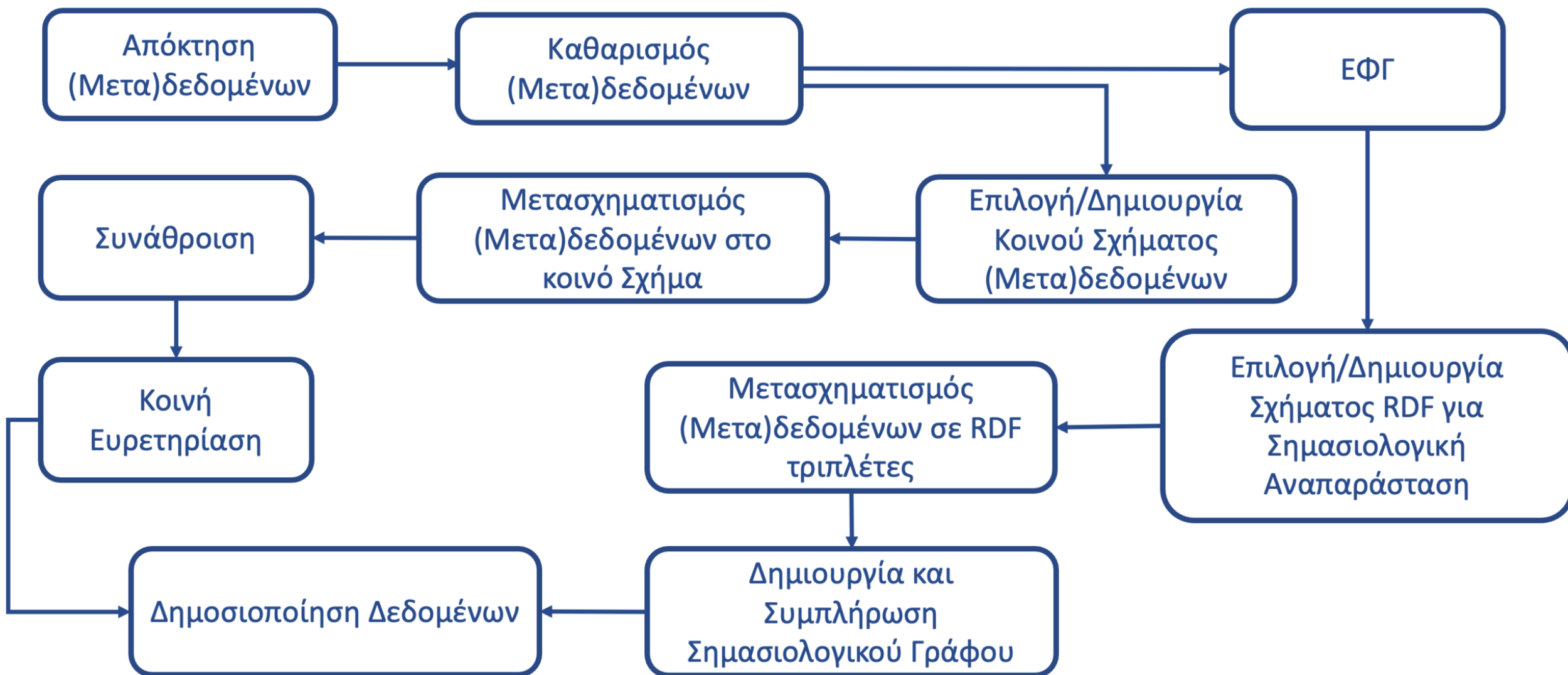
Προκλήσεις:

- διαφορετική συντακτική οργάνωση πληροφορίας (μορφότυπος)
- διαφορετική σημασιολογική οργάνωση πληροφορίας (σχήματα τεκμηρίωσης)
- χρήση διαφορετικών λεξιλογίων / ορολογίας
- τυπογραφικά και ορθογραφικά σφάλματα
- έλλειψη τυποποίησης στη συμπλήρωση των πεδίων μεταδεδομένων

Σκοπός: δημιουργία ενιαίας συλλογής

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

Χάρτης Εργασιών Ψηφιακής Επιμέλειας



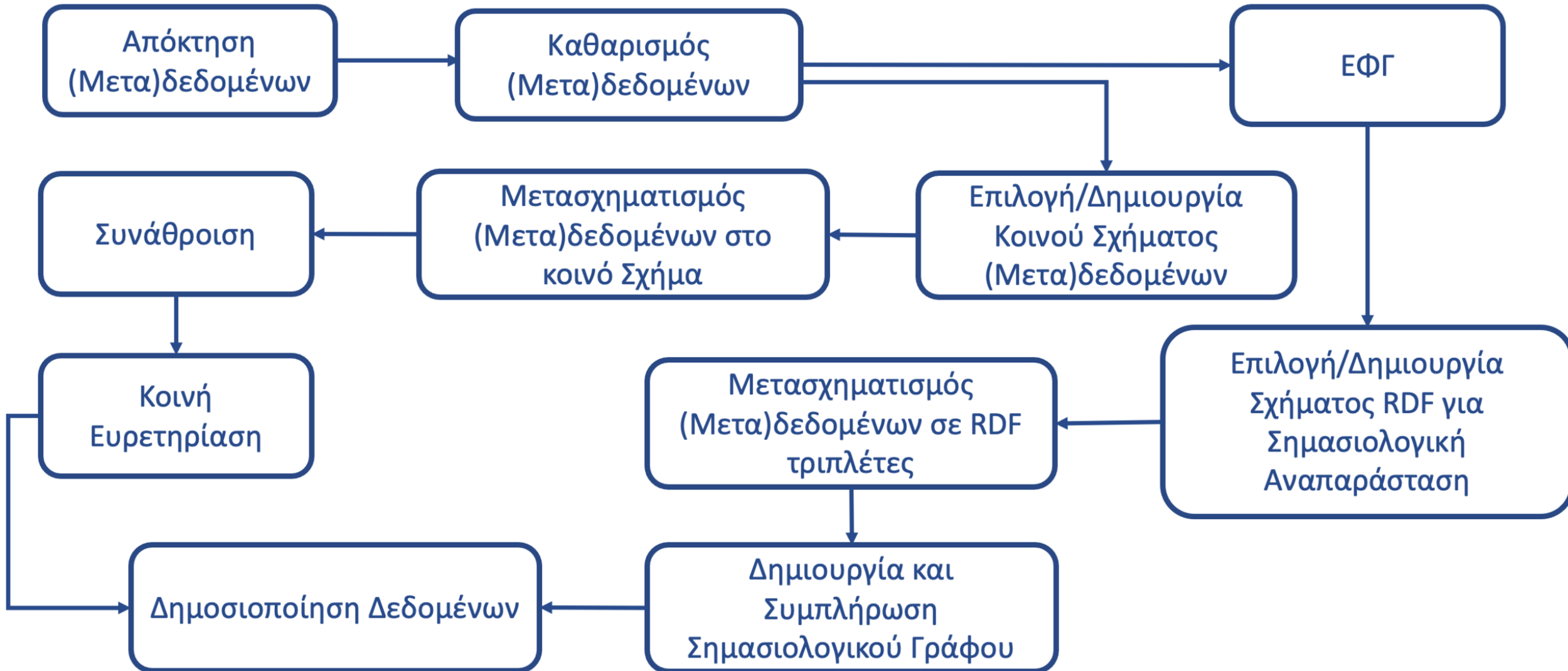
Πρόσκτηση αρχείων (Acquisition)

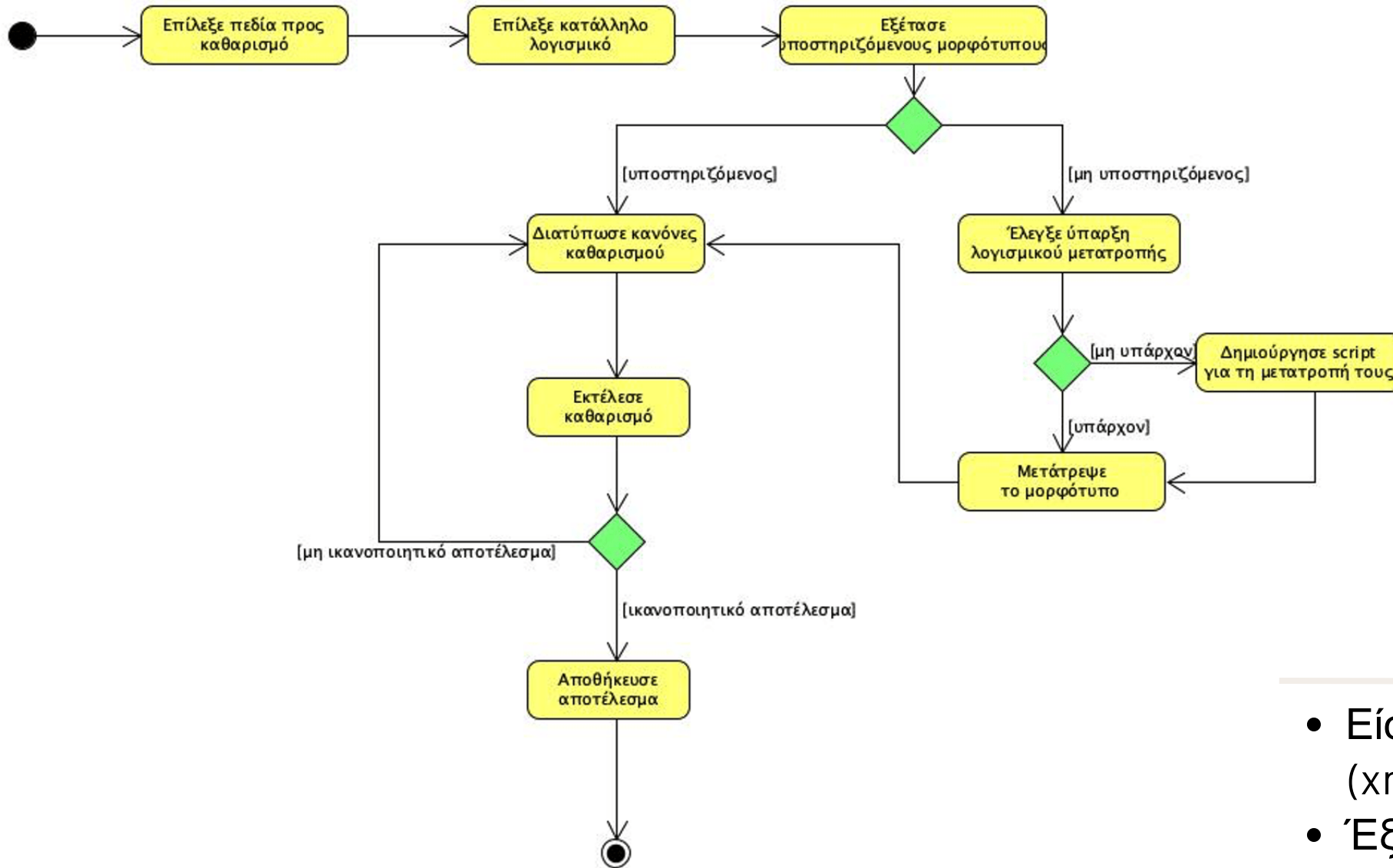
- Πρόσκτηση αρχείων από φορείς
 - Μορφότυπος
 - Ακαδημία Αθηνών: βάση δεδομένων Microsoft Access
 - Αρχεία Σύγχρονης Κοινωνικής Ιστορίας: xml
 - Ανωτάτη Σχολή Καλών Τεχνών: json
 - Διεύθυνση Ιστορίας Στρατού: xml
 - Εβραϊκό Μουσείο Ελλάδος: html
 - Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών: json

Πρόσκτηση αρχείων (Acquisition) II

- Πρόσκτηση αρχείων από φορείς
 - Μελέτη μορφής και μετατροπή μορφότυπου
 - ΑΑ: Εξαγωγή δεδομένων σε csv μέσω εκτέλεσης κατάλληλων SQL ερωτήσεων
 - ΑΣΚΙ: Ενοποίηση πεδίων ανά εγγραφή – μη αυτόματη διεργασία, δημιουργία script
 - ΑΣΚΤ, ΔΙΣ, Εβραϊκό Μουσείο, ΕΚΠΑ : Αυτόματη μετατροπή δεδομένων σε csv μέσω λογισμικού

Χάρτης Εργασιών Ψηφιακής Επιμέλειας





Καθαρισμός δεδομένων

- Είσοδος: αρχεία συλλογών (xml, xls, tsv, json)
- Έξοδος: «καθαρισμένα» αρχεία συλλογών

Επαναλαμβανόμενα Πεδία - Τεμαχισμός

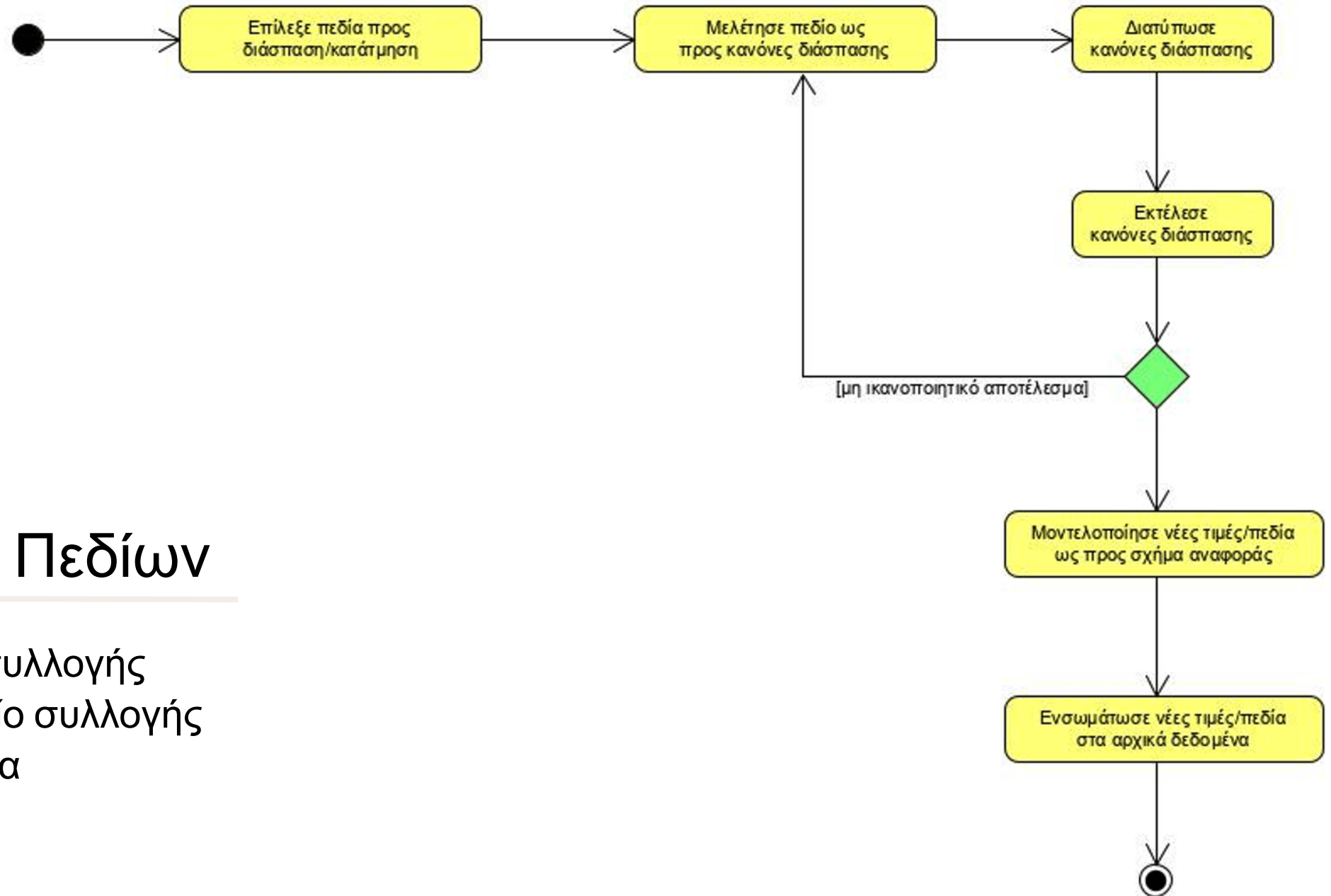
ID	Subject
AA_1207	Ε.Λ.Α.Σ., Α. Βελουχιώτης (Αρης) Καπετάνιος Γενικού Στρατηγείου Ε.Λ.Α.Σ., Myers, 5/42 Σύνταγμα Ευζώνων

Επαναλαμβανόμενα Πεδία - Τεμαχισμός

ID	Subject
AA_1207	Ε.Λ.Α.Σ., Α. Βελουχιώτης (Αρης) Καπετάνιος Γενικού Στρατηγείου Ε.Λ.Α.Σ., Myers, 5/42 Σύνταγμα Ευζώνων



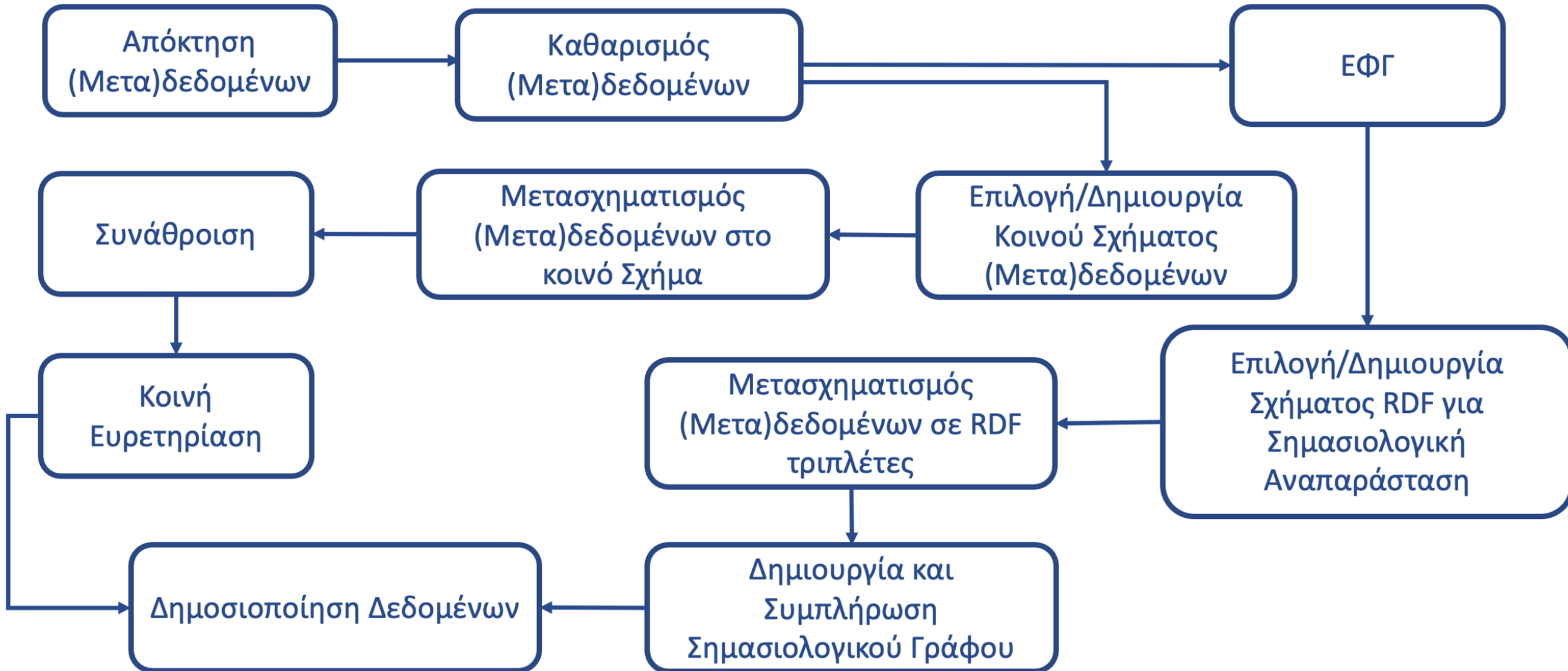
ID	Subject
AA_1207	Ε.Λ.Α.Σ.
AA_1207	Α. Βελουχιώτης (Αρης) Καπετάνιος Γενικού Στρατηγείου Ε.Λ.Α.Σ.
AA_1207	Myers
AA_1207	5/42 Σύνταγμα Ευζώνων



Τεμαχισμός Πεδίων

- Είσοδος: αρχείο συλλογής
- Έξοδος: νέο αρχείο συλλογής με πρόσθετα πεδία

Χάρτης Εργασιών Ψηφιακής Επιμέλειας



Δομική και Σημασιολογική Ανομοιογένεια

- Πρόσκτηση: 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

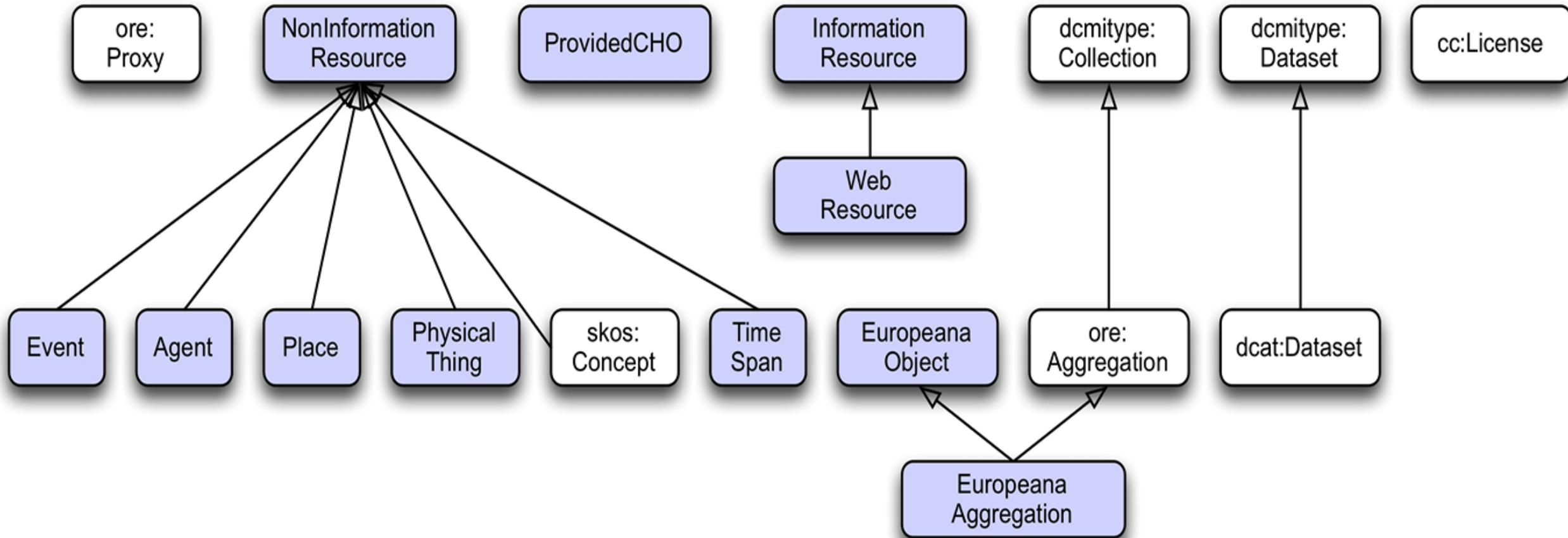
- Ανάγκη υιοθέτησης κοινού σχήματος (εννοιολογικού μοντέλου)
 - Ομογένεια
 - Απαλοιφή αμφισημιών
- Προϋποθέτει λεπτομερή μελέτη των υφισταμένων σχημάτων ΚΑΙ των τιμών που τα πεδία περιέχουν
 - Ανάδειξη αμφισημιών, λανθασμένης ονοματολογίας πεδίων
- Βέλτιστη πρακτική: υιοθέτηση προτύπου / κοινώς αποδεκτού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



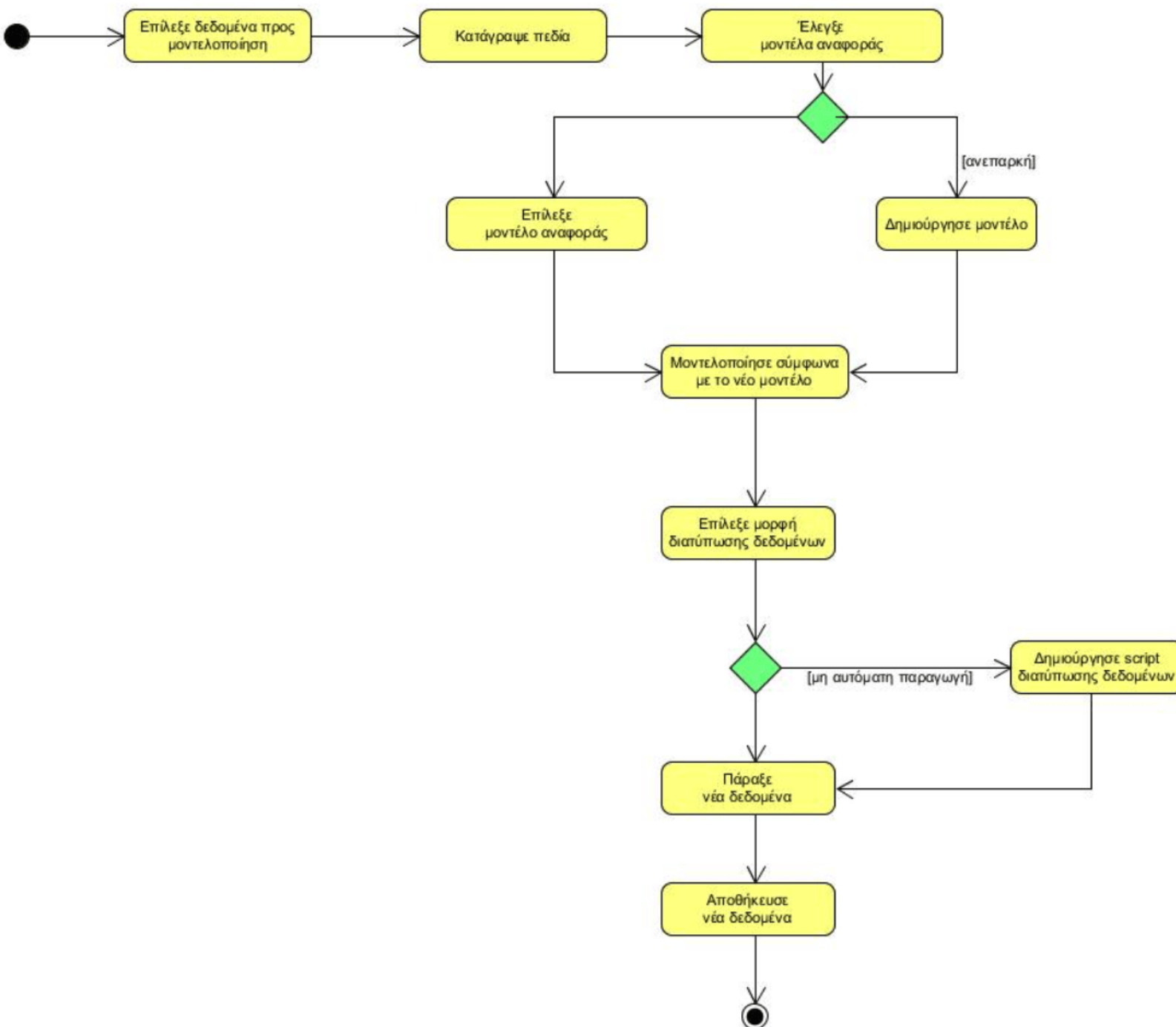
```

<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">Αρχαία Σύγχρονης Κοινωνικής Ιστορίας</edm:provider>
    <edm:dataProvider>ΑΣΚΙ</edm:dataProvider>
    <edm:dataset xml:lang="el">Δεκαετία 1940</edm:dataset>
  </ore:Aggregation>
</rdf:RDF>

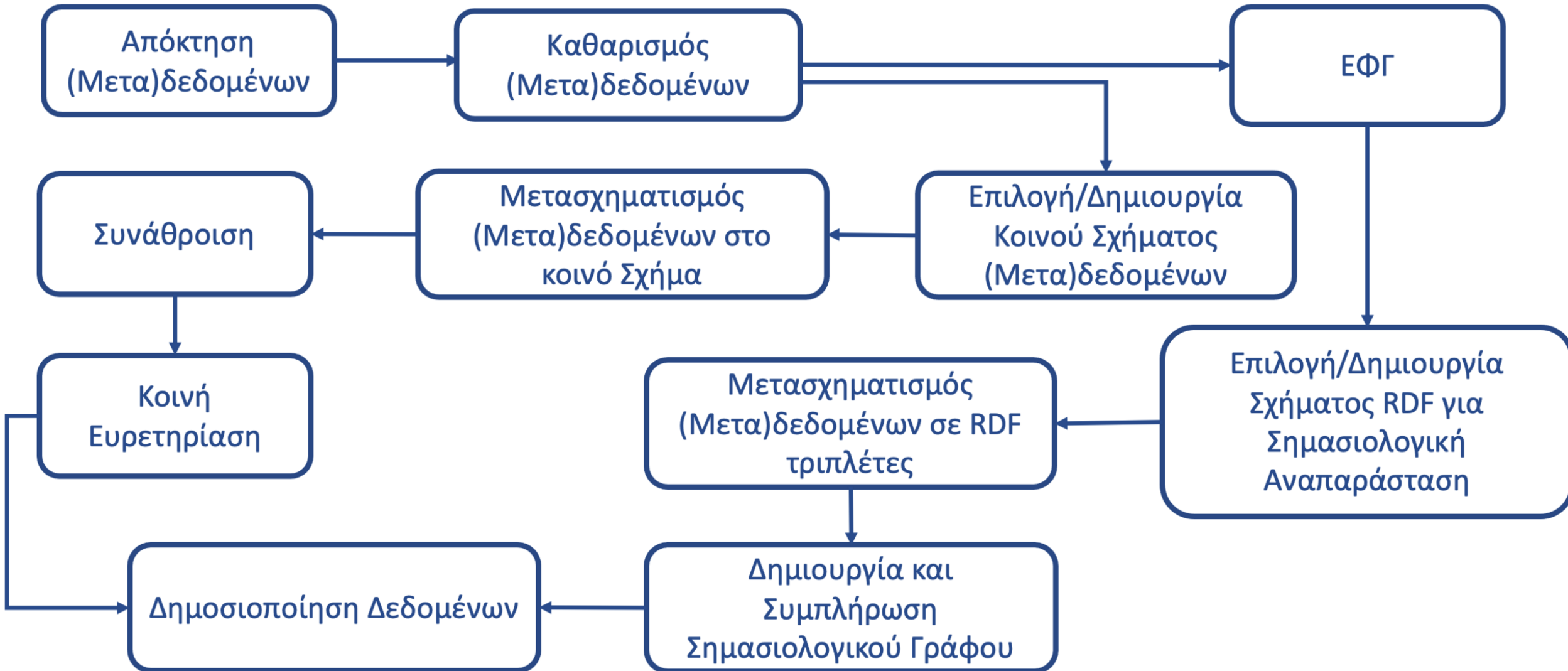
```

Εννοιολογική μοντελοποίηση

- Μη αυτόματη διεργασία
 - Απαιτεί νοητική εργασία
- Δεκαετία '40: EDM σχήμα
 - Εμπλουτισμένο
- Είσοδος: πολλαπλά xml σχήματα
- Έξοδος: ενιαίο xml σχήμα + αντιστοιχίσεις

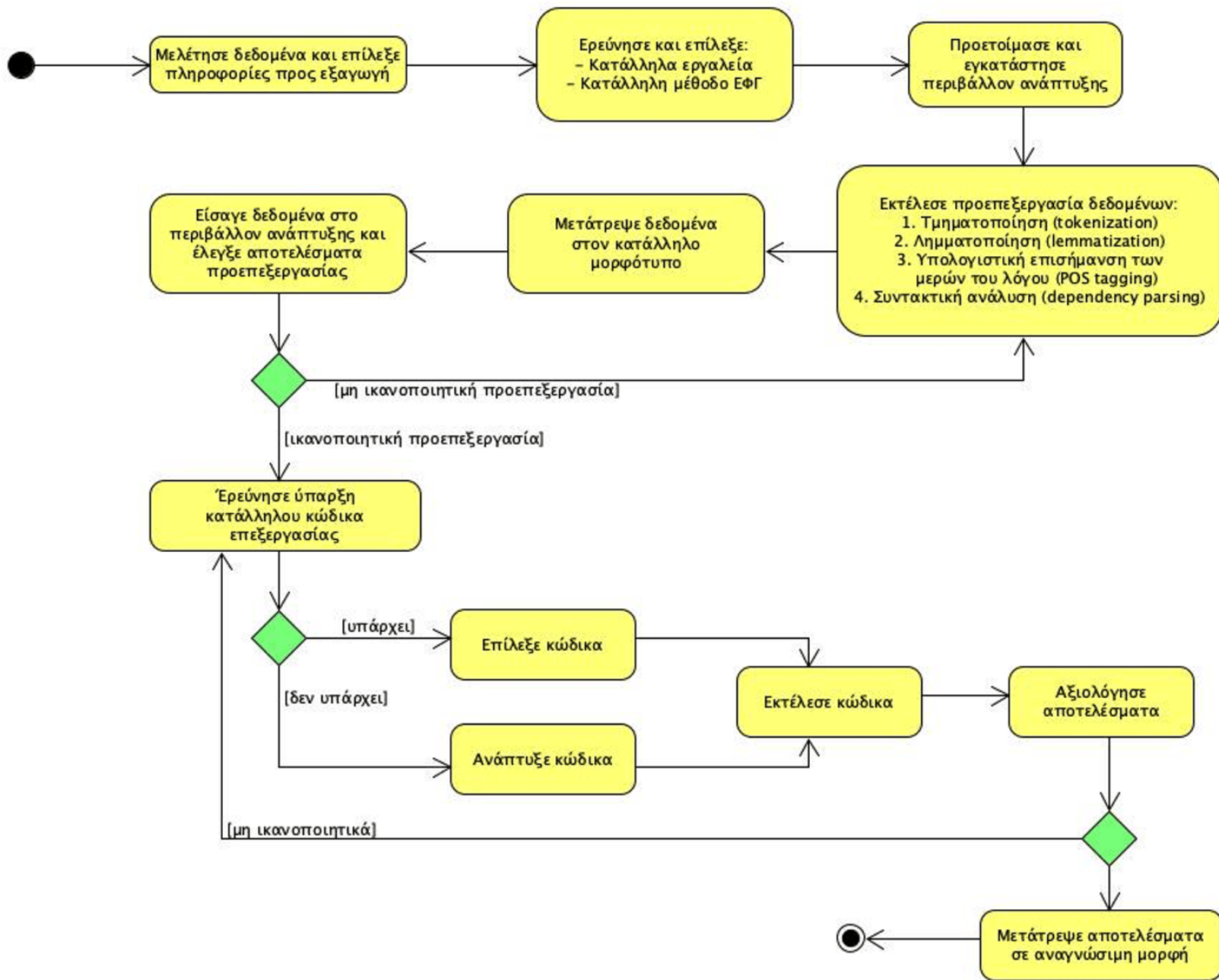


Χάρτης Εργασιών Ψηφιακής Επιμέλειας



Επεξεργασία Φυσικής Γλώσσας

- Σύμπλοκα πεδία (υπό το πρίσμα του περιεχομένου τους)
 - Περιέχουν τιμές άλλων πεδίων
 - Παραδείγματα
 - Πεδίο ‘Τίτλος’ της ΔΙΣ, πεδίο ‘Περιγραφή’ της ΑΑ
 - Εξαγωγή πληροφορίας με χρήση Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing – NLP)
 - Η εξαγόμενη πληροφορία χρήζει μοντελοποίησης εκ νέου

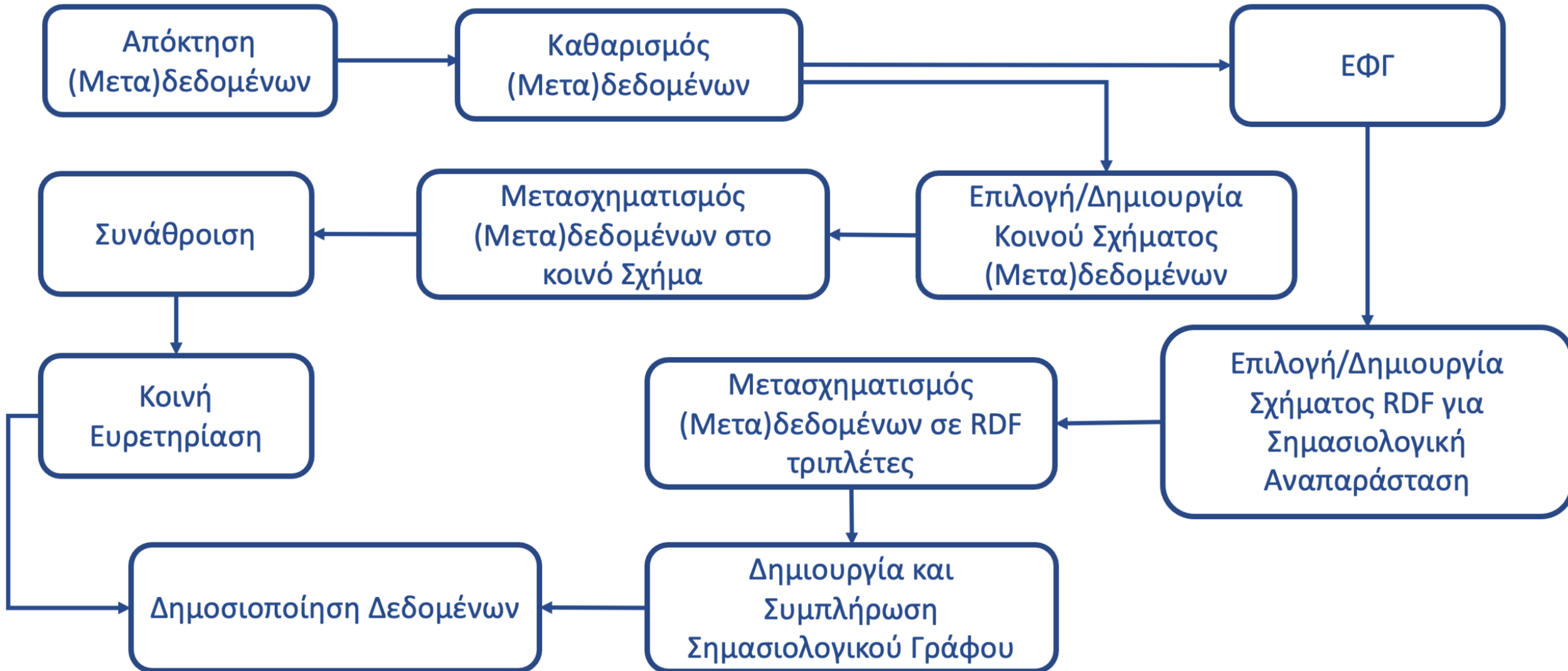


Επεξεργασία Φυσικής Γλώσσας (NLP)

Αποτελέσματα ΕΦΓ στη Συλλογή ΔΙΣ

- Επιτυχής αναγνώριση των οντοτήτων:
 - Δράστες – 15.723 επισημειώσεις (Πρόσωπα, Ομάδες και Στρ. Μονάδες)
 - Θέματα – 3.851 επισημειώσεις
 - Ημερομηνίες – 4.545 επισημειώσεις
 - Τόποι – 2.125 επισημειώσεις
 - Τύποι εγγράφων – 9.949 επισημειώσεις
- Οι επισημειώσεις αυτές θεωρήθηκαν πρόσθετη γνώση
 - Επιλέχθηκε να οργανωθεί σε ξεχωριστή δομή
- Οργάνωση & αναπαράσταση πρόσθετης γνώσης σε σημασιολογικό γράφο
 - Βάση γνώσης: πληροφορία οργανωμένη κάτω από κατάλληλο σχήμα (οντολογία)

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

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

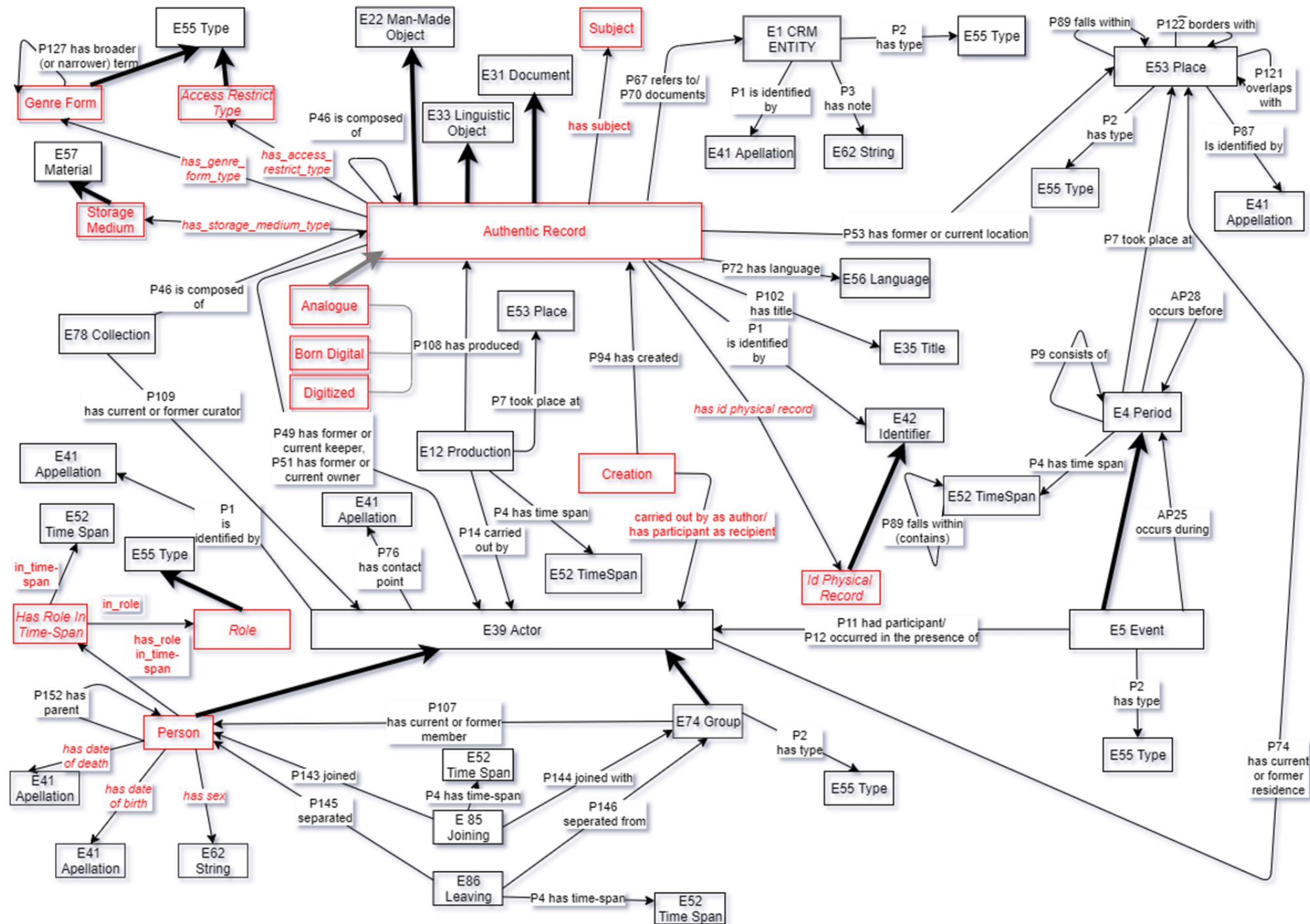
Γεγονός: ΔΙΑΛΥΣΗ ΤΟΥ 5/42 ΣΥΝΤΑΓΜΑΤΟΣ ΕΥΖΩΝΩΝ

Γεγονός: ΘΑΝΑΤΟΣ ΤΟΥ ΣΥΝΤΑΓΜΑΤΑΡΧΗ ΨΑΡΡΟΥ

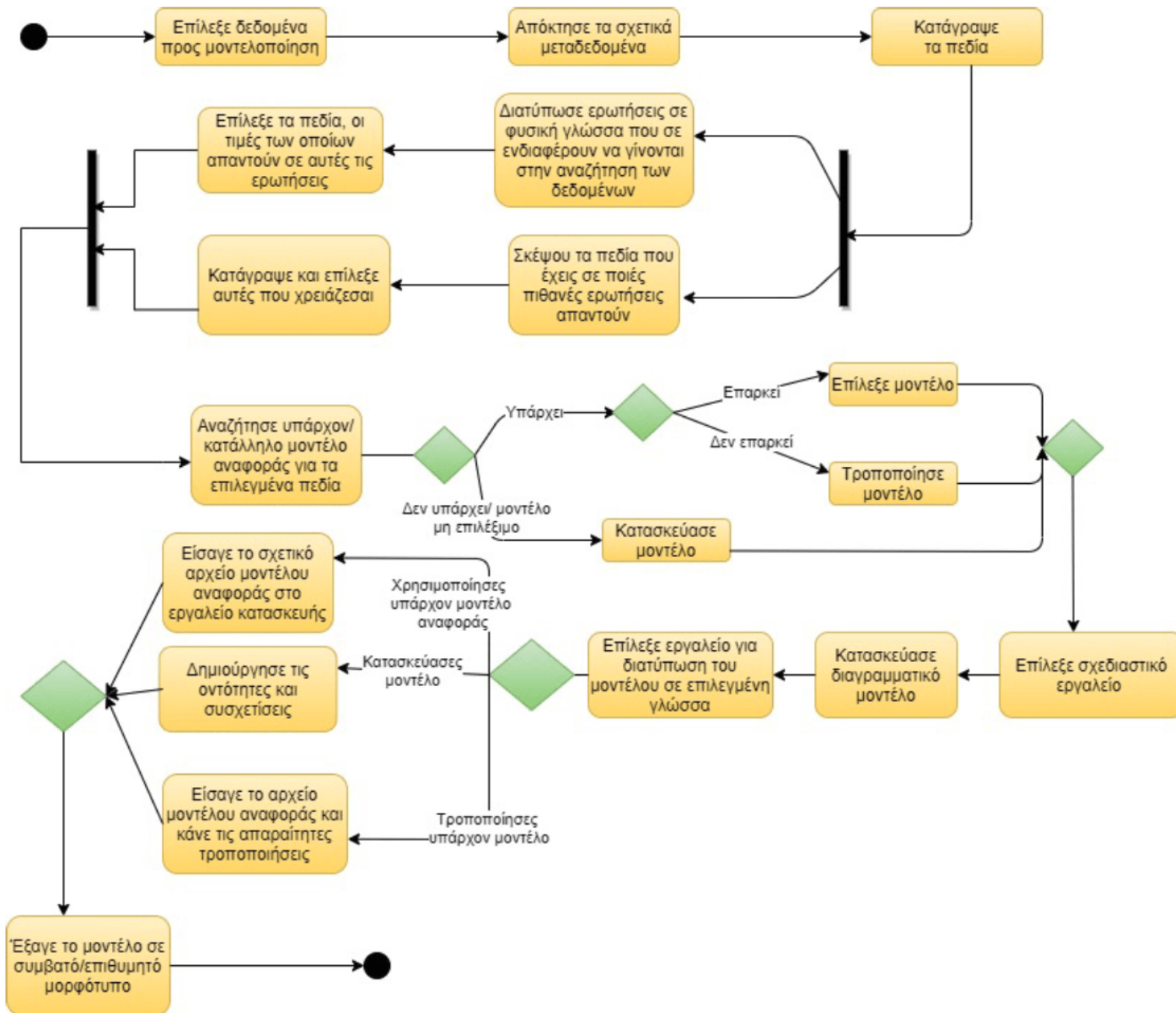
Γεγονός: ΘΑΝΑΤΟΣ ΤΟΥ ΤΑΓΜΑΤΑΡΧΗ ΚΑΠΕΤΣΩΝΗ

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](#)
 - 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



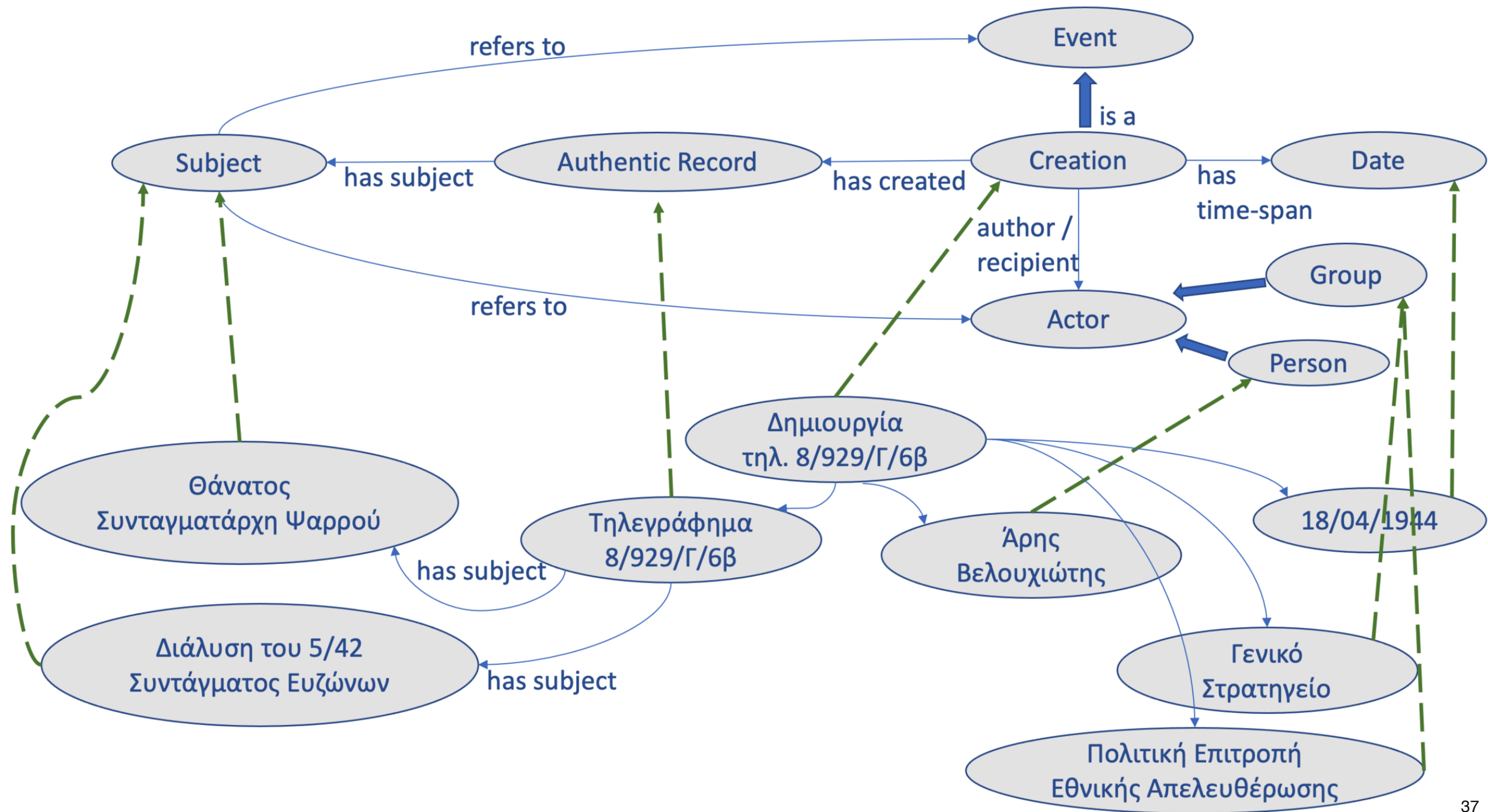
Εννοιολογική Μοντελοποίησης Σχήματος Βάσης Γνώσης

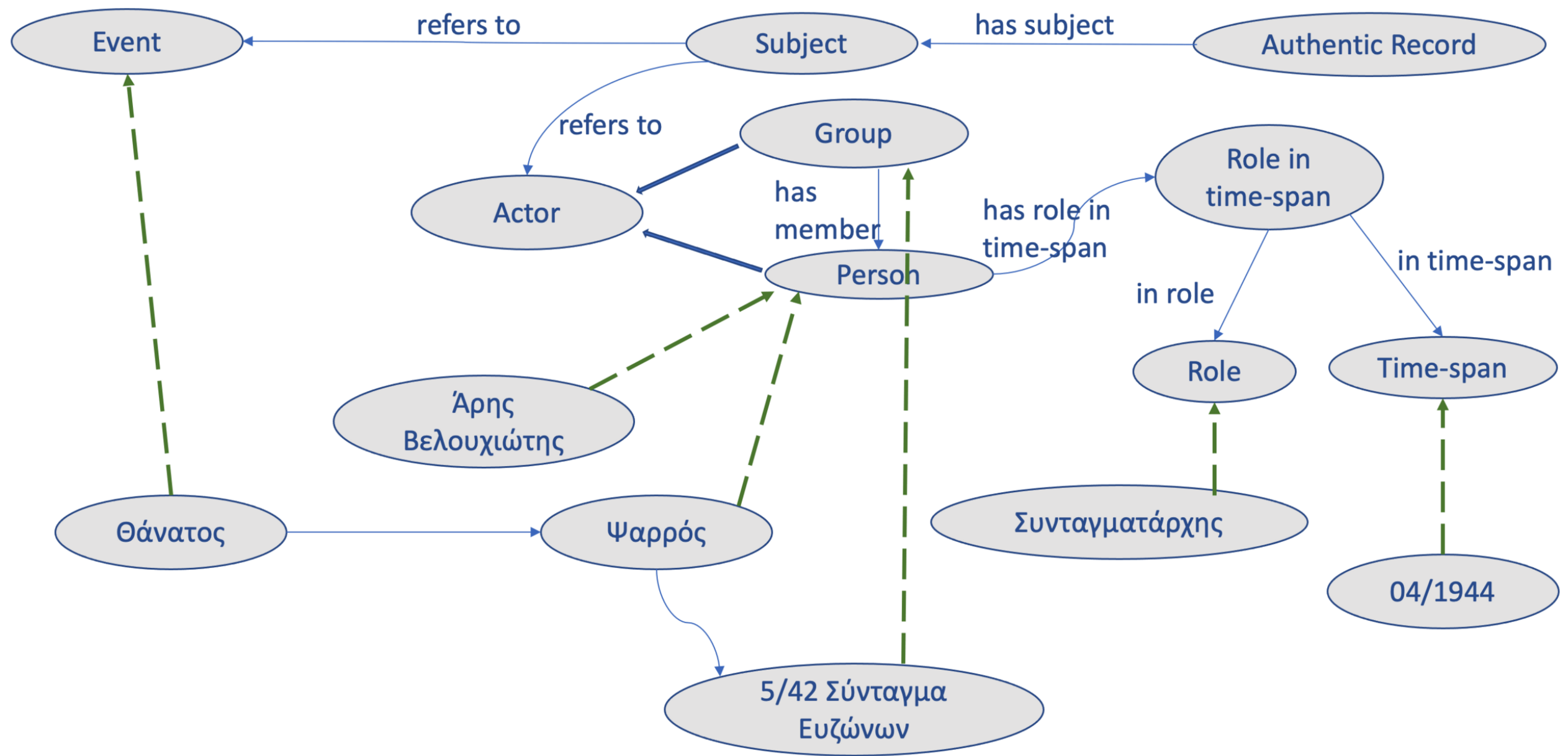


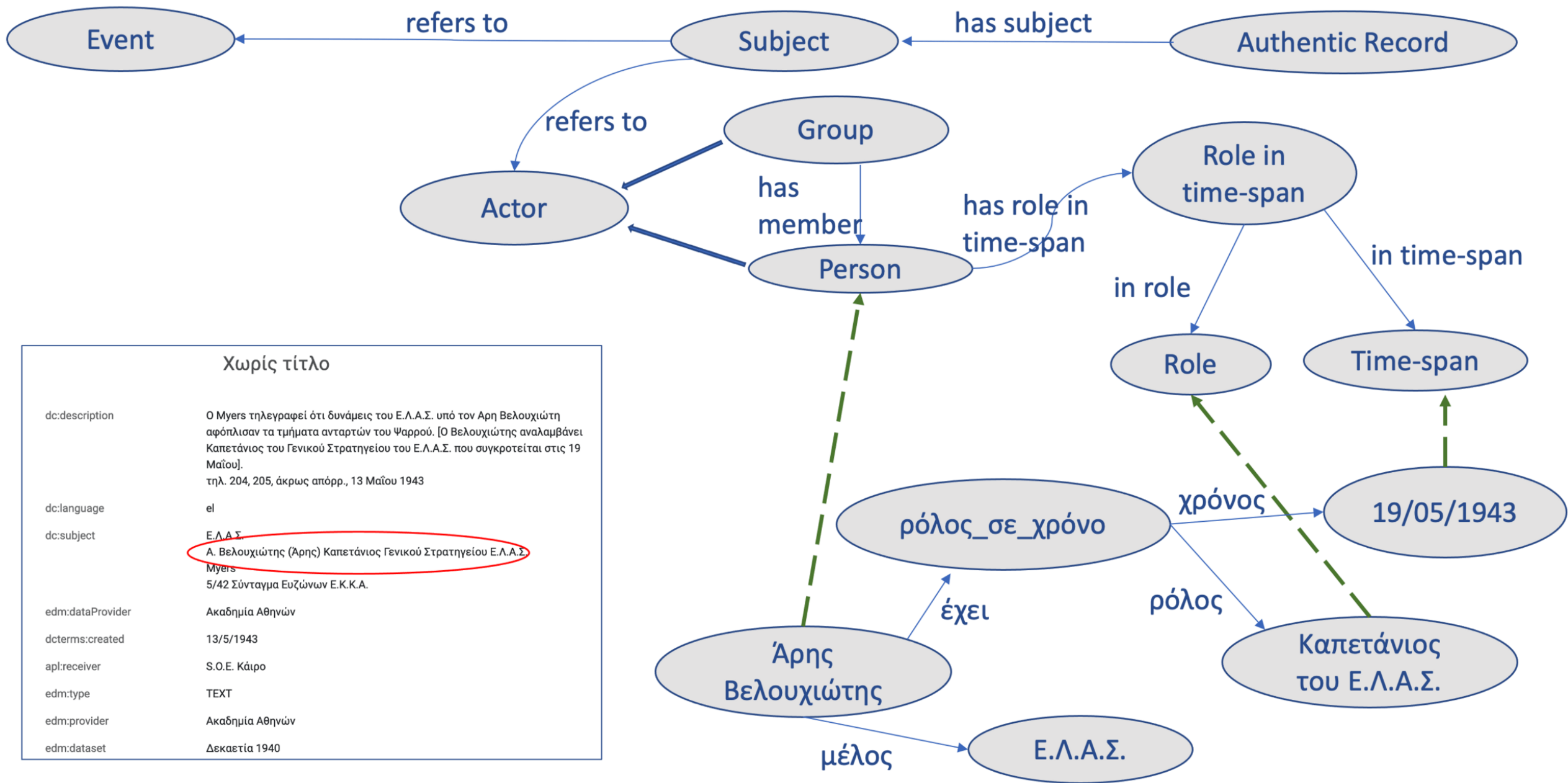
Οφέλη Χρήσης Σχήματος Βάσης Γνώσης

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

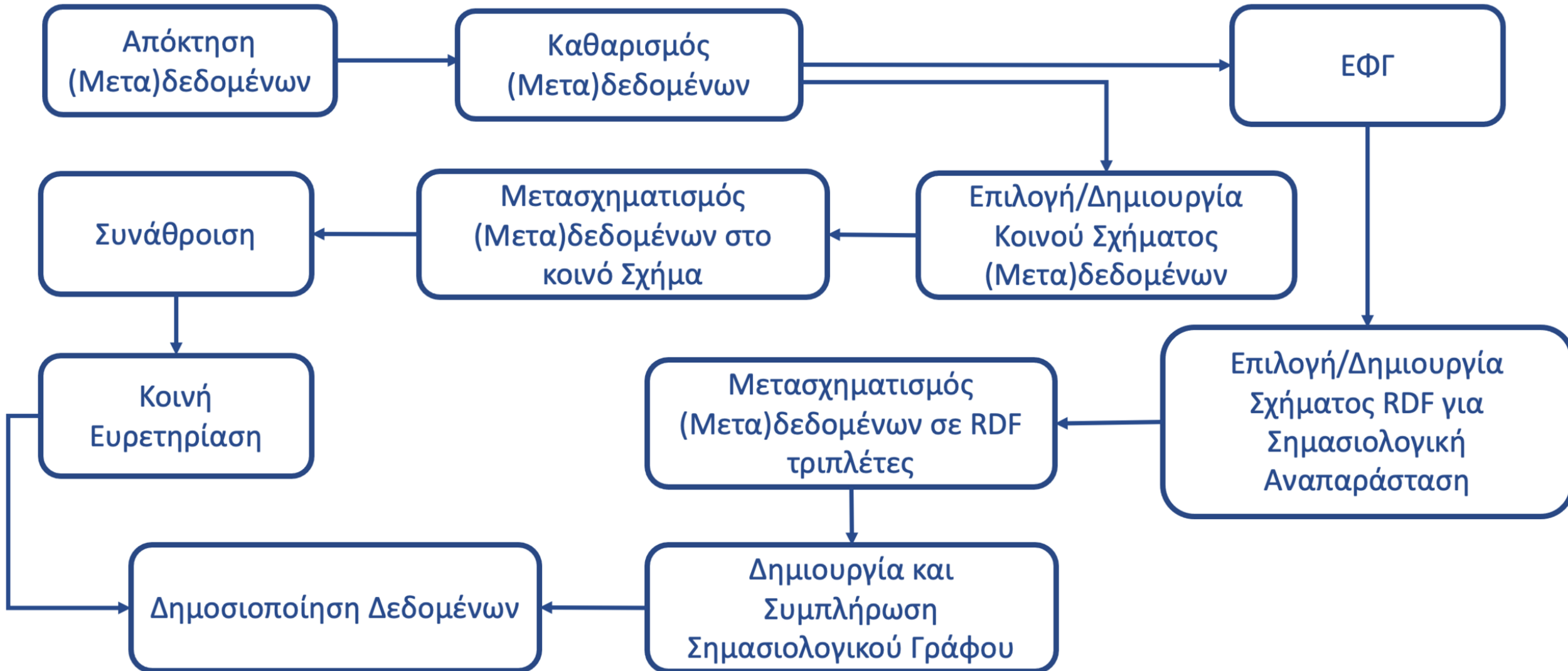
ΤΗΛΕΓΡΑΦΗΜΑ ΤΟΥ ΑΡΗ ΒΕΛΟΥΧΙΩΤΗ ΤΗΣ
18/04/1944 ΠΡΟΣ ΤΗΝ ΠΟΛΙΤΙΚΗ ΕΠΙΤΡΟΠΗ
ΕΘΝΙΚΗΣ ΑΠΕΛΕΥΘΕΡΩΣΗΣ ΚΑΙ ΤΟ ΓΕΝΙΚΟ
ΣΤΡΑΤΗΓΕΙΟ, ΣΧΕΤΙΚΑ ΜΕ ΤΗ ΔΙΑΛΥΣΗ ΤΟΥ 5/42
ΣΥΝΤΑΓΜΑΤΟΣ ΕΥΖΩΝΩΝ ΚΑΙ ΤΟ ΘΑΝΑΤΟ ΤΟΥ
ΣΥΝΤΑΓΜΑΤΑΡΧΗ ΨΑΡΡΟΥ ΚΑΙ ΤΟΥ ΤΑΓΜΑΤΑΡΧΗ
ΚΑΠΕΤΣΩΝΗ (ΑΡΧΕΙΟ ΔΟΥΚΑ)







Χάρτης Εργασιών Ψηφιακής Επιμέλειας



Παραγωγή RDF Δεδομένων

- Για τη διατύπωση της πληροφορίας χρησιμοποιήθηκε το W3C πρότυπο **RDF** (Resource Description Framework)
 - 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

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

Βάση γνώσης Δεκαετίας '40 (blazegraph)

Συμπεράσματα & Παρατηρήσεις

- Καταγραφή και αποτύπωση των εργασιών που εκτελέστηκαν σε διαγράμματα ροής
 - Διαμοιρασμός τους με την ερευνητική κοινότητα και ενδιαφερόμενους φορείς
 - Επαναχρησιμοποιήσιμες
 - Εξυπηρετούν ως οδηγοί για την εκτέλεση εργασιών επιμέλειας από ενδιαφερόμενους φορείς
- Η ακολουθία των βημάτων δεν είναι απόλυτα αυστηρή
 - Εξαρτάται από τον επιμελητή και το ίδιο το περιεχόμενο
 - Πλην κομβικών βημάτων, π.χ. απόκτηση, δημιουργία τελικής συλλογής

Συμπεράσματα & Παρατηρήσεις II

- Ενδιάμεσα αποτελέσματα μπορεί να απαιτούν την επανεκτέλεση προηγούμενων βημάτων
 - Διάσπαση τιμών και καθαρισμός
 - Διάσπαση τιμών και μοντελοποίηση
 - NLP και μοντελοποίηση
- Τελικά βήματα πριν τη δημοσιοποίηση απαιτούν προγραμματιστικές ικανότητες
 - Εισαγωγή σε σχεσιακή βάση ή συσσωρευτή xml, ευρετηρίαση, ανάπτυξη διεπαφής ή
 - Παραγωγή σημασιολογικού γράφου rdf, εκτέλεση πολύπλοκων ερωτημάτων (π.χ. SPARQL)