



Αρχεία: Και μετά την ψηφιοποίηση, τι;

Τεχνολογίες Ανάλυσης περιεχομένου και Europeana

ΣΤΕΦΑΝΟΣ ΚΟΛΛΙΑΣ

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ

30 Σεπτεμβρίου 2010

Περιεχόμενα

- Διαλειτουργικότητα και Ευρωπαϊκό Περιεχόμενο (Πολιτιστικό, Εκπαιδευτικό)
- Europeana - Θεματικοί Συσσωρευτές περιεχομένου (Εργαλείο εισαγωγής & αντιστοίχισης δεδομένων)
- Νέες τεχνολογίες - Συνδεδεμένα Δεδομένα
- Το Ελληνικό πεδίο - Εθνική Πολιτική

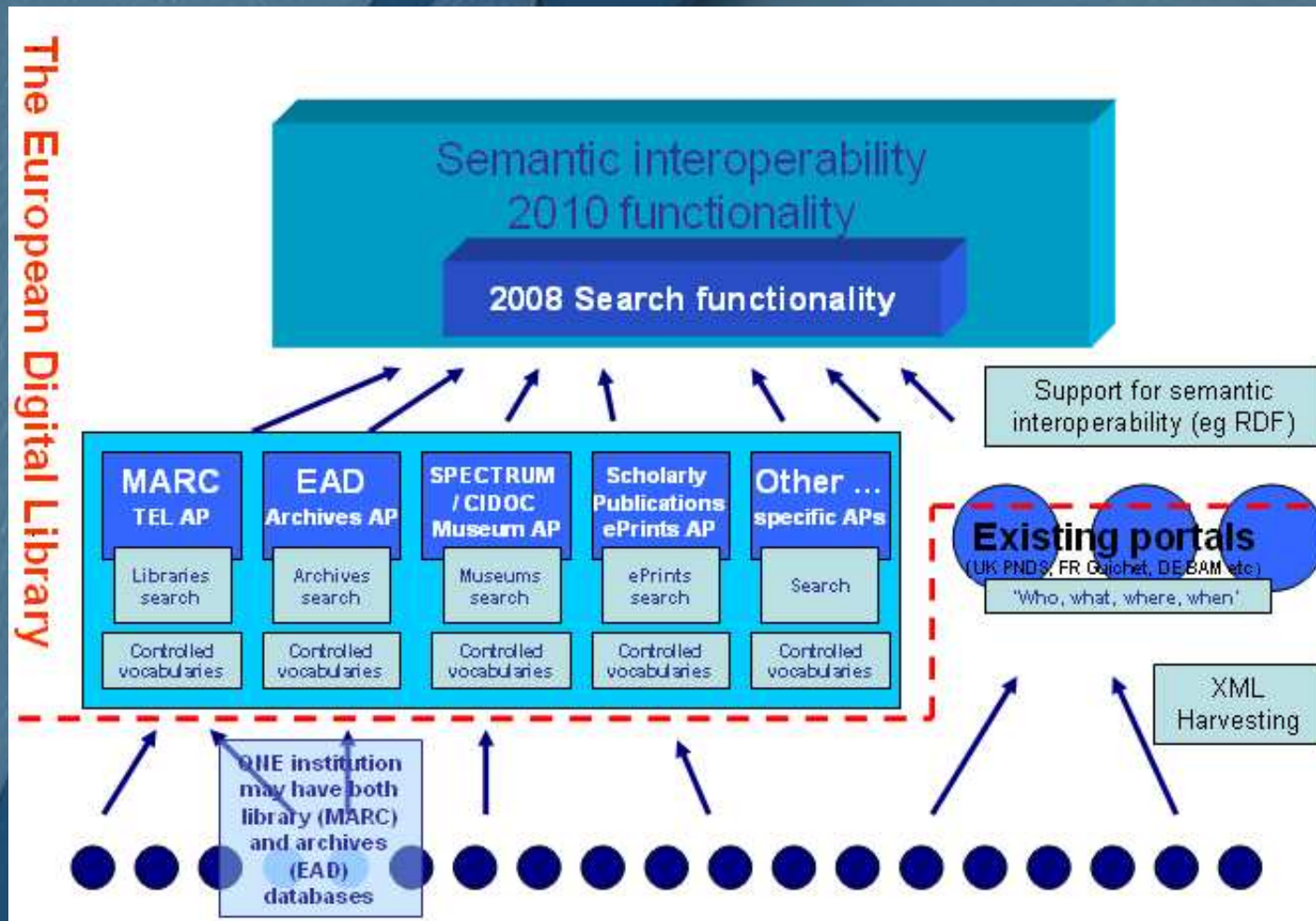
Εισαγωγή

- Η οργάνωση του περιεχομένου με στόχο την αποτελεσματική πρόσβαση έχει μεγάλη σπουδαιότητα σε εθνικό και ευρωπαϊκό - διεθνές επίπεδο
- Το 2007 η Ευρωπαϊκή Επιτροπή όρισε για τον πολιτιστικό τομέα την Ομάδα για την Ευρωπαϊκή Διαλειτουργικότητα με σκοπό την δημιουργία ενός roadmap για την προώθηση της διαλειτουργικότητας στο πλαίσιο των εθνικών και ευρωπαϊκών εξελίξεων.
- Στην ομάδα αυτή προτείναμε ένα πλαίσιο για την εξέλιξη τόσο για το άμεσο όσο και για το απώτερο μέλλον όσον αφορά στη διαλειτουργικότητα.
- Σήμερα το πλαίσιο διαλειτουργικότητας εξελίσσεται στην Europeana και ειδικότερα στα έργα Europeana v1.0/v2.0 and Europeana Connect.

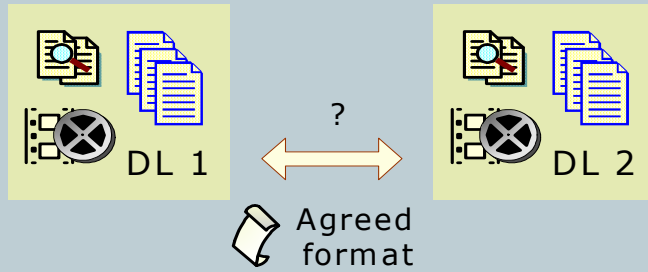
Europeana και Συσώρευση Περιεχομένου

- Η συσώρευση σε εθνικό και θεματικό επίπεδο , η πρόσβαση και η διατήρηση του περιεχομένου αποτελούν ένα από τους κύριους στόχους του Ευρωπαϊκού πολιτιστικού χώρου σήμερα.
- Στο ορόσημο Rhine 2010 περί τα 12.000.000 τεκμήρια θα είναι προσβάσιμα από τον κόμβο της EUROPEANA.
- Η συσώρευση περιεχομένου & η διαλειτουργικότητα με κατάλληλα μοντέλα μεταδεδομένων (ESE, EDM) είναι το μέσο για την επίτευξη του στόχου.
- Την περίοδο 2011-2013 η σημασιολογική πρόσβαση και χρήση των συνδεδεμένων δεδομένων θα αποτελέσουν τα κύρια σημεία των εξελίξεων.

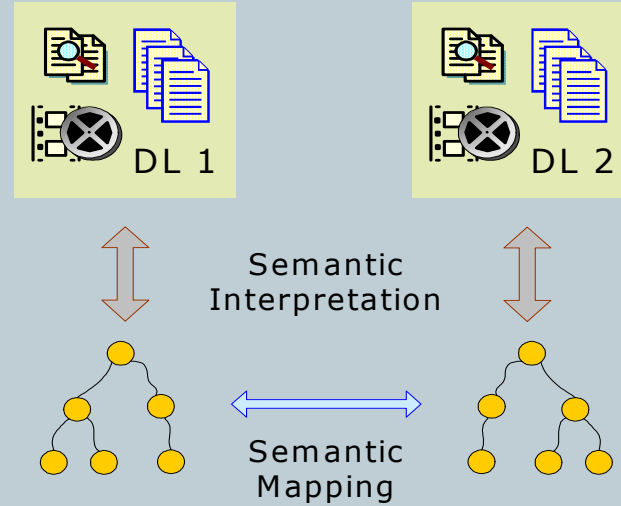
Το όραμα διαλειτουργικότητας (2007)



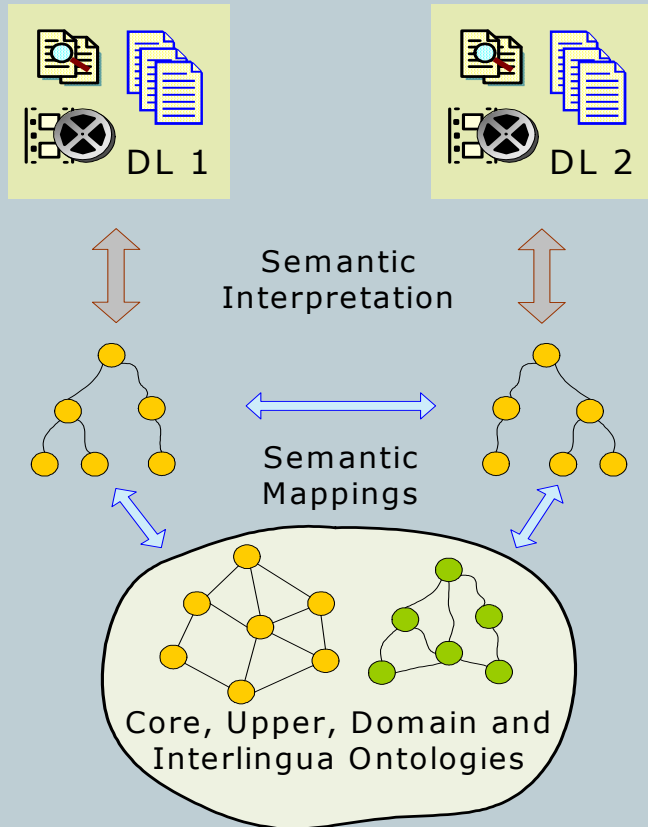
Scenario 1



Scenario 2



Scenario 4



Scenario 3

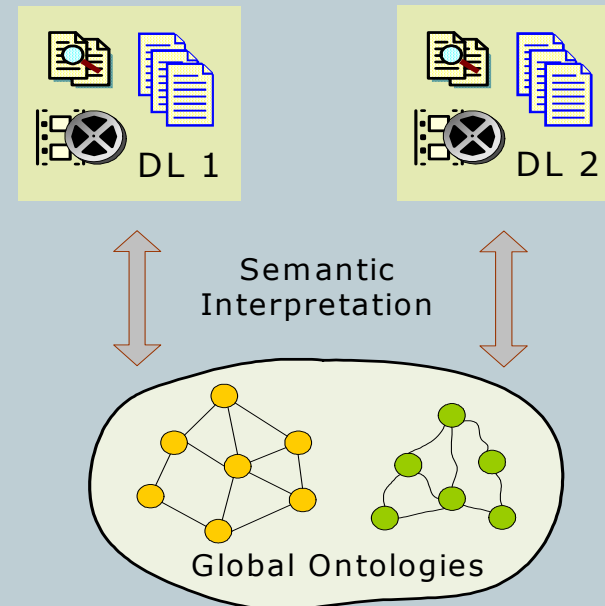


Figure 1. Achieving Semantic Interoperability

Αρχή 1: Σημασιολογική Τεκμηρίωση

- Περιγραφή αντικειμένων με έννοιες από κοινά λεξιλόγια



Original Abstract
Interview of the painter Salvador Dali.
Owner

Date created 
1962-01-19
Languages
French

...

...

...

Αρχή 2: Σημασιολογική Διασύνδεση

- Αναζήτηση αντικειμένων που συνδέονται εννοιολογικά
- Χρήση θησαυρών/λεξιλογίων
 - TGN for Geographic Names
 - ULAN for famous artists
 - IPTC for News, Sports etc.
 - Other thesauri and/or vocabulary

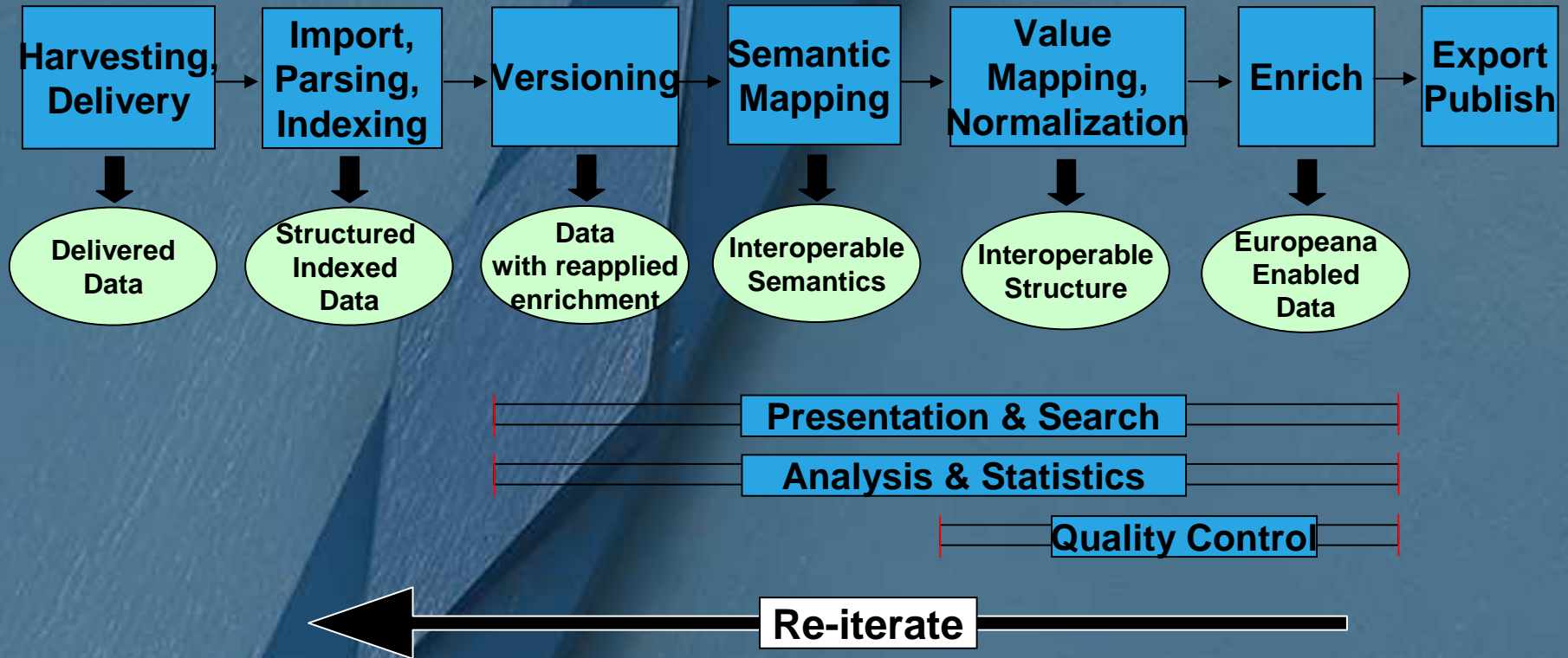
Σημασιολογική Οργάνωση Περιεχομένου

- Η σημασιολογική συσώρευση περιεχομένου και ο εμπλουτισμός του αποτελούν σημαντικές τεχνικές έννοιες. Το περιεχόμενο δεν ανήκει σε ανεξάρτητες πηγές, αλλά σε συνδεδεμένες γνωσιακές πηγές από όπου μπορεί να ανακτηθεί στο σύνολό του.
- Οι χρήστες μπορούν να μοιραστούν & να επαναχρησιμοποιήσουν το περιεχόμενο στο εννοιολογικό πλαίσιο που επιθυμούν και σε διαφορετικά σενάρια.
- Το περιεχόμενο που διαθέτει κάποιος συσσωρευτής περιέχει μεταδεδομένα, αλλά και θέματα διανοητικών δικαιωμάτων & πολυμεσική πληροφορία που του δίνουν περισσότερες δυνατότητες από την απλή ανταλλαγή - προσφορά πληροφοριών.

Διαλειτουργικότητα και Εργαλεία Συσώρευσης - Προσφοράς Περιεχομένου στη Europeana

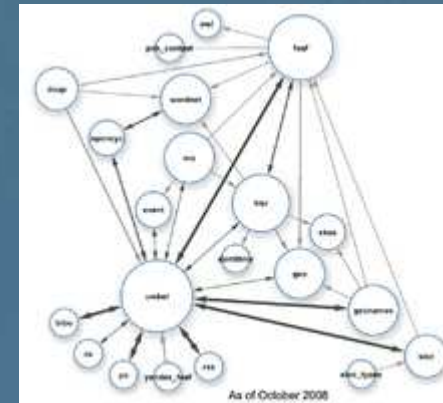
- Το ΕΜΠ έχει αναπτύξει ένα εργαλείο - Ingestion Tool (στο πλαίσιο του έργου ΑΘΗΝΑ) με το οποίο, περιεχόμενο κύρια από τον χώρο των μουσείων αντιστοιχίζεται με το πρότυπο LIDO (& το πρότυπο ESE)
- Αυτό χρησιμοποιείται εκτός από το ΑΘΗΝΑ και από άλλα έργα που συσσωρεύουν περιεχόμενο για τη Europeana: CARRARE (αρχαιολογία), EUSCREEN (οπτικοακουστικό), Judaica (πολιτιστικό), ECLAP (θέατρο), DSA (τέχνη), LINKED HERITAGE (πολιτιστικό).
- Επεκτάσεις αφορούν: στη σύνδεσή του με το πρότυπο EDM της EUROPEANA, στο σημασιολογικό εμπλουτισμό και στα συνδεδεμένα δεδομένα (LINKED HERITAGE).

Architecture & Phases



Συνδέοντας Ανοικτά Δεδομένα

- Ο σκοπός του W3C [Semantic Web Education and Outreach](#) group's [Linking Open Data](#) community project είναι η επέκταση του Παγκόσμιου Ιστού με την πρόσβαση σε ανοικτά δεδομένα από διαφορετικές πηγές. Το 2007 υπήρχαν πάνω από 2 δις RDF triples, συνδέοντας πάνω από 2 εκατομμύρια συνδέσμους RDF. Το 2009 ο αριθμός αυξήθηκε σε 4.2 δις RDF triples και 142 εκατομμύρια συνδέσμους.



Συνδεδεμένα δεδομένα και η Πολιτιστική Κληρονομιά

1. Τα δεδομένα περιλαμβάνουν όλα τα πολυμεσικά είδη, όπως και περιεχόμενο που δημιουργείται από τους χρήστες.
2. Τα συνδεδεμένα δεδομένα συνδέουν τον πολιτιστικό τομέα με όλο τον δημόσιο τομέα. Ως αποτέλεσμα οι επενδύσεις στους συναφείς τομείς και στη βιομηχανία μπορούν να επεκταθούν και στον πολιτιστικό τομέα.
3. Τα συνδεδεμένα δεδομένα μπορούν να προσθέσουν αξία στο υπάρχον περιεχόμενο.

Συνδεδεμένα δεδομένα και η Πολιτιστική κληρονομιά (2)

4. Τα συνδεδεμένα δεδομένα αποτελούν τη νέα φάση προσαρμοστικότητας στον Παγκόσμιο Ιστό.
5. Ο Πολιτιστικός τομέας πρέπει να εγγυάται Ανοικτά Δεδομένα, δηλαδή προσβάσιμα δεδομένα μέσω των ανοικτών και γνωστών προτύπων.

Πολιτιστική Κληρονομιά και e-Infrastructures

Τα έργα DC-NET, INDICATE :

- Διερευνούν και προετοιμάζουν την ενσωμάτωση των εφαρμογών της πολιτιστικής κληρονομιάς σε πλατφόρμες e-infrastructures (ΥΠΠΟΤ, ΥΠΕΠΘ, ΕΜΠ)
- Σχεδιάζουν ένα πρωτότυπο e-culture science gateway (eCSG)

Το Ελληνικό Πεδίο - Εθνική Πολιτική

- Υπάρχον Ψηφιοποιημένο Περιεχόμενο (κύρια Γ ΚΠΣ - επεκτάσεις στο Δ ΚΠΣ)
- Διαλειτουργική του παρουσίαση σε εθνικό και Ευρωπαϊκό επίπεδο - Europeana (στην τωρινή φάση η Ελλάδα συμβάλλει στο 1,5% περίπου του περιεχομένου)
- Ανάγκη επέκτασης - διατήρησης - εμπλουτισμού - διαλειτουργικότητας με εξελίξεις (πχ με πρότυπο EDM)
- Ανάγκη εθνικού πλαισίου ανάπτυξης και προβολής του περιεχομένου (εθνικού συσσωρευτή)
- Ενέργειες και ρόλος του ΥΠΕΠΘ και του ΥΠΠΟΤ.



ΣΑΣ ΕΥΧΑΡΙΣΤΩ