

Οι ακαδημαϊκές βιβλιοθήκες ως πάροχοι δεδομένων στο σημασιολογικό ιστό

Η μετάβαση από τη μία εγγραφή ανά πόρο στα δίκτυα σχέσεων μεταξύ
αυτόνομων οντοτήτων

Μανόλης Πεπονάκης

Εθνικό Κέντρο Τεκμηρίωσης, Εθνικό Ίδρυμα Ερευνών
epepo@ekt.gr

22^ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών
24-25 Οκτωβρίου 2013, Πάτρα

Οι αδιαίρετες περιγραφές των αδιαίρετων αντικειμένων

- Στη διάρκεια του 20^{ου} αιώνα οι βιβλιοθήκες περιγράφουν αδιαίρετα φυσικά αντικείμενα
 - κατά κανόνα βιβλία
- Για να περιγράψουν τα “αδιαίρετα αντικείμενα” δημιουργούν “αδιαίρετες περιγραφές” (εγγραφές)
 - Όσο εξελίσσεται η καταλογογράφηση οι “αδιαίρετες περιγραφές” αρχίζουν να αποτελούνται από διακριτά τμήματα (εγγραφές καθιερωμένων αποδόσεων, στοιχεία από ταξινομικά συστήματα, κτλ)
 - Όμως τα παραπάνω ενσωματώνονται στην εγγραφή, άρα ο κανόνας είναι: μία εγγραφή ανά πόρο



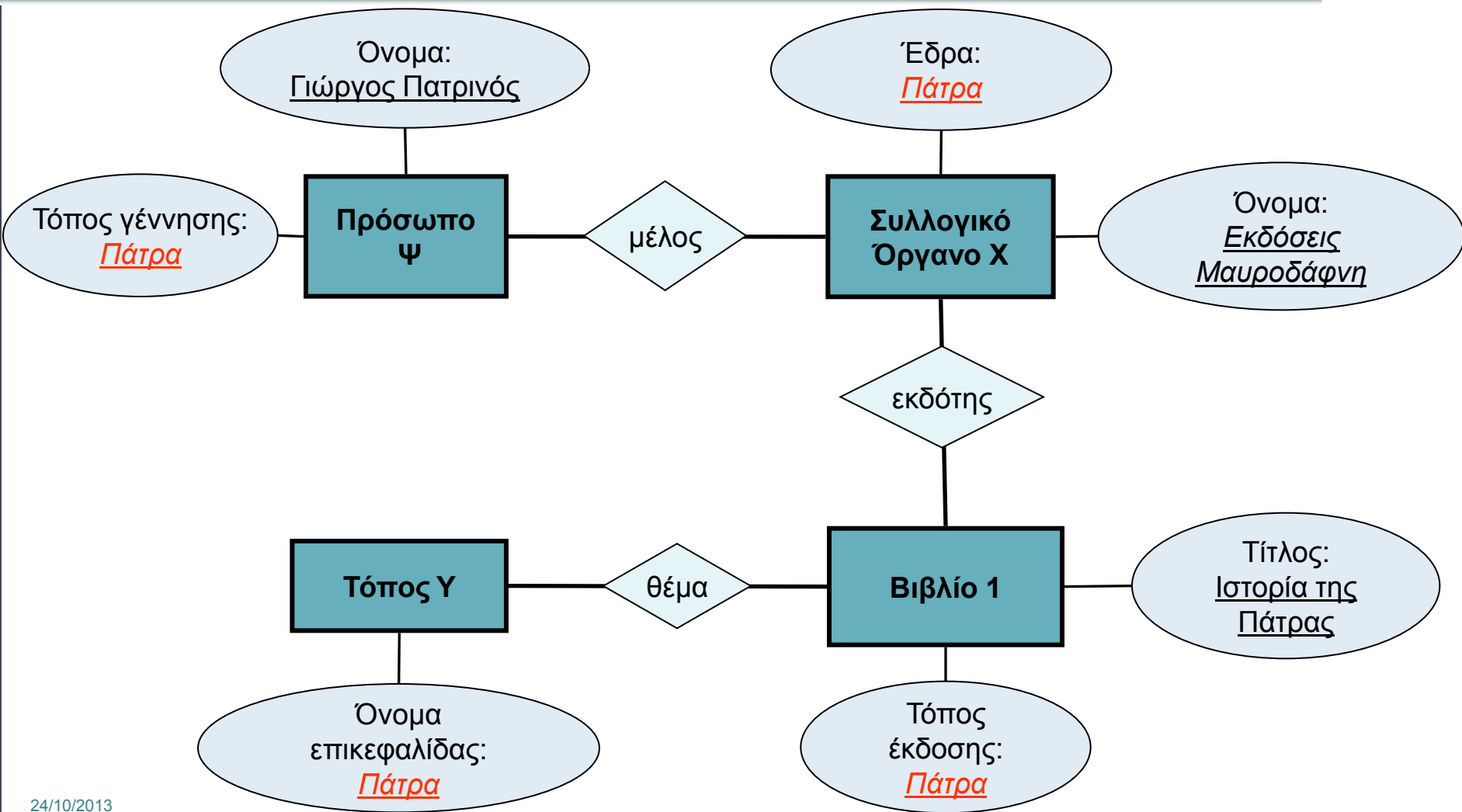
Η καινοτόμα θεώρηση των FRBR

- Το κέντρο της περιγραφής δεν είναι ο πόρος (είτε ως φυσικό αντικείμενο, είτε ως ψηφιακό)
 - η εγγραφή περιέχει σημαντικά στοιχεία για την ερμηνεία του πόρου τα οποία δεν προέρχονται (κατ' ανάγκη) από τον ίδιο τον πόρο:
 - η ημερομηνία δημιουργίας του Έργου
 - ο τίτλος του Έργου
 - τα συμφραζόμενα (context) μέσα στα οποία γράφτηκε το Έργο
 - ο τόπος γέννησης του συγγραφέα
 - ...
- Η εγγραφή δεν είναι αδιαίρετη
 - η εγγραφή δεν είναι μια αδιαίρετη περιγραφή ενός αδιαίρετου πόρου αλλά ένα δίκτυο συνδέσεων μεταξύ “αυτόνομων” οντοτήτων



Παράδειγμα Αναπαράστασης FRBR / FRAD

Entity Relationship based

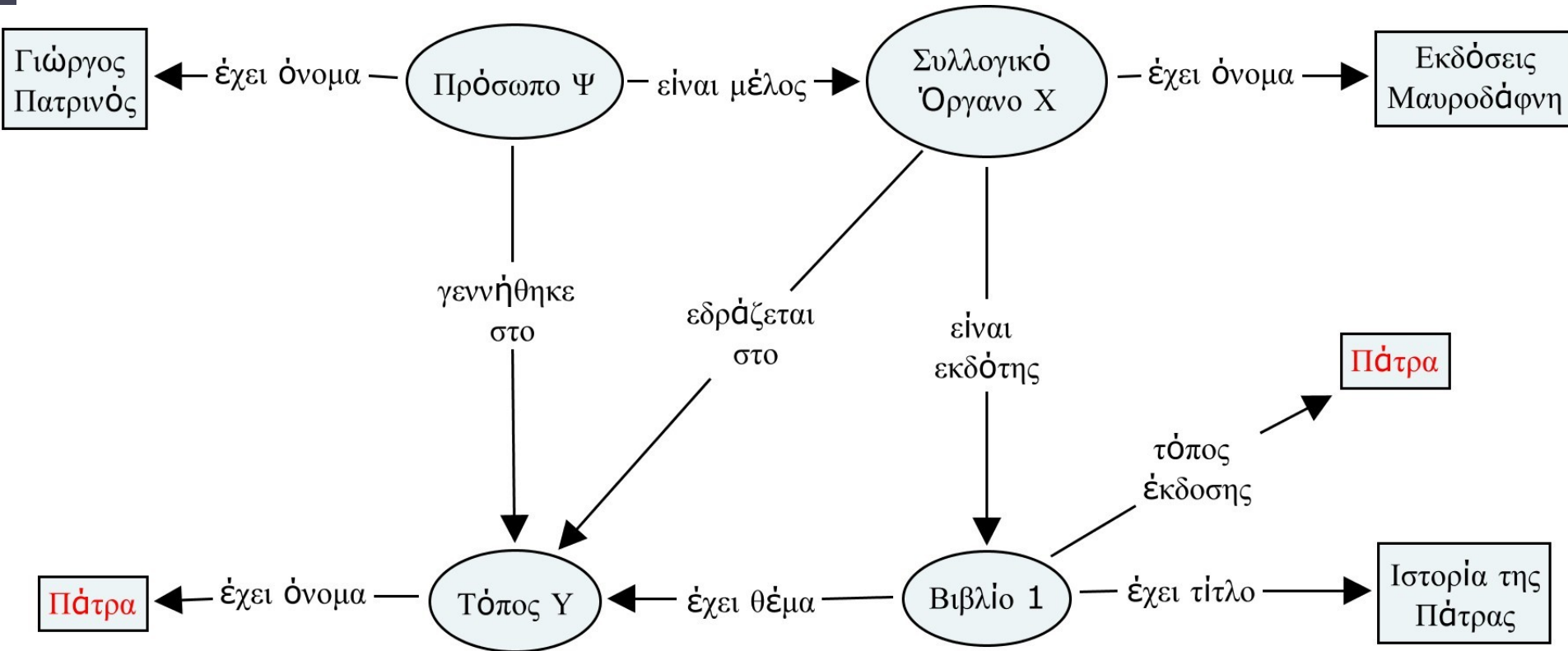


Η «οπισθοδρόμηση» των FRBR

- Η επιλογή του Entity Relationship (ER) μοντέλου
 - Η αναπαράσταση των FRBR σε ER δημιουργεί προβλήματα στην έκφραση των διασυνδέσεων, ειδικά στο σύγχρονο τεχνολογικό περιβάλλον
 - παράδειγμα, ο τόπος υπάρχει ως οντότητα αλλά και ως ιδιότητα σε εννιά περιπτώσεις:
 - τόπος γέννησης του Προσώπου
 - τόπος διαμονής του Προσώπου
 - τόπος προέλευσης του Έργου
 - ...
- ❖ Νέες λογικές εκφρασμένες σε “ξεπερασμένη” γλώσσα



Τα semantics του προηγούμενου παραδείγματος εκφρασμένα στη λογική της RDF



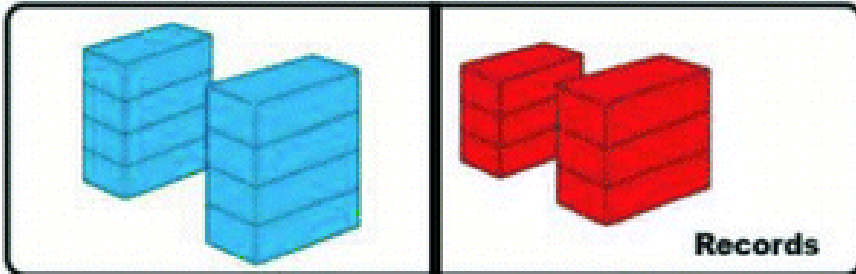
Οι 4 αναφορές της Πάτρας, που υπήρχαν στο ER, έγιναν 2
Δεν έγιναν μία διότι σε αυτή την περίπτωση θα καταργούσαμε το περιγραφικό τμήμα



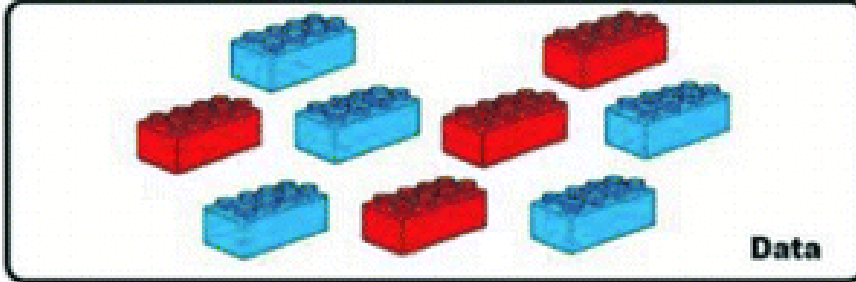
Από τις εγγραφές στα δεδομένα

- Πρώτος στόχος: να κατακερματίσουμε τις εγγραφές στα «συστατικά τους τμήματα»
 - π.χ. στις οντότητες των FRBR
 - Δεύτερος στόχος: να βρούμε τις συνδέσεις μεταξύ αυτών των «αυτόνομων» τμημάτων
 - εντός και εκτός του χώρου των βιβλιοθηκών
- ❖ *ο πόρος πρέπει να αποδομηθεί στις βασικές του οντότητες και η σύνθεση αυτών των οντοτήτων θα επιτρέψει μια καρποφόρα αναπαράσταση του πόρου και των συνδέσεών του με άλλες οντότητες*

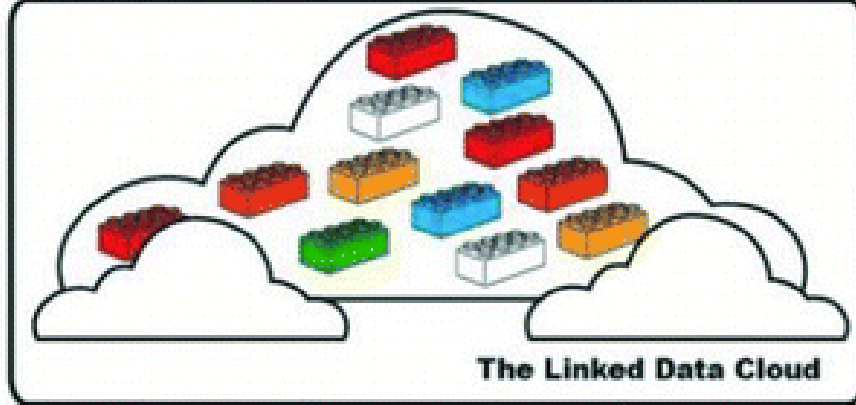




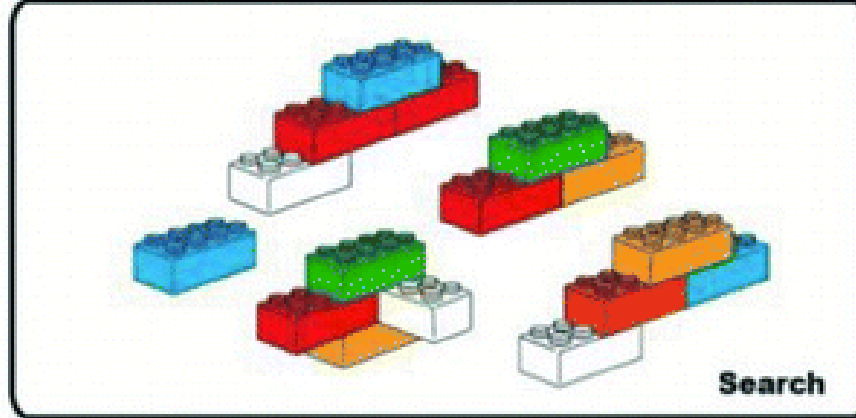
Transforming records into data



Publishing data



Linking data as you search or browse



Σχήμα από: Lampert, C. K., & Southwick, S. B. (2013). Leading to Linking: Introducing Linked Data to Academic Library Digital Collections. *Journal of Library Metadata*, 13(2-3), 230–253.

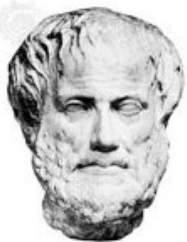
Linked data και τα οφέλη των URIs

- Η κάθε οντότητα, το κάθε «τουβλάκι» της προηγούμενης διαφάνειας, χαρακτηρίζεται μοναδικά και «παγκόσμια» από ένα http URI
 - Δυνητικά μπορούμε να διασυνδέσουμε κάθε οντότητα με μια «δική μας»
 - Κατά την αναζήτηση αξιοποιούνται οι σχέσεις μεταξύ των οντοτήτων
- Δεν χρειάζεται, πλέον, να κατεβάζουμε εγγραφές
 - Αρκεί μια δήλωση του τύπου «αυτό εκεί»
 - Με την προϋπόθεση πως πίσω από το «εκεί» κρύβεται μια ακριβής και επαρκής περιγραφή
 - π.χ. αντί για Ελύτης, <http://viaf.org/viaf/24601064/>



Ποιος δημιουργεί;

- Αφού το URI και μόνο είναι αρκετό για να παραπέμψουμε, τότε ποιος είναι εκείνος που θα δώσει την «πρώτη/βασική» περιγραφή που όλοι οι άλλοι θα μπορούν να χρησιμοποιούν;
- Πόσο αξιόπιστη είναι μια περιγραφή ώστε να «χτίσουμε» πάνω της;
 - Μήπως ο Αριστοτέλης ζει;
 - Πόση αλήθεια κρύβει η πληροφορία του Google Scholar;
 - Σύμφωνα με το Google Scholar ο Αριστοτέλης έχει “Verified email at buffalo.edu”!
 - <http://scholar.google.com/citations?user=4rO87mQAAAAJ>



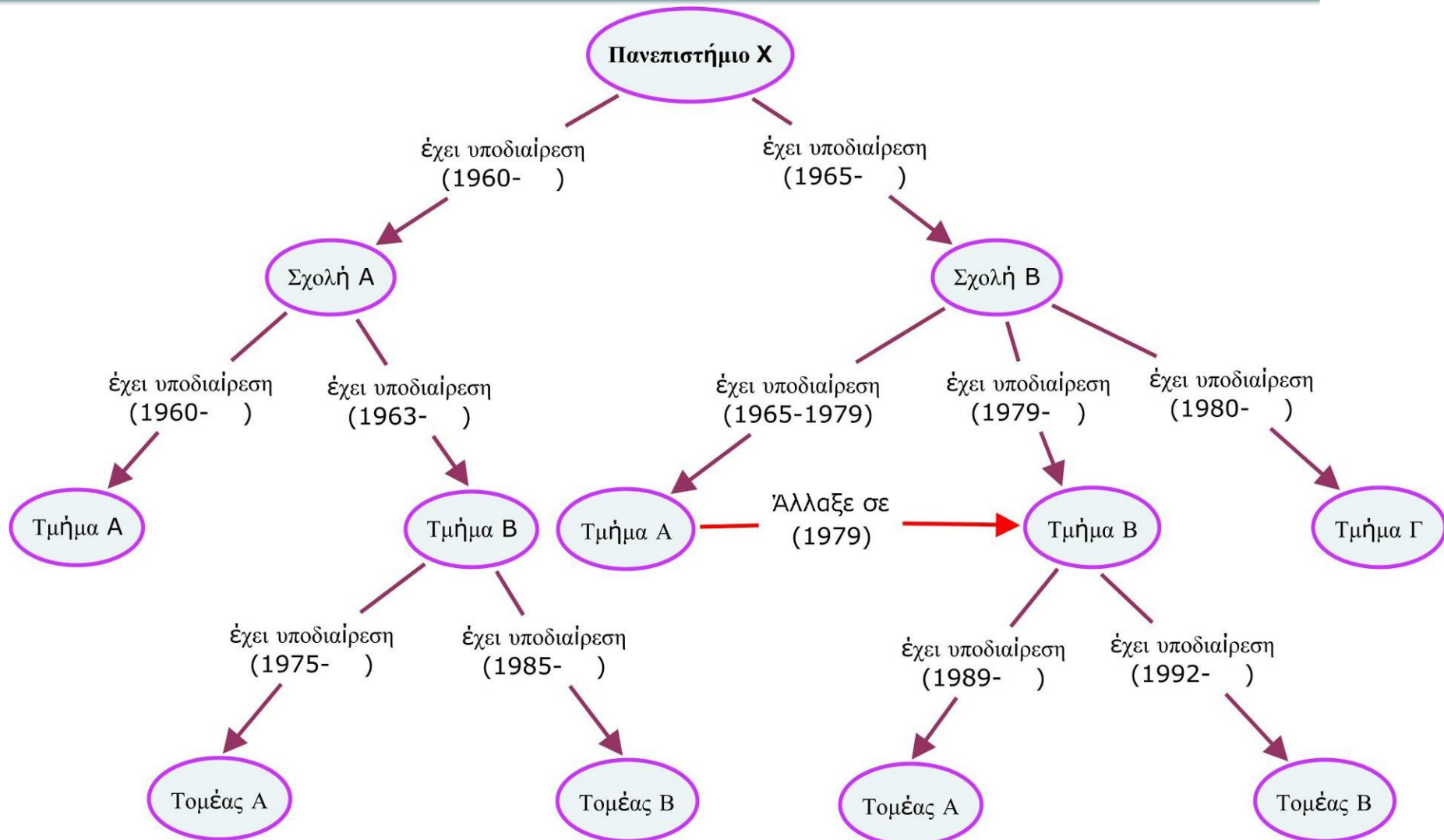
Στην ακαδημαϊκή κοινότητα

- Οι ακαδημαϊκές βιβλιοθήκες είναι υπεύθυνες για την καταγραφή μεγάλου μέρους της πληροφορίας που παράγει το οικείο τους ίδρυμα:
 - της ακαδημαϊκής δομής του Πανεπιστημίου
 - των μελών ΔΕΠ
 - των μεταπτυχιακών εργασιών
 - ...
- Διότι η βιβλιοθήκη:
 - έχει την τεχνογνωσία της μοντελοποίησης και κωδικοποίησης της πληροφορίας
 - έχει πρόσβαση στην πρωτογενή πληροφορία
- Έτσι εξασφαλίζεται:
 - η ακεραιότητα (integrity) της πληροφορίας και
 - μέσω της πιστοποιημένης προέλευσής (provenance) της
 - αυξάνεται η εμπιστοσύνη (trust) των τρίτων



Ένα παράδειγμα:

Η δομή του πανεπιστημίου ως τμήμα των Authorities



Επαναχρησιμοποίηση

- Την προηγούμενη δομή μπορεί να χρησιμοποιήσει όποιος θέλει:
 - να μοντελοποιήσει τους φοιτητές ανά τμήμα
 - να δει την κατανομή του γένους των καθηγητών ανά σχολή
 - να κατανείμει τις μεταπτυχιακές εργασίες ανά τομέα
 - ...

Αυτό σημαίνει επαναχρησιμοποίηση (reuse) των δεδομένων στο σημασιολογικό ιστό
- ❖ Αν δεν υπάρχει λεπτομερής και αξιόπιστη δομή καθένας αναγκάζεται να φτιάχνει μια αντίστοιχη (πιθανότατα ελλιπή και μειωμένης αξιοπιστίας) για να τη χρησιμοποιήσει



Εν κατακλείδι

- Η καταλογογράφηση αλλάζει, οι κατάλογοι δεν αποτελούνται από αδιαίρετες εγγραφές που περιγράφουν αδιαίρετους πόρους
 - Οι κατάλογοι γίνονται δίκτυα σχέσεων μεταξύ αυτόνομων οντοτήτων
- Οι βιβλιοθήκες έχουν δομημένη πληροφορία η οποία πρέπει να ενσωματωθεί στα δεδομένα του σημασιολογικού ιστού
 - Διαλειτουργικότητα
- Τα δεδομένα των βιβλιοθηκών (πρέπει να) είναι αξιόπιστα ώστε «τρίτοι» να στηριχθούν σε αυτά και να τα αξιοποιήσουν-επεκτείνουν
 - Υπηρεσίες προστιθέμενης αξίας

