

Διεργασίες ψηφιακής επιμέλειας

Πάνος Κωνσταντόπουλος
Οικονομικό Πανεπιστήμιο Αθηνών και Ε.Κ. «Αθηνά»

19ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών,
Πάντειο Πανεπιστήμιο, Αθήνα, 5/11/2010

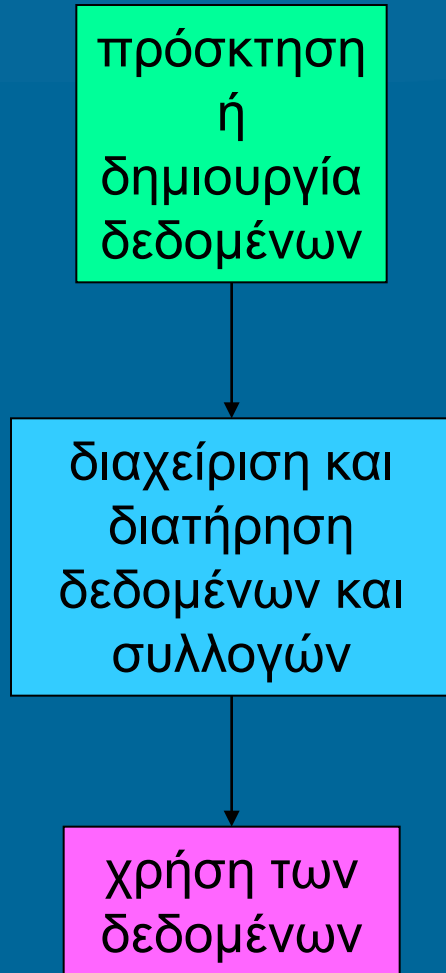


Δυνάμει ενιαίος ψηφιακός χώρος

- Γεννάται χάρη στη δυνατότητα ενιαίας πρόσβασης σε αυτοτελείς ψηφιακές συλλογές
- Πολλαπλασιαστικής αξίας
- Προϋπόθεση: συντακτική και σημασιολογική διαλειτουργικότητα
- Διατήρηση
 - Διαδικασίες, μεταδεδομένα
- Σύγκλιση(;) πληροφοριακών λειτουργιών αρχείων, βιβλιοθηκών και μουσείων



Κύκλος ζωής των ψηφιακών πόρων

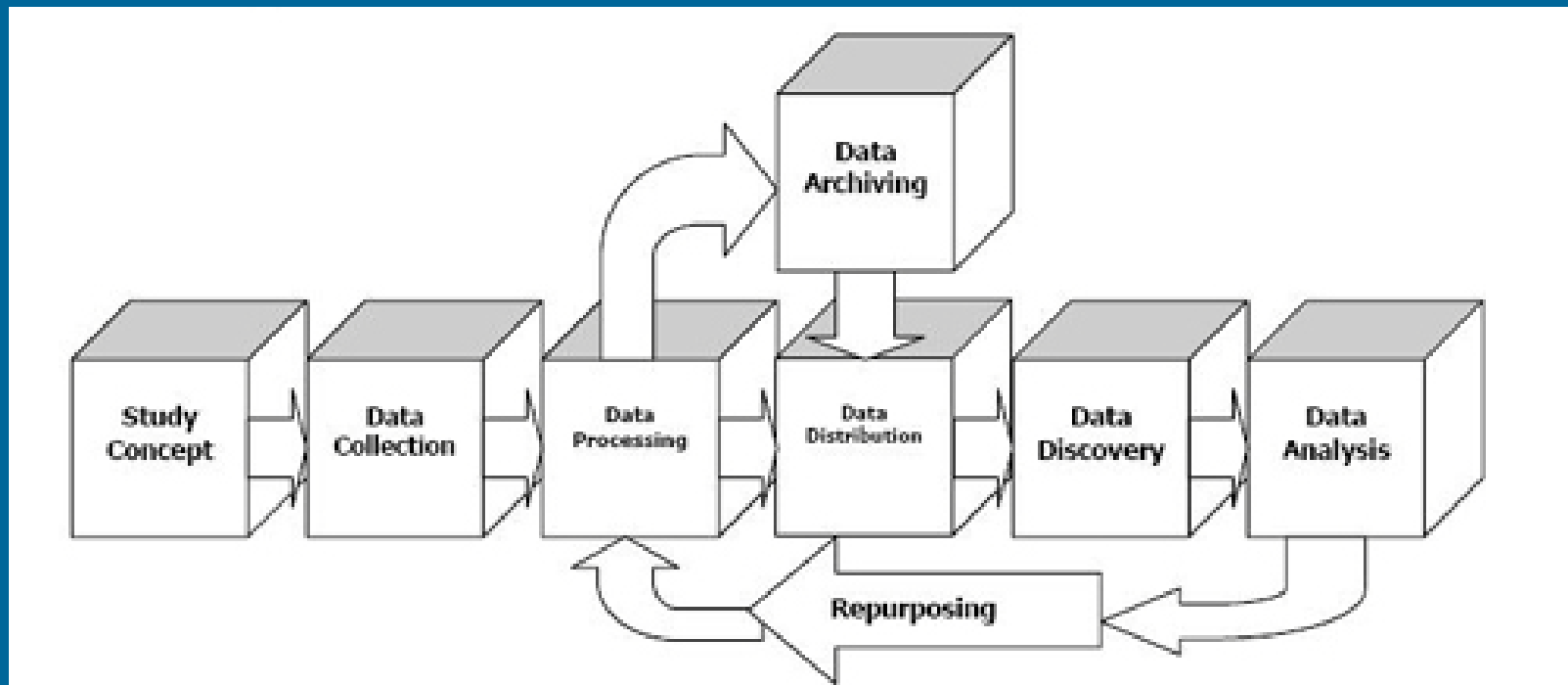


- Αρχικοποίηση
 - Αρχική μαζική καταχώριση ψηφιακού υλικού
- Συλλεκτική πολιτική
- Ποιότητα δεδομένων
 - Εγκυρότητα, πιστότητα, πληρότητα
- Διαλειτουργικότητα
 - Διασφάλιση της διασύνδεσης πληροφοριών από αυτοτελή συστήματα
- Βιωσιμότητα
 - Διασφάλιση της ακεραιότητας και αναγνωσιμότητας των πληροφοριών έναντι σφαλμάτων και τεχνολογικών αλλαγών
- Σύνθεση, ανανοηματοδότηση, έκδοση

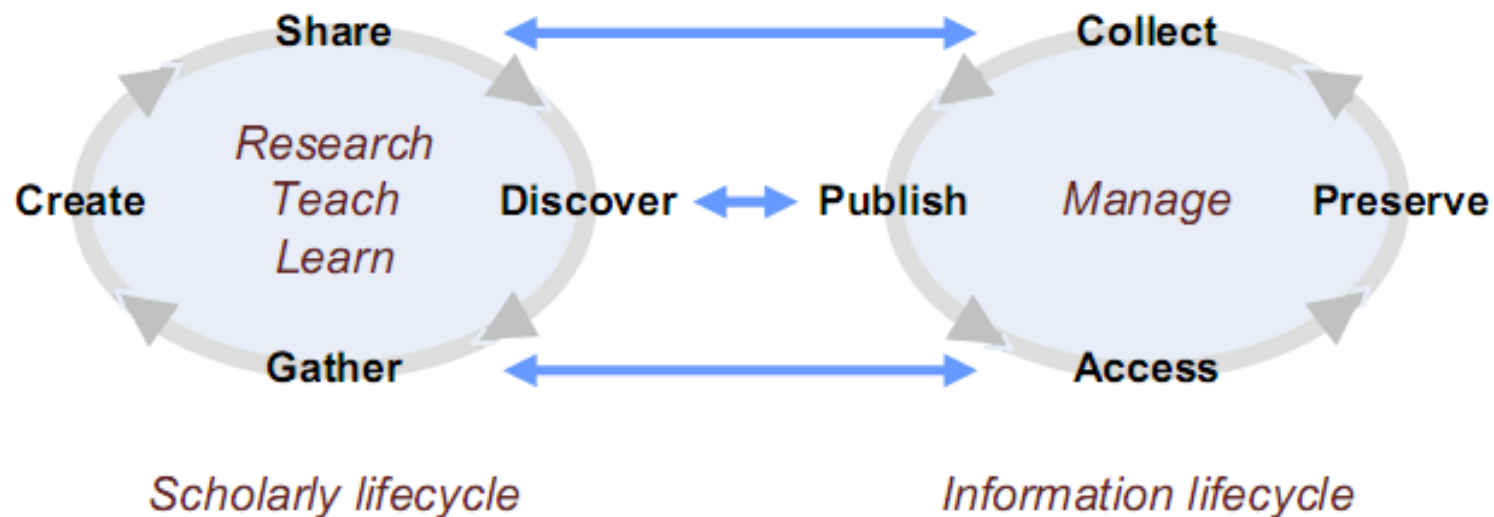


Κύκλος ζωής ερευνητικών δεδομένων

Πηγή: DDI Structural Reform Group. "DDI Version 3.0 Conceptual Model." DDI Alliance. 2004. Accessed on 11 August 2008. <<http://www.icpsr.umich.edu/DDI/committee-info/Concept-Model-WD.pdf>>.



Σύνδεση ερευνητικής και πληροφοριακής δραστηριότητας



Πηγή: UC Curation Center / California Digital Library, Merritt: An Emergent Micro-services Approach to Digital Curation Infrastructure, Rev. 0.6 – 2010-03-25



Ποιότητα των δεδομένων

- Εγκυρότητα
 - Προέλευση
 - Δεν υπάρχουν «ουδέτερα» δεδομένα
- Πιστότητα
 - Πιστή απόδοση του πρωτοτύπου: διαβαθμίσεις κατά χρήση
 - Ακριβής απόδοση της πραγματικότητας: μέσα στο ερμηνευτικό πλαίσιο
 - Αβέβαια και αντιφατικά δεδομένα
- Πληρότητα:
 - Δεν υπάρχει «πλήρης τεκμηρίωση»
- Πολυπλοκότητα:
 - Σύνθετα αντικείμενα, σχέσεις και συμφραζόμενα

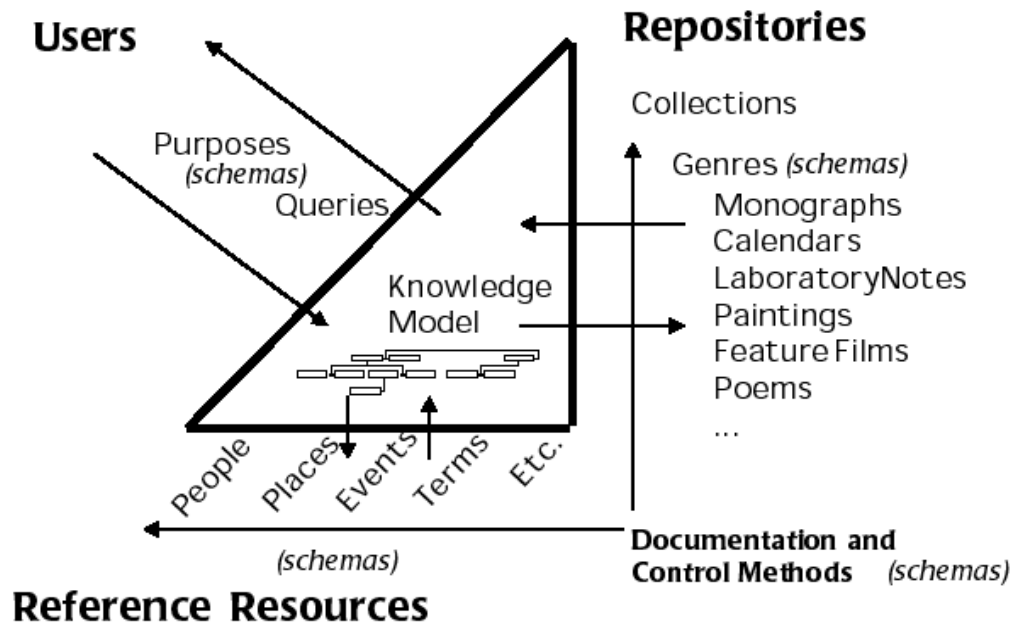


Πώς θα διασφαλιστεί η χρησιμότητα των ψηφιακών συλλογών;

- Συγκρότηση και διαχείριση σύνθετων ψηφιακών αντικειμένων αποτελούμενων από
 - Ψηφιακά υποκατάστατα
 - Συμβολικές παραστάσεις με βάση
 - Περιγραφικούς κώδικες και κανόνες
 - Θεματικές γλώσσες, οντολογίες πεδίου
- Η επιτυχία των ψηφιακών συλλογών συναρτάται με την πολλαπλότητα πρόσβασης, ερμηνείας και χρήσης
- Η προσέγγιση της «συγκέντρωσης» δεν είναι εύστοχη



Παράγοντες των πληροφοριακών διεργασιών



Πηγή: D. Bearman & J. Trant, <http://www.archimuse.com/papers/ukoln98paper/index.html>



Ψηφιακοί πόροι

Είδη πόρων

- **Δεδομένα**, πρωτογενή και δευτερογενή
- **Μοντέλα**
 - Οδηγοί ερμηνείας και συσχετισμού δεδομένων
- **Ορολογία**

Απαιτήσεις πόρων

Για όλους τους ψηφιακούς πόρους πρέπει να διασφαλισθεί:

- Η **ποιότητα**
- Η **διατήρηση**
- Η **χρησιμότητα** στο μέλλον και σε νέες συνθήκες



Τα μείζονα ζητούμενα

- Ψηφιακή αποταμίευση δεδομένων
- Πολλαπλή ερμηνεία – νοηματοδότηση, πλαισίωση
- Εξελισσόμενη ερμηνεία, αναπλαισίωση
- Υποστήριξη της γνωστικής διεργασίας: τοπική, υπερτοπική, συνδυαστική
- *«Η διατήρηση και επαύξηση της αξίας ενός σώματος δόκιμων πληροφοριών για χρήση τώρα και στο μέλλον»*



www.dcc.ac.uk/about/what/index.html

Ψηφιακή επιμέλεια



Ανάγκη για ψηφιακή επιμέλεια

- Κίνδυνος απώλειας ή αχρηστίας σημαντικών επενδύσεων σε ψηφιακό υλικό
- Ταχύτητα αυξανόμενοι όγκοι δεδομένων από όλα τα επιστημονικά πεδία, εκπαιδευτικό υλικό, αρχειακό υλικό και ποικίλα πολιτισμικά τεκμήρια
- Ασύμφορο ή και αδύνατο ένας φορέας να αναπτύξει μόνος τις απαραίτητες δράσεις
- Ανάγκη συνεργασίας σε ευρύτερα σχήματα με τη συμμετοχή εξειδικευμένων επιστημονικών μονάδων
- Παγκόσμια πραγματικότητα η υστέρηση φορέων ως προς την μακροπρόθεσμη διατήρηση και αξιοποίηση του ψηφιακού υλικού



Ψηφιακή επιμέλεια

Παραγωγή καλής ποιότητας, αξιόπιστων ψηφιακών πόρων

- Καλές πρακτικές ψηφιοποίησης και τεκμηρίωσης
 - Μελέτες – οδηγίες ΚΤΠ
- Ανίχνευση προέλευσης
- Αυθεντικότητα και νομική εγκυρότητα



Ψηφιακή επιμέλεια

Οργάνωση και αρχειοθέτηση

- Εύκολη πρόσβαση
- Αναχρησιμοποίηση
- Συνδυαστική χρήση
- Πρότυπα
- Συστήματα οργάνωσης γνώσης
 - Π.χ., θησαυροί όρων, θεματικοί χάρτες



Ψηφιακή επιμέλεια

Μακροπρόθεσμη διατήρηση

- των ιδίων των ψηφιακών πόρων
- του εννοιολογικού υποβάθρου αυτών
- Απαραίτητη όταν τα δεδομένα είναι πολύτιμα και δεν μπορούν να ανακτηθούν
- Οι ψηφιακοί πόροι απαιτείται να είναι διαθέσιμοι και ασφαλείς παρά τις όποιες μελλοντικές αστοχίες εξοπλισμού καθώς και τις αλλαγές της τεχνολογίας.



Διατήρηση

- Οι ψηφιακοί πόροι απαιτείται να είναι διαθέσιμοι και ασφαλείς παρά τις όποιες μελλοντικές αποτυχίες εξοπλισμού καθώς και τις αλλαγές της τεχνολογίας.

Παράγοντας κινδύνου	Χαρακτήρας
Γήρανση και αστοχία μέσω αποθήκευσης	Συστηματική, τυχαία
Παλαίωση μορφοτύπων	Συστηματική
Παλαίωση εξοπλισμού	Συστηματική
Ανθρώπινο σφάλμα	Τυχαίο, αναπόφευκτο
Σφάλμα λογισμικού	Δύσκολο να μοντελοποιηθεί
Εξωτερικό γεγονός	Τυχαίο



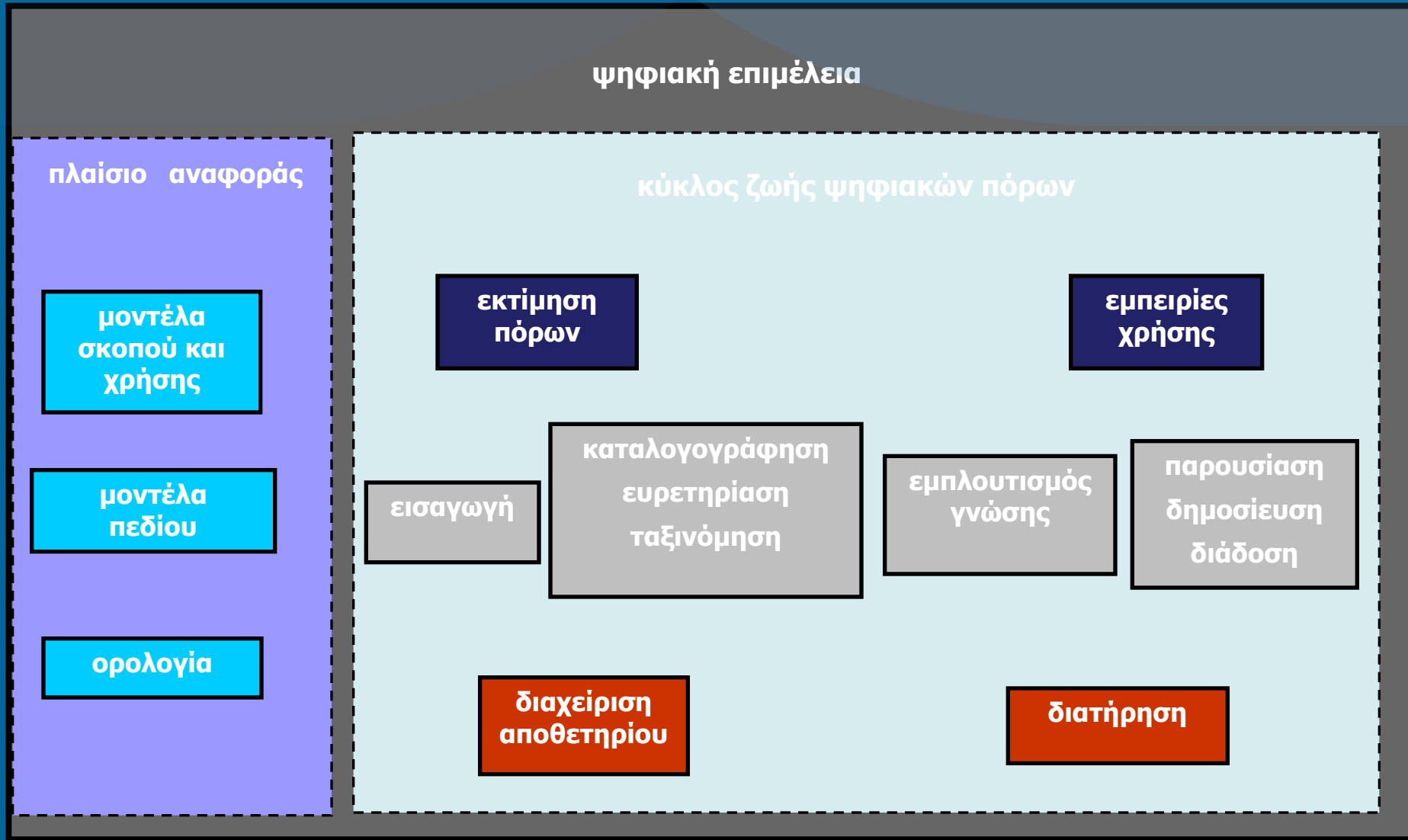
Ψηφιακή επιμέλεια

Παραγωγή νέας γνώσης από τους ψηφιακούς πόρους
Κοινοποίηση, αξιοποίηση και χρήση των πόρων

- Συνδυαστική χρήση πόρων
- Δυνατή χάρη στην κατανόηση και μοντελοποίηση του πλαισίου αναφοράς και χρήσης
 - Διεπιστημονική προσέγγιση
- Νέες χρήσεις για παλιότερο υλικό
 - Σημασιολογική ανανέωση, κοινότητες ερμηνευτών



Διεργασίες Ψηφιακής Επιμέλειας



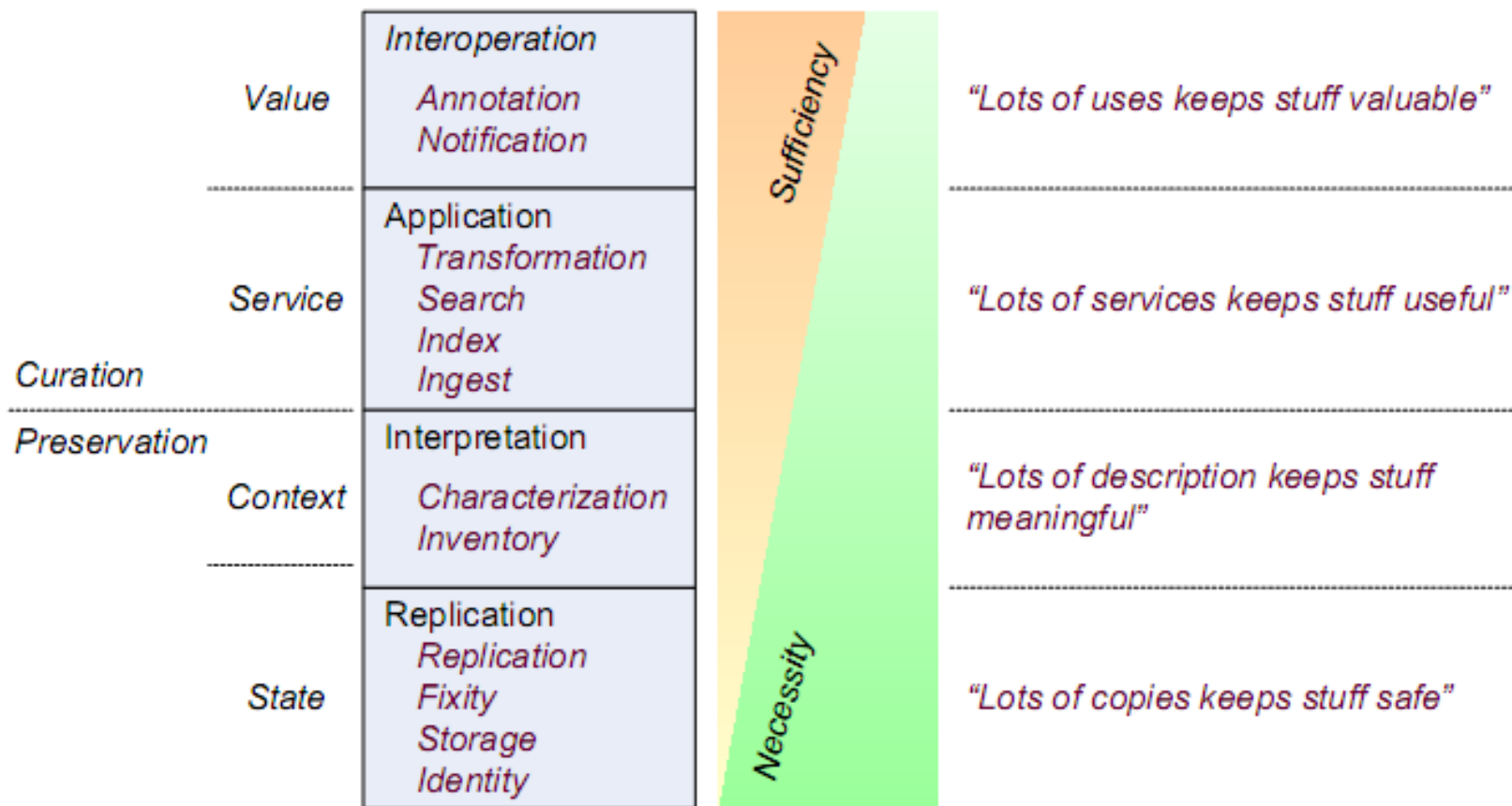
Merritt

University of California Curation Center

- Το Merritt είναι ένα υπό ανάπτυξη περιβάλλον ψηφιακής επιμέλειας, που βασίζεται στις μικρο-υπηρεσίες και τη σύνθεση αυτών.
- Μικρο-υπηρεσίες: αυτοτελείς, μικρές, διαλειτουργικές
 - Ευκολότερο να αναπτυχθούν, χρησιμοποιηθούν, συντηρηθούν, αντικατασταθούν
- Σύνθεση: παράγει λειτουργίες επιμέλειας κατά τις ανάγκες
- Το αποθετήριο ως λογικός, όχι φυσικός, τόπος
- Θεμελιώδεις απαιτήσεις
 - Ασφάλεια δεδομένων
 - Διατήρηση νοήματος
 - Διευκόλυνση χρήσης
 - Επαύξηση της αξίας



«Μικρο-υπηρεσίες» επιμέλειας του Merritt



Θεμελιώδεις υπηρεσίες: identity, ingest, storage, inventory



ΜΟΨΕΑΣ

Μονάδα Ψηφιακής Επιμέλειας – ΙΠΣΥΠ/Ε.Κ. «Αθηνά»

Υποδομή Επιμέλειας και Μακροπρόθεσμης διατήρησης ψηφιακών συλλογών

- Ο ΜΟΨΕΑΣ είναι ένα σύστημα αποθετηρίου και επιμέλειας ψηφιακών πόρων και αναπτύσσεται με τη χρήση προγραμμάτων ανοικτού λογισμικού.
- Στόχος του ΜΟΨΕΑ είναι να καλύψει τις βασικές λειτουργίες ανάπτυξης και διαχείρισης ενός αποθετηρίου και να αποτελέσει ένα ισχυρό εργαλείο για τη μακρόχρονη διατήρηση ψηφιακού υλικού.



ΜΟΨΕΑΣ : βασικές λειτουργίες

- Αποτελεσματική διαδικασία μετάπτωσης δεδομένων από διαφορετικά αποθετήρια
- Ενσωματωμένοι μηχανισμοί διατήρησης δεδομένων – μεταδεδομένων
- Σηματολογική συσχέτιση
- Μηχανισμός αυτόματης αντιστοίχισης – μετασχηματισμών μεταδεδομένων
- Συμβατότητα με πρότυπα διαλειτουργικότητας (όπως OAI-PMH)
- Αποδοτικοί μηχανισμοί ευρετηρίασης μεγάλου όγκου δεδομένων
- Διαχείριση σύνθετων αντικειμένων (συλλογές, πολλαπλές ψηφιακές παραστάσεις δεδομένων)
- Μηχανή ροών εργασίας
- Σύνδεση με περιβάλλον σχεδιασμού διατήρησης - PLATO



ΜΟΨΕΑΣ : Διατήρηση

Αρχεία

- Δυνατότητα κλιμάκωσης με χρήση iRODS grids
- Δυνατότητα χρήσης πολλαπλών RDBMS για ευρετηρίαση
- Εύκολα backups
- Disaster recovery
 - Μηχανισμός αναδημιουργίας των ευρετηρίων από το file system

Μεταδεδομένα

- Κάθε εγγραφή στο αποθετήριο δημιουργεί
 - Νέα έκδοση στο datastream των μεταδεδομένων
 - Έγγραφή PREMIS η οποία συνοδεύει και περιγράφει την αλλαγή που έγινε
 - Έγγραφή σε FOXML η οποία συνοδεύει και περιγράφει την αλλαγή
- Για κάθε αντιστοίχιση μεταξύ σχημάτων μεταδεδομένων
 - Δημιουργείται νέο datastream μεταδεδομένων
 - Διατηρείται το αρχικό σχήμα των μεταδεδομένων και ο μηχανισμός αντιστοίχισης (provenance)



ΜΟΨΕΑΣ : Εισαγωγή και διαχείριση μεταδεδομένων

- Δυναμικός ορισμός σχημάτων μεταδεδομένων
 - Ο χρήστης περιγράφει σε απλή XML το σχήμα μεταδεδομένων που θέλει
 - Η αντίστοιχη υπηρεσία «μεταφράζει» την περιγραφή του σχήματος σε εύχρηστες φόρμες HTML
- Δυνατότητα περιγραφής περίπλοκων σχημάτων
 - Απεριόριστες ιεραρχίες
 - Απεριόριστα γνωρίσματα (attributes)
- Η περιγραφή του σχήματος παριστάνεται ως ψηφιακό αντικείμενο και αποθηκεύεται στον ΜΟΨΕΑ

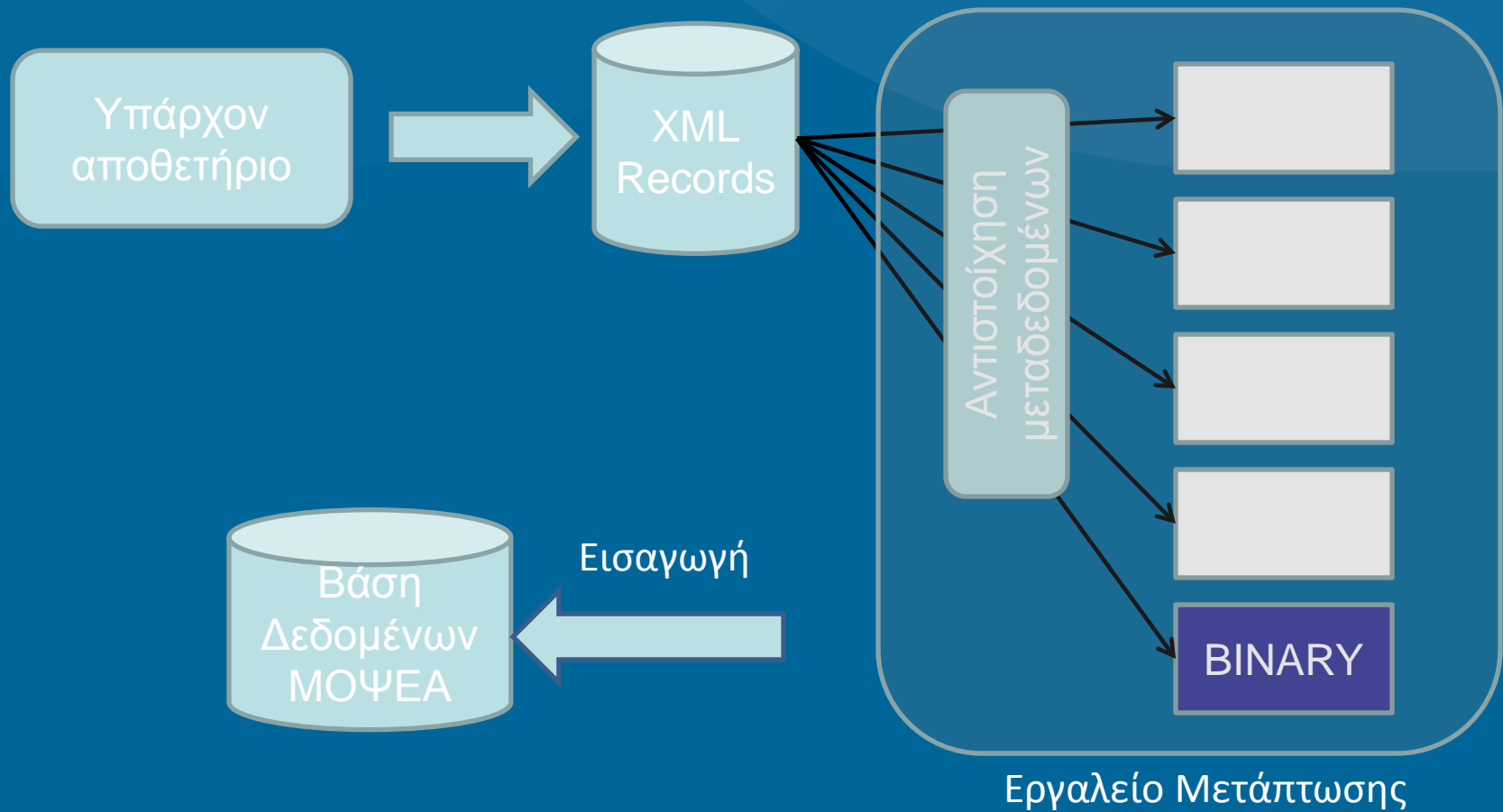


ΜΟΨΕΑΣ : Αυτόματος συγχρονισμός με σχεσιακή βάση δεδομένων

- Υπάρχον μοντέλο ευρετηρίων του Fedora-commons
 - Ευρετηριάζει μόνο το unqualified Dublin Core (15 πεδίων)
 - Δεν παρέχει ευελιξία
- Μοντέλο ΜΟΨΕΑ
 - Δυναμικά ευρετήρια για οποιοδήποτε πρότυπο μεταδεδομένων
 - Υποστηρίζονται ευρετήρια XML + RDF
 - Συγχρονισμός με MySQL (ή οποιαδήποτε RDBMS)
 - Η περιγραφή των ευρετηρίων σε Xpath
 - Ο μηχανισμό ευρετηρίασης παριστάνεται ως ψηφιακό αντικείμενο (μέρος του ΜΟΨΕΑ)



ΜΟΨΕΑΣ : Διαδικασία μετάπτωσης



Η τρέχουσα υλοποίηση έχει δοκιμαστεί με επιτυχία για το αποθετήριο Dspace.



Αντιμετώπιση της ετερογένειας

Συντακτική ετερογένεια:

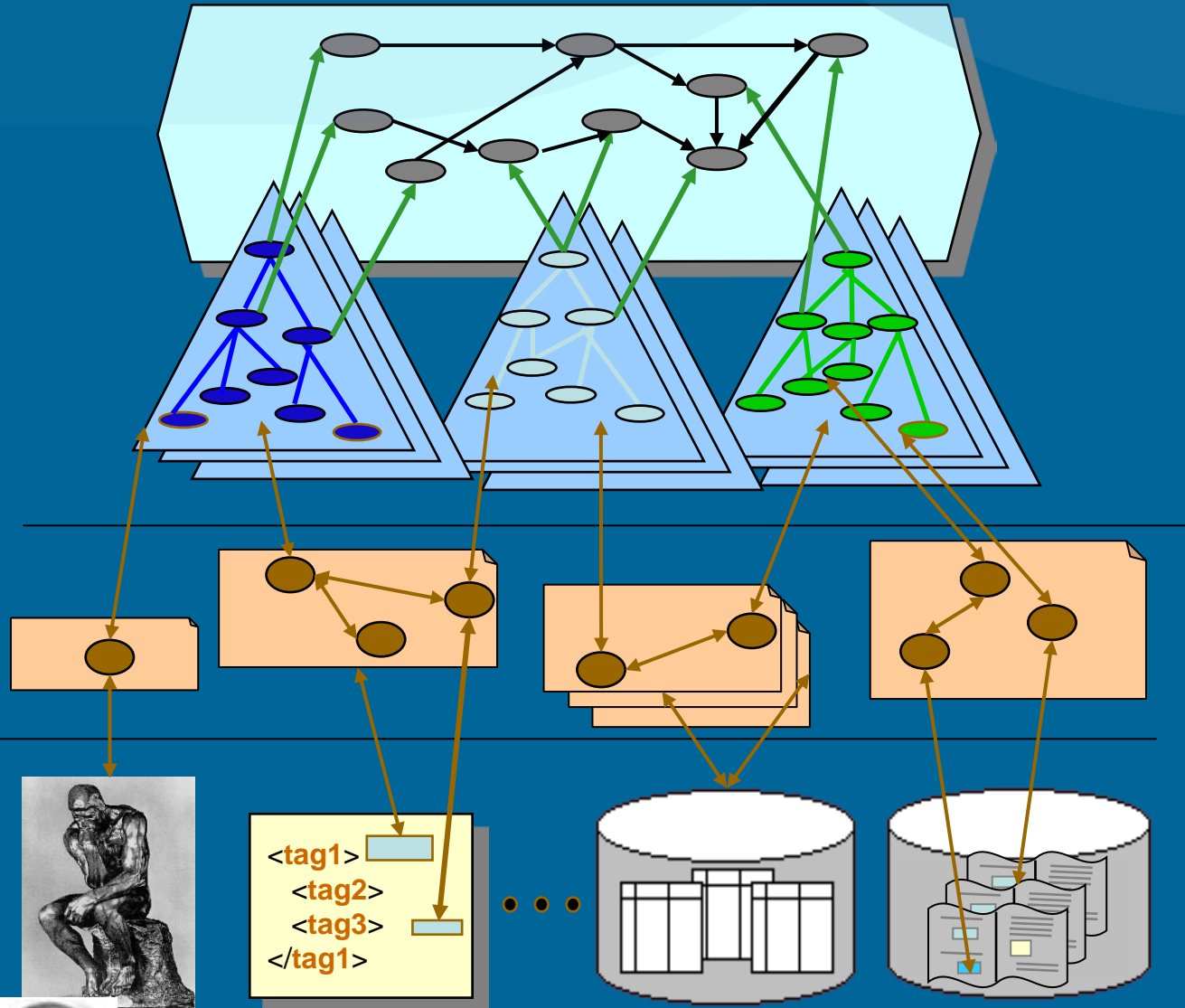
- Πλατφόρμες και λειτουργικά συστήματα
- Παραστάσεις δεδομένων και συμβάσεις πρόσβασης

Σημασιολογική ετερογένεια: η δυσχερέστερη στην αντιμετώπιση

- Προσέγγιση:
 - Διατήρηση της ποικιλίας και της τοπικότητας
 - Προώθηση συμβάσεων και σημασιολογικής συμβατότητας
- Μέσα:
 - Παράσταση γνώσεων
 - Οντολογίες και διαχείριση ορολογίας
 - Πλαίσια αναφοράς / συμφραζόμενα



Σημασιολογική ολοκλήρωση και διαλειτουργικότητα



Οντολογίες
Θησαυροί όρων

Σημασιολογικές
περιγραφές πόρων

Ποικιλία
πληροφοριακών
πόρων



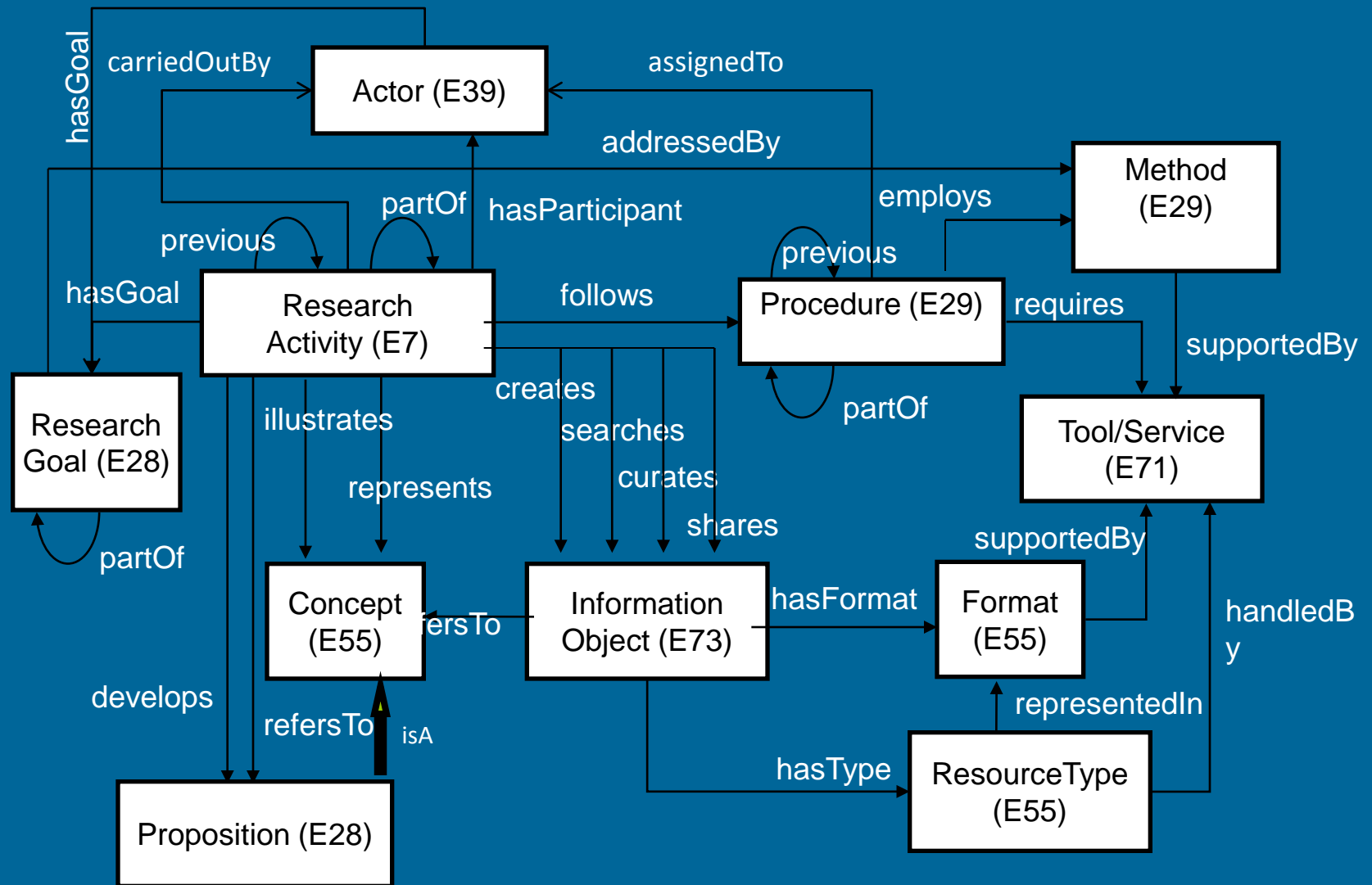
Μοντελοποίηση ερευνητικής δραστηριότητας

Benardou, A., et al., "Understanding the information requirements of arts and humanities scholarship: implications for digital curation". International Journal of Digital Curation, 5, no. 1, 2010

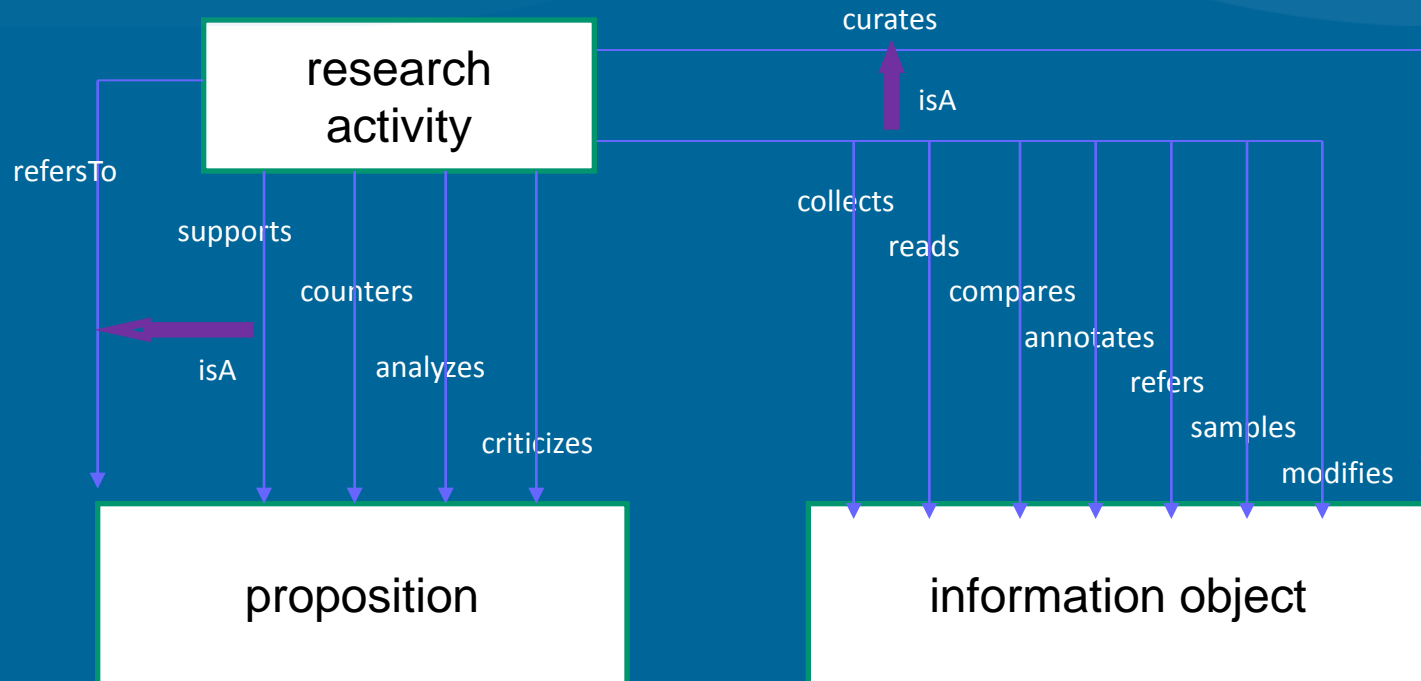
- Σκόπιμη διεργασία
- Εμπλέκει οντότητες διαφορετικών ειδών:
- **Υλικά αντικείμενα, φυσικά ή τεχνητά**
 - Αντικείμενα ή εργαλεία της εργασίας
- **Νοητικά αντικείμενα**
 - Έννοιες που δημιουργούνται, παριστάνονται, αναφέρονται
 - Λογικές προτάσεις που διατυπώνονται, αποδεικνύονται, αντικρούονται, κλπ.
- **Πληροφοριακά αντικείμενα**
 - Ειδική κατηγορία νοητικών αντικειμένων με αντίστοιχους υλικούς φορείς
 - Αναφέρονται σε και παριστάνουν υλικά και νοητικά αντικείμενα
 - Δημιουργούνται, αναζητούνται, κοινωνούνται και γίνονται αντικείμενο επιμέλειας κατά τη διάρκεια ερευνητικών δραστηριοτήτων



Μοντέλο ερευνητικής δραστηριότητας



«Πρωτογενείς» ενέργειες (scholarly primitives)



Επιμέλεια...

