

Διαλειτουργικότητα μεταδεδομένων

Σταύρος Αγγελής
Μονάδα Ψηφιακής Επιμέλειας, ΙΠΣΥ – Ε.Κ. Αθηνά
{s.angelis@dcu.gr}

Ανάγκη για διαλειτουργικότητα μεταδεδομένων

- Ετερογενή περιβάλλοντα, κοινότητες χρηστών, τύποι πόρων, δεδομένα και σχήματα μεταδεδομένων
- Δεν μπορούν όλα τα συστήματα να υποστηρίξουν διαδικασίες όπως η φόρτωση, η αποθήκευση και η εξαγωγή για διαφορετικά σχήματα μεταδεδομένων
- Περιορισμοί σχημάτων μεταδεδομένων στην κατάλληλη αναπαράσταση της πληροφορίας
- Ανάγκη για κεντρική αναζήτηση σε συλλογές με διαφορετικά σχήματα μεταδεδομένων

Είδη ετερογένειας

- Ετερογένεια συστήματος: ασυμβατότητες μεταξύ υλικού και λειτουργικών συστημάτων.
- Συντακτική ετερογένεια: ασυμφωνίες από τα διαφορετικά πρωτόκολλα, τις κωδικοποιήσεις και τις γλώσσες που χρησιμοποιούνται από τις πληροφοριακές πηγές (π.χ. γλώσσες ερωτημάτων, μορφότυπα μεταδεδομένων, κ.ά.)
- Ετερογένεια δομής: διαφορές που προκύπτουν από τα διαφορετικά μοντέλα δεδομένων, τις δομές ετερογένειας και τα σχήματα.
- Σημασιολογική ετερογένεια: αφορά στις σημασιολογικές διαφορές που προκύπτουν από το γεγονός ότι το νόημα των δεδομένων μπορεί να εκφραστεί με πολλούς διαφορετικούς τρόπους (π.χ. ομώνυμα και συνώνυμα, απεικονίσεις, κ.ά.)

Σε τι απαντάει η διαλειτουργικότητα μεταδεδομένων

- Ποιο σχήμα μεταδεδομένων είναι κατάλληλο για τη συλλογή μου;
- Πρέπει να χρησιμοποιήσω ένα υπάρχον σχήμα ή να δημιουργήσω ένα νέο;
- Πρέπει να αλλαχθεί το σχήμα;
- Μπορεί να χρησιμοποιηθεί υπάρχουσα πληροφορία στη νέα ψηφιακή συλλογή;
- Θέλω να ενώσω αντικείμενα από άλλη συλλογή που χρησιμοποιούν άλλο σχήμα μεταδεδομένων, μπορώ;
- Θέλω να κάνω τα δεδομένα μου διαθέσιμα σε ένα κεντρικό αποθετήριο (π.χ. europeana), μπορώ;
- Οι συλλογές του οργανισμού χρησιμοποιούν περισσότερα από ένα σχήματα μεταδεδομένων. Μπορώ να έχω ένα κεντρικό σύστημα αναζήτησης;

Διαλειτουργικότητα μεταδεδομένων

- Η διαλειτουργικότητα μεταδεδομένων είναι η δυνατότητα δύο ή περισσότερων πληροφοριακών συστημάτων να ανταλλάξουν μεταδεδομένα με την ελάχιστη δυνατή απώλεια πληροφορίας
- Στόχος η επαναχρησιμοποίηση και ανταλλαγή μεταδεδομένων
- Κανόνες που αφορούν το περιεχόμενο και το βαθμό λεπτομέρειας της πληροφορίας
- Χρειάζεται συμφωνία για τα σχήματα μεταδεδομένων και τα πρωτόκολλα ανταλλαγής που θα χρησιμοποιηθούν

Είδη διαλειτουργικότητας

- Συντακτική διαλειτουργικότητα: πώς οι δημιουργοί μεταδεδομένων αποθηκεύουν, ανταλλάσσουν και χρησιμοποιούν εγγραφές μεταδεδομένων από διαφορετικές πηγές (π.χ. XML, name spaces).
- Διαλειτουργικότητα δομής: πώς εκφράζουν τα διαφορετικά σχήματα μεταδεδομένων τις ίδιες πληροφορίες, χρησιμοποιώντας διαφορετική δομή, ενώ η διαλειτουργικότητα επιτυγχάνεται με τη χρήση των κατάλληλων μηχανισμών μετατροπής (π.χ. authorities).
- Σημασιολογική διαλειτουργικότητα: η ικανότητα των πληροφοριακών συστημάτων να μεταδίδουν κωδικοποιημένες πληροφορίες με την επιδιωκόμενη σημασία.

Εμπόδια διαλειτουργικότητας

- Σημασιολογική διαφορά ανάμεσα σε σχήματα μεταδεδομένων.
 - DC – MARC
- Διαφορετική εφαρμογή στην πράξη
 - Βιβλιοθήκες – Μουσεία
- Μη καθιερωμένος τρόπος τεκμηρίωσης
 - Κακριδής, Ιωάννης Θ. vs Κακριδής, Ι.Θ.
- Διαφορετικά λεξιλόγια
 - Χρήση διαφορετικών πηγών θεματικής ευρετηρίασης (θεματικές επικεφαλίδες, θησαυροί)

Εμπόδια διαλειτουργικότητας

- Τεκμήρια και συλλογές
 - Απώλεια πληροφοριών από τη συνδυασμένη αναζήτηση σε απλά σχήματα μεταδεδομένων (DC) με σύνθετα και πολυεπίπεδα σχήματα (EAD).
- Πολλαπλές εκδόσεις υλικού
 - Προβλήματα με την περιγραφή των αναπαραγωγών διαφόρων έργων, π.χ. ψηφιακές φωτογραφίες.
- Πολυγλωσσία
 - Παραδοσιακά: χρήση ελεγχόμενων λεξιλογίων από πολύγλωσσους θησαυρούς.
 - Τώρα: ανάπτυξη πολύγλωσσων μητρώων σχημάτων μεταδεδομένων καθιέρωση αντιστοιχίσεων μεταξύ των ονομάτων των στοιχείων.

Επίπεδα διαλειτουργικότητας μεταδεδομένων

- σε επίπεδο σχήματος
 - derivation
 - προφίλ εφαρμογής
 - διασταυρώσεις (crosswalks)
 - πλαίσια διαλειτουργικότητας
 - μητρώα μεταδεδομένων
- σε επίπεδο αρχείου μεταδεδομένων
 - μετατροπή δεδομένων
- σε επίπεδο αποθετηρίου
 - ΟΑΙ-ΡΜΗ
 - εμπλουτισμός μεταδεδομένων

Διαλειτουργικότητα σε επίπεδο σχήματος

- Η διαλειτουργικότητα σε επίπεδο σχήματος εστιάζει σε στοιχεία των σχημάτων μεταδεδομένων που είναι ανεξάρτητα από εφαρμογές
- Οι δράσεις εκτελούνται κατά τη διάρκεια του σχεδιασμού ενός έργου πριν αρχίσει η καταγραφή των δεδομένων
- Προκύπτουν νέα σύνολα από στοιχεία ή κωδικοποιημένα σχήματα, crosswalks, προφίλ εφαρμογής και μητρώα σχημάτων μεταδεδομένων

Derivation

- Ένα νέο σχήμα προκύπτει από ένα υπάρχον ανάλογα με τις ανάγκες του έργου
- Ένα πλούσιο σχήμα χρησιμοποιείται ως «πηγή» και από αυτό προκύπτει ένα πιο απλό σχήμα
- Μέθοδοι derivation μπορεί να είναι η προσαρμογή, τροποποίηση, επέκταση, μερική προσαρμογή, μετάφραση κτλ
- Το νέο σχήμα που δημιουργείται (στόχος) εξαρτάται πάντα από το σχήμα πηγή
- Το νέο σχήμα έχει όμοια βασική δομή και κοινά στοιχεία με το σχήμα πηγή
- παραδείγματα : TEI -> TEI Lite, MARC21 -> MARC Lite, MARC21 -> MODS κ.α.

Προφίλ εφαρμογής 1/2

- Ένα προφίλ εφαρμογής αποτελεί μια δήλωση σχετικά με ποιους όρους μεταδεδομένων χρησιμοποιεί στα μεταδεδομένα του ένας οργανισμός, ένας πληροφοριακός πόρος, μια εφαρμογή ή μια κοινότητα χρηστών*
- Αντιμετωπίζει το πρόβλημα ιδιαίτερων χαρακτηριστικών και απαιτήσεων χρήσης μέσα σε κοινότητες χρηστών
- Η χρήση προφίλ εφαρμογής εξασφαλίζει κοινή βασική δομή και κοινά στοιχεία αλλά παράλληλα επιτρέπει διαφορετικά επίπεδα λεπτομέρειας και βάθους ανάλογα με τις ανάγκες διαφορετικών κοινοτήτων χρηστών

*Baker, T. (2003). DCMI Usage Board review of application profiles. Available: <<http://dublincore.org/usage/documents/profiles/index.shtml>>.

Προφίλ εφαρμογής 2/2

- Τα προφίλ εφαρμογής αποτελούνται από στοιχεία μεταδεδομένων που προκύπτουν από ένα ή περισσότερα σχήματα μεταδεδομένων που συνδυάζονται σε ένα νέο συμπαγές σχήμα και βελτιστοποιούνται για μια συγκεκριμένη κοινότητα και ανάγκη
- Ένα προφίλ εφαρμογής δεν μπορεί να ορίσει νέα στοιχεία, αλλά μπορεί να ορίσει επιτρεπόμενες τιμές ή επιτρεπόμενες εφαρμογές και να εξειδικεύσει τον ορισμό των στοιχείων
- παραδείγματα: DC -> DCCAP, DC -> DC-Lib

Διασταυρώσεις 1/3

- Μια διασταύρωση αποτελεί μια αντιστοίχιση των στοιχείων, της σημασιολογίας και του συντακτικού ενός σχήματος μεταδεδομένων σε ένα άλλο*
- Αποτελούν την πιο διαδεδομένη προσέγγιση και υπάρχουν διαθέσιμες για τα περισσότερα διαδεδομένα σχήματα μεταδεδομένων
- Η μέθοδος που χρησιμοποιείται είναι η απευθείας αντιστοίχιση ισοδύναμων ή σχεδόν ισοδύναμων στοιχείων ανάμεσα στα σχήματα

*NISO (2004) Understanding Metadata. Available: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>.

Διασταυρώσεις 2/3

- Σημαντικά χαρακτηριστικά:
 - ο σημασιολογικός ορισμός του κάθε στοιχείου μεταδεδομένων
 - αν το στοιχείο είναι υποχρεωτικό, υποχρεωτικό υπό προϋποθέσεις ή προαιρετικό
 - αν το στοιχείο είναι πολλαπλό
 - περιορισμοί που προκύπτουν από την οργάνωση των στοιχείων μεταδεδομένων, π.χ. ιεραρχικές σχέσεις
 - περιορισμοί που επιβάλλονται στην τιμή ενός στοιχείου, π.χ. κείμενο, αριθμός, ημερομηνία κ.α.
 - υποστήριξη σε τοπικά καθορισμένα στοιχεία μεταδεδομένων



Διασταυρώσεις 3/3

- Απόλυτη αντιστοίχιση
 - απαιτεί απόλυτη αντιστοίχιση ανάμεσα στα στοιχεία του σχήματος “πηγή” και του σχήματος “στόχο”
 - όπου δεν υπάρχει απόλυτη ισοδυναμία δεν πραγματοποιείται αντιστοίχιση
 - αποτέλεσμα η απώλεια πληροφορίας (συνήθως από πιο “πλούσια” σε πιο “φτωχά” σχήματα μεταδεδομένων)
- Σχετική αντιστοίχιση
 - απαιτεί την αντιστοίχιση όλων των στοιχείων του σχήματος “πηγή” στο σχήμα “στόχο”
 - έχει καλύτερα αποτελέσματα όταν πάμε από ένα σύνθετο σχήμα μεταδεδομένων σε ένα πιο απλό αλλά δεν ισχύει το αντίθετο

Πλαίσιο διαλειτουργικότητας μεταδεδομένων

- Ένα πλαίσιο μεταδεδομένων αποτελεί μια κοινή βάση πάνω στην οποία μπορούν να αναπτυχθούν σχήματα μεταδεδομένων
- Ένα πλαίσιο μπορεί είτε να αναπτυχθεί πριν από τα σχήματα και τις εφαρμογές ή να βασιστεί πάνω σε υπάρχοντα σχήματα
- Αποτελεί ένα σύνολο από γενικούς κανόνες που οδηγούν σε σχήματα με κοινά χαρακτηριστικά για διαφορετικές κοινότητες
- παράδειγμα : OAI5

Μητρώο μεταδεδομένων 1/3

- Ένα μητρώο μεταδεδομένων αποτελεί ένα μητρώο με πληροφορίες σχετικά με σχήματα μεταδεδομένων
- Τα μητρώα μεταδεδομένων αποτελούν εργαλεία που βοηθούν στην αναγνώριση και αναφορά καθιερωμένων σχημάτων και προφίλ εφαρμογών
- Τα μητρώα μπορούν να έχουν τα χαρακτηριστικά ενός ηλεκτρονικού λεξικού
- Τα μητρώα μεταδεδομένων που αναπτύσσονται σήμερα βασίζονται στο πρότυπο ISO 11179
- Χρησιμοποιούν ελεγχόμενα λεξιλόγια και θησαυρούς

Μητρώα μεταδεδομένων 2/3

- Τα μητρώα μπορεί να περιέχουν πληροφορία σχετικά με:
 - μοντέλα δεδομένων
 - στοιχεία
 - σύνολα στοιχείων
 - σχήματα κωδικοποίησης
 - προφίλ εφαρμογής
 - πληροφορία για τη χρήση των στοιχείων
 - διασταυρώσεις σχημάτων



Μητρώα μεταδεδομένων 3/3

- Μητρώα δεδομένων ανάλογα με το σκοπό:
 - Cross-domain, cross-schema μητρώο
 - Domain-specific, cross-schema μητρώο
 - Project-specific μητρώο
 - Schema-specific μητρώο

Διαλειτουργικότητα σε επίπεδο αρχείου μεταδεδομένων

- Η διαλειτουργικότητα σε επίπεδο αρχείων μεταδεδομένων έχει ως στόχο την ολοκλήρωση μεταδεδομένων μέσα από αντιστοίχιση της σημασιολογίας των στοιχείων τους
- Οι δράσεις εκτελούνται σε προχωρημένο στάδιο ενός έργου όπου έχουν ήδη δημιουργηθεί αντικείμενα
- Προκύπτουν αρχεία μεταδεδομένων μετά από μετατροπή και νέα αρχεία μεταδεδομένων μετά από συνδυασμό τιμών από υπάρχοντα αρχεία μεταδεδομένων

Μετατροπή αρχείων μεταδεδομένων

- Ο στόχος αυτής της διαδικασίας είναι να ελαχιστοποιηθεί η απώλεια ή η παραμόρφωση πληροφορίας
- Έχουν αναπτυχθεί διάφορα εργαλεία γι αυτό το σκοπό, π.χ. για τη μετατροπή ανάμεσα σε MARC και MODS, ανάμεσα σε DC και MODS κ.α. (<http://www.loc.gov/standards/mods/>)
- Οι αντιστοιχήσεις που στηρίζονται σε πραγματικά δεδομένα αντί για σχήματα δεδομένων παρουσιάζουν μεγαλύτερη δυσκολία

Διαλειτουργικότητα σε επίπεδο αποθετηρίου

- Η πληροφορία που περιέχεται σε ένα αποθετήριο πρέπει να έχει συνέπεια ώστε να είναι εύκολη τόσο η ανάκτηση όσο και η παρουσίαση
- Ένα αποθετήριο μπορεί να περιέχει πληροφορία η οποία προέρχεται από διαφορετικούς παρόχους περιεχομένου, τεκμηριωμένη με διαφορετικό σχήμα μεταδεδομένων
- Η διαλειτουργικότητα σε ένα αποθετήριο εμπλέκεται σε διάφορες από τις διεργασίες που εκτελούνται όπως η συγκομιδή δεδομένων, η υποστήριξη πολλαπλών τύπων αντικειμένων, το aggregation, οι υπηρεσίες αντιστοίχισης, ο εμπλουτισμός, η ανάκτηση πληροφορίας, η εξαγωγή δεδομένων κ.α.

ΟΑΙ-ΡΜΗ

- Αποτελεί ένα μηχανισμό για διαλειτουργική συγκομιδή μεταδεδομένων ανάμεσα σε αποθετήρια μέσα από τεχνολογίες διαδικτύου
- Στόχος της Open Archives Initiative είναι η ανάπτυξη και η παροχή προτύπων διαλειτουργικότητας για τη διευκόλυνση διάδοσης του περιεχομένου
- Αποθετήρια με το ρόλο του παρόχου πληροφορίας κάνουν διαθέσιμη δομημένη πληροφορία μέσω ΟΑΙ-ΡΜΗ
- Πάροχοι υπηρεσιών εκτελούν ερωτήματα στην υπηρεσία ΟΑΙ-ΡΜΗ για να συλλέξουν αυτή την πληροφορία

Εμπλουτισμός μεταδεδομένων

- Εμπλουτισμός με θεματικές επικεφαλίδες
- Εμπλουτισμός με βάση το κοινό περιεχόμενο
- Ημιαυτόματος εμπλουτισμός χωρικής πληροφορίας
- Ημιαυτόματος εμπλουτισμός χρονικής πληροφορίας
- Αρχικά αρχεία μεταδεδομένων => εμπλουτισμένα αρχεία μεταδεδομένων

Συμπεράσματα

- Η διαλειτουργικότητα μεταδεδομένων εξασφαλίζει την επαναχρησιμοποίηση περιεχομένου
- Επιτρέπει την ομογενοποίηση δεδομένων,
 - επιτρέπει την κεντρική αναζήτηση διαφορετικού τύπου αντικειμένων
 - επιτρέπει την ομοιογενή παρουσίαση διαφορετικού τύπου αντικειμένων
 - επιτρέπει την ανταλλαγή δεδομένων (απελευθερώνει την πληροφορία από το σύστημα στο οποίο περιέχεται)
- Επιτρέπει την αξιοποίηση περιεχομένου έξω από το σκοπό δημιουργίας του
- Συνεισφέρει στην ψηφιακή διατήρηση
- Οδηγεί στη δημιουργία καλύτερων συστημάτων– αποθετηρίων

Πηγές

- Baker, T. (2003). DCMI Usage Board review of application profiles. Διαθέσιμο στο: <http://dublincore.org/usage/documents/profiles/index.shtml>
- NISO (National Information Standards Organization). (2004). Understanding metadata. Bethesda, MD: NISO Press. Διαθέσιμο στο: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
- Zeng, M., & Chan, L. (2006). Metadata Interoperability and Standardization – A Study of Methodology Part I. D-Lib Magazine, 12(6). Διαθέσιμο στο: <http://www.dlib.org/dlib/june06/chan/06chan.html>
- Zeng, M., & Chan, L. (2006). Metadata Interoperability and Standardization – A Study of Methodology Part II. D-Lib Magazine, 12(6). Διαθέσιμο στο: <http://www.dlib.org/dlib/june06/zeng/06zeng.html>