

ΜΕΡΟΣ ΙΙΙ:
Πρότυπα Κωδικοποίησης
βασισμένα στην XML

Πίνακας Περιεχομένων

ΚΕΦΑΛΑΙΟ 7:	85
Κωδικοποίηση Κειμένου - Το πρότυπο TEI.....	85
7.1. Εισαγωγή	85
7.2. Βασική δομή.....	86
7.3. Βασικά μεταδεδομένα στο teiHeader	89
7.3.1. Το στοιχείο titleStmt	90
7.3.2. Το στοιχείο publicationStmt	90
7.3.3. Το στοιχείο sourceDesc	90
7.4. Στοιχεία για την κωδικοποίηση κειμένου	91
7.4.1. Δομικά στοιχεία	92
7.4.2. Στοιχεία Φράσεων	94
7.4.3. Γραμμές και σελίδες	94
7.4.4. Αναφορές και σημειώσεις	95
7.4.5. Βιβλιογραφικές Αναφορές	97
7.5. Εξώφυλλο, οπισθόφυλλο και παραρτήματα	97
7.6. Άλλα στοιχεία	99
7.7. Ένα ολοκληρωμένο παράδειγμα	99
ΚΕΦΑΛΑΙΟ 8: MARC XML	103
8.1. Εισαγωγή	103
8.2. Χαρακτηριστικά, πλεονεκτήματα - μειονεκτήματα	104
8.3. Δομή	105
8.3.1. Στοιχεία, Σύνθετοι τύποι και απλοί τύποι	105
8.3.2. Συλλογές και εγγραφές	105
8.4. Πεδία και υπο-πεδία	107
8.5. Ένα ολοκληρωμένο Παράδειγμα	111
8.6. Άλλες υλοποιήσεις	115
Βιβλιογραφία	116
Ευρετήριο όρων	0

ΚΕΦΑΛΑΙΟ 7:

Κωδικοποίηση Κειμένου - Το πρότυπο TEI

7.1. Εισαγωγή

Με τον όρο *κωδικοποίηση κειμένου* (text markup, text encoding) εννοούμε τη διαδικασία διάκρισης δομικών ή σημασιολογικών (semantic) χαρακτηριστικών κειμένου με βάση κάποιους κανόνες. Το πλέον διαδεδομένο πρότυπο κωδικοποίησης κειμένου, το οποίο παρουσιάζεται στις ακόλουθες παραγράφους, ονομάζεται *Text Encoding Initiative* (TEI). Η ανάπτυξή του TEI ξεκίνησε τα τέλη της δεκαετίας του '80, ενώ το 1999 σχηματίστηκε το TEI consortium ως ένα έργο των ανθρωπιστικών επιστημών επιχορηγούμενο από γνωστούς και μεγάλους επαγγελματικούς Οργανισμούς (<http://www.tei-c.org>).

Στόχος του προτύπου είναι να δημιουργήσει ένα περιβάλλον για την κωδικοποίηση κειμένων ακαδημαϊκού ενδιαφέροντος, έτσι ώστε να μπορούν να μεταγράφονται και να διατηρούνται ανεξάρτητα από την εκάστοτε τεχνολογία. Φιλοδοξεί να καλύψει όλες τις μορφές του γραπτού λόγου (πεζός λόγος-prose, ποίηση-verse, δράμα-drama), οποιασδήποτε γλώσσας και χρονικής περιόδου. Το TEI έχει γίνει αποδεκτό διεθνώς και σήμερα αποτελεί το κατ' εξοχήν πρότυπο για την ηλεκτρονική κωδικοποίηση κειμένων και επομένως την ανάπτυξη της Ηλεκτρονικής Εκδοτικής (e-publishing).

Το TEI αποτελεί ένα SGML-DTD που απεικονίζει πρωτίστως τη *λογική* που διέπει τη δόμηση ενός κειμένου και δευτερευόντως τη *μορφοποίησή* (τον τρόπο εμφάνισής) του. Για παράδειγμα η πρόταση:

*There are very few **risqué** passages in **Paradise Lost***

κωδικοποιείται από την HTML, η οποία είναι μια γλώσσα *μορφοποίησης* κειμένου, ως εξής:

There are very few <i>risqué</i> passages in <i>Paradise Lost</i>

δηλώνοντας με το στοιχείο <i> ότι η λέξη *risqué* και η έκφραση *Paradise Lost* γράφονται με πλάγια γραμματοσειρά (italics).

Αντίθετα το TEI κωδικοποιεί την ίδια πρόταση ως εξής:

There are very few <foreign>risqué</foreign> passages in <title>Paradise Lost</title>

αναδεικνύοντας τα σημασιολογικά της χαρακτηριστικά της πρότασης. Συγκεκριμένα αναφέρεται ότι η λέξη «*risqué*» ανήκει σε άλλη γλώσσα από τις υπόλοιπες λέξεις της πρότασης (χρησιμοποιώντας το στοιχείο *<foreign>*), ενώ η έκφραση *Paradise Lost* είναι τίτλος (χρησιμοποιώντας το στοιχείο *<title>*).

Συνοπτικά το TEI DTD κωδικοποιεί:

- Τη δομή ενός κειμένου (παράγραφοι, σελίδοποίηση, διάλογοι, υποσημειώσεις, παραπομπές, βιβλιογραφικές αναφορές, στοιχεία του λόγου)
- Γλωσσική επεξεργασία (διάλεκτοι, ονόματα, προτάσεις, λέξεις, εκφράσεις, στοιχεία μετάφρασης)
- Μεταδεδομένα (βιβλιογραφικά στοιχεία, εκδοτικό ιστορικό, παρεμβάσεις σχολιαστών, εκδοτών κ.λπ.)
- Την εμφάνιση του κειμένου.

7.2. Βασική δομή

Τα βασικά στοιχεία (elements) του TEI-DTD είναι:

- το στοιχείο **teiHeader**, το οποίο δεν αποτελεί μέρος του υπο κωδικοποίηση κειμένου, αλλά παρέχει πληροφορίες (μεταδεδομένα) για αυτό. Αυτό το στοιχείο αποτελείται από επί τα μέρους στοιχεία:
 - **fileDesc** υποχρεωτικό, το οποίο περιέχει στοιχεία τίτλου, έκδοσης και προέλευσης του κωδικοποιημένου κειμένου
 - **encodingDesc** προαιρετικό, το οποίο περιγράφει τις μεθόδους και τις εκδοτικές αρχές στις οποίες βασίστηκε η μεταγραφή και κωδικοποίηση του κειμένου
 - **profileDesc** προαιρετικό, το οποίο παρέχει πληροφορίες για τον τρόπο δημιουργίας του κειμένου, τη γλώσσα του και τη θεματική του ταξινόμηση
 - **revisionDesc** προαιρετικό, το οποίο καταγράφει τις αλλαγές που έγιναν στο κωδικοποιημένο κείμενο
- το στοιχείο **text** το οποίο περιλαμβάνει το κωδικοποιημένο κείμενο. Αυτό το στοιχείο αποτελείται από τα επιμέρους στοιχεία:
 - **front**, (προαιρετικό) το οποίο περιέχει προκαταρκτικό περιεχόμενο (επικεφαλίδες, σελίδες τίτλων, πρόλογοι κ.λπ.) που βρίσκονται πριν την αρχή του κανονικού κειμένου
 - **body** (υποχρεωτικό) το οποίο περιέχει το σώμα ενός μοναδικού κειμένου εκτός του front και back περιεχομένου.
 - **back** (προαιρετικό) το οποίο περιέχει παραρτήματα κ.λπ. που ακολουθούν το κυρίως κείμενο

Έτσι η δομή ενός κωδικοποιημένου κειμένου σύμφωνα με το πρότυπο TEI έχει τη μορφή:

```

<TEI.2>
<teiHeader> <!--περιεχόμενο του στοιχείου --> </teiHeader>
<text>
    <front>
        <!-- εξώφυλλο (front matter). -->
    </front>
    <body>
        <!-- κυρίως κείμενο. -->
    </body>
    <back>
        <!-- οπισθόφυλλο (back matter). -->
    </back>
</text>
</TEI.2>

```

Οι XML DTD δηλώσεις που αντιστοιχούν στα παραπάνω στοιχεία είναι:

```

<!ELEMENT TEI.2
    (teiHeader, text) >
<!ELEMENT teiHeader
    (fileDesc, encodingDesc*, profileDesc*, revisionDesc?) >
<!ELEMENT text
    ((index | interp | interpGrp | lb | milestone | pb | gap
    | anchor)*, (front, (index | interp | interpGrp | lb | milestone
    | pb | gap | anchor)*)?, (body | group), (index | interp
    | interpGrp | lb | milestone | pb | gap | anchor)*, (back,
    (index | interp | interpGrp | lb | milestone | pb | gap
    | anchor)*)? ) >
<!ELEMENT group
    ((argument | byline | docAuthor | docDate | epigraph | head
    | opener | salute | signed | index | interp | interpGrp
    | lb | milestone | pb | gap | anchor)*, (text | group),
    (text | group | index | interp | interpGrp | lb | milestone
    | pb | gap | anchor)*, ((byline | closer | epigraph | salute
    | signed | trailer), (index | interp | interpGrp | lb |
    milestone | pb | gap | anchor)*)* ) >

```

Τα βασικότερα γνωρίσματα (attributes) τα οποία εφαρμόζονται σε όλα τα στοιχεία του TEI-DTD είναι τα:

- **id** (γνώρισμα ταυτότητας δηλ. κωδικό όνομα ή αριθμός)
- **n** (αριθμός ή όνομα)
- **lang** (γλώσσα, π.χ. **lang="FR"** για δήλωση γαλλικών)
- **rend** (τυπογραφική αναπαράσταση π.χ. **rend="italics"** για δήλωση πλάγιας γραφής).

Με βάση τα παραπάνω η ολοκληρωμένη δήλωση του στοιχείου `teiHeader` του TEI-DTD είναι η ακόλουθη:

```
<!ELEMENT teiHeader
    (fileDesc, encodingDesc*, profileDesc*, revisionDesc?) >
<!ATTLIST teiHeader
    ana IDREFS #IMPLIED
    corresp IDREFS #IMPLIED
    next IDREF #IMPLIED
    prev IDREF #IMPLIED
    id ID #IMPLIED
    n CDATA #IMPLIED
    lang IDREF #IMPLIED
    rend CDATA #IMPLIED
    type CDATA "text"
    creator CDATA #IMPLIED
    status (new | update) "new"
    date.created CDATA #IMPLIED
    date.updated CDATA #IMPLIED
    TEIform CDATA "teiHeader" >
```

Τέλος το TEI καλύπτει την κωδικοποίηση συλλογών κειμένων παρέχοντας το στοιχείο `group`. Η κωδικοποίηση μιας συλλογής κειμένων θα είχε τη μορφή:

```
<TEI.2>
  <teiHeader>
    [περιεχόμενο του στοιχείου teiHeader της συλλογής]
  </teiHeader>
  <text>
    <front>
      [περιεχόμενο του στοιχείου front της συλλογής ]
    </front>
  <group>
```

```

<text> <!-- Εδώ αρχίζει το πρώτο κείμενο -->
  <front>
    [περιεχόμενο του στοιχείου front του πρώτου κειμένου]
  </front>
  <body>
    [περιεχόμενο του στοιχείου body του πρώτου κειμένου]
  </body>
  <back>
    [περιεχόμενο του στοιχείου back του πρώτου κειμένου]
  </back>
</text>
<text> <!-- Εδώ αρχίζει το δεύτερο κείμενο -->
  <front>
    [περιεχόμενο του στοιχείου front του δεύτερου κειμένου]
  </front>
  <body>
    [περιεχόμενο του στοιχείου body του δεύτερου κειμένου]
  </body>
  <back>
    [περιεχόμενο του στοιχείου back του δεύτερου κειμένου]
  </back>
</text>
  [σε αυτό το σημείο μπορεί να ακολουθούν και περισσότερα
   κείμενα]
</group>
<back>
  [περιεχόμενο του στοιχείου back της συλλογής]
</back>
</text>
</TEI.2>

```

Μια από τις πολλές XML υλοποιήσεις του TEI-DTD, γνωστή ως TEI Lite, δίνεται στη διεύθυνση: <http://www.tei-c.org/Lite/DTD/teixlite.dtd>.

7.3. Βασικά μεταδεδομένα στο **teiHeader**

Το στοιχείο **fileDesc** είναι υποχρεωτικό υποστοιχείο του **teiHeader** και μπορεί να περιέχει ένα και μόνο ένα στοιχείο **titleStmt**, ένα **publicationStmt** και

τουλάχιστον ένα **sourceDesc** στοιχείο. Τα υπόλοιπα στοιχεία είναι προαιρετικά. Η δήλωση του **fileDesc** στοιχείου έχει ως εξής:

```
<!ELEMENT fileDesc
  (titleStmt, editionStmt?, extent?, publicationStmt, seriesStmt?,
  notesStmt?, sourceDesc+) >
```

7.3.1. Το στοιχείο **titleStmt**

Το υποχρεωτικό στοιχείο του τίτλου πρέπει να είναι μοναδικό. Η δήλωση του στοιχείου **titleStmt** είναι η ακόλουθη:

```
<!ELEMENT titleStmt
  ((index | interp | interpGrp | lb | milestone | pb | gap
  | anchor)*, (title, (index | interp | interpGrp | lb | milestone
  | pb | gap | anchor)*), ((author | editor | sponsor | funder
  | principal | respStmt), (index | interp | interpGrp | lb
  | milestone | pb | gap | anchor)*))* >
```

7.3.2. Το στοιχείο **publicationStmt**

Το στοιχείο **publicationStmt** παρέχει πληροφορία για την έκδοση τη διανομή (distribution) ενός ηλεκτρονικού ή συμβατικού κειμένου. Το μοναδικό στοιχείο του **publicationStmt** (που εμφανίζεται υποχρεωτικά μια ή περισσότερες φορές) είναι το **p** (που δηλώνει παράγραφο) καθώς επίσης και ένα από τα ακόλουθα: **publisher**, **distributor**, **authority**, **pubPlace**, **address**, **idno**, **availability**, **date**. Η δήλωση αυτού του στοιχείου είναι η ακόλουθη:

```
<!ELEMENT publicationStmt
  ((p, (index | interp | interpGrp | lb | milestone | pb |
  gap | anchor)*)+ | ((publisher | distributor | authority
  | pubPlace | address | idno | availability | date), (index
  | interp | interpGrp | lb | milestone | pb | gap | anchor)*))+ >
```

7.3.3. Το στοιχείο **sourceDesc**

Το στοιχείο **sourceDesc** παρέχει πληροφορία για την πηγή ή τις πηγές από τις οποίες προέκυψε το κωδικοποιημένο κείμενο. Η πληροφορίες αυτές μπορεί να έχουν τη μορφή είτε πεζού λόγου είτε βιβλιογραφικών αναφορών.

Παράδειγμα 7.1.

Ένα ολοκληρωμένο παράδειγμα του στοιχείου **teiHeader** και των υποστοιχείων του είναι το ακόλουθο:

```
<teiHeader>
  <filedesc>
    <titleStmt>
      <title>The Riders of the Purple Sage</title>
      <author>Zane Grey</author>
    </titleStmt>
    <publicationStmt>
      <publisher>Oxford University Press</publisher>
      <pubPlace>Oxford</pubPlace>
      <date>1989</date>
      <idno type="ISBN"> 0-19-254705-5</idno>
      <availability>Copyright 1989, Oxford University Press</availability>
    </publicationStmt>
    <sourcedesc>
      <bibl>
        <author> Zane Grey</author>
        <title>The Riders of the Purple Sage</title>
        <date>12 Jun 1989</date>
      </bibl>
    </sourceDesc>
  </filedesc>
</teiHeader>
```

7.4. Στοιχεία για την κωδικοποίηση κειμένου

Στην παράγραφο αυτή παρουσιάζονται τα βασικότερα στοιχεία για την κωδικοποίηση ενός κειμένου. Τα στοιχεία αυτά περικλείονται κυρίως στο στοιχείο **body**. Τα στοιχεία αυτά ομαδοποιούνται στις ακόλουθες κατηγορίες:

- Δομικά στοιχεία: **div, head**
- Στοιχεία παραγράφων **p, cit, q, l, lg, sp**
- Λίστες, πίνακες και σχήματα: **list, item, table, row, cell, figure, figDesc**
- Στοιχεία φράσεων: **date, emph, foreign, hi, name, num, soCalled, term, title**
- Στοιχεία χωρισμού σελίδων και γραμμών: **milestone, pb, lb**
- Στοιχεία για σύνδεση στοιχείων: **ref, rs, ptr, xref, xptr**
- Βιβλιογραφικά στοιχεία: **bibl, author, editor, publisher, respStmt, resp, pubPlace**

Σε γενικές γραμμές οι κανόνες χρήσης των παραπάνω στοιχείων είναι:

- Το στοιχείο **body** ενός κειμένου περιλαμβάνει στοιχεία **div** (υποδιαίρεσεις, π.χ. μέρη, κεφάλαια, τμήματα κ.λπ.).
- Τα στοιχεία **div** περιλαμβάνουν στοιχεία **p** (παράγραφος), **q** (εδάφιο με εισαγωγικά), **l** (γραμμή), **lg** (ομάδα γραμμών), **sp** (λόγος) και **speaker** (ομιλητής).
- Το στοιχείο **p** και τα άλλα στοιχεία παραγράφων περιλαμβάνουν κείμενο (**#PCDATA**), το οποίο κωδικοποιείται από στοιχεία φράσεων και χωρισμού γραμμών και σελίδων (**data**, **name**, **pb**, κ.λπ.).

Στις επόμενες παραγράφους δίνεται μια επιλεκτική και συνοπτική περιγραφή των παραπάνω στοιχείων.

7.4.1. Δομικά στοιχεία

div (υποδιαίρεσεις): το πρότυπο προβλέπει υποδιαίρεσεις μέχρι 7 επίπεδα. Σε περίπτωση που κάποιο κείμενο διαθέτει περισσότερα επίπεδα υποδιαίρεσεων πρέπει να γίνει τροποποίηση του TEI-DTD. Το στοιχείο αυτό χρησιμοποιείται και στα στοιχεία **front** και **back**. Τα σημαντικότερα γνωρίσματα αυτού του στοιχείου είναι:

- **type** με τιμές: 'Book', 'Chapter', 'Part', κ.λπ.
- **id**: μοναδικός κωδικός υποδιαίρεσης
- **n**: όνομα ή αριθμός υποδιαίρεσης

Βασικά υποστοιχεία μιας υποδιαίρεσης είναι το **head** που κωδικοποιεί τον τίτλο της υποδιαίρεσης και το **p** που δηλώνει την ύπαρξη παραγράφου.

Παράδειγμα 7.2.

```
<div1 type="Act" n="I">
  <head>ACT I</head>
  <div2 type="Scene" n="1">
    <head>SCENE I</head>
    <p>I fully appreciate Gen. Pope's splendid... </p> ...
  </div2>
</div1>
```

Το TEI σκοπεύοντας να καλύψει όλα τα είδη του γραπτού λόγου παρέχει τα στοιχεία **speaker** και **stage** για τα πεζά και θεατρικά κείμενα και τα στοιχεία **l** και **lg** για τον έμμετρο λόγο, ενώ το στοιχείο **sp** χρησιμοποιείται τόσο σε πεζά, όσο και έμμετρα κείμενα και δηλώνει ένα τμήμα ομιλίας. Βασικό γνώρισμα αυτού του στοιχείου είναι το **who** με το οποίο προσδιορίζεται ο ομιλητής.

Στον πεζό και θεατρικό λόγο ο ομιλητής που έχει το λόγο καθορίζεται από το στοιχείο **speaker**, ενώ με το στοιχείο **stage** δηλώνονται σκηνοθετικές κατευθύνσεις, που προσδιορίζονται από το γνώρισμα **type**, το οποίο έχει τιμές **entrance, exit, setting, delivery**

Στον έμμετρο λόγο το στοιχείο **lg** δηλώνει ότι ακολουθεί μια ομάδα στίχων, ενώ το στοιχείο **l** περιλαμβάνει ως δεδομένα μια γραμμή (στίχο).

Παράδειγμα 7.3.

```
<stage rend="italic">
```

```
    Enter Barnardo and Francisco, two Sentinels, at several doors
```

```
</stage>
```

```
<sp>
```

```
    <speaker>Barn</speaker>
```

```
    <l>Who's there?</l>
```

```
</sp>
```

```
<sp>
```

```
    <speaker>Fran</speaker>
```

```
    <lg>
```

```
        <l>But why drives on that ship so fast</l>
```

```
        <l>Withouten wave or wind?</l>
```

```
    </lg>
```

```
</sp>
```

```
<sp>
```

```
    <speaker>Barn</speaker>
```

```
    <l>Long live the King!</l>
```

```
</sp>
```

```
<sp>
```

```
    <speaker>Fran</speaker>
```

```
    <l>Barnardo?</l>
```

```
</sp>
```

```
<sp>
```

```
    <speaker>Barn</speaker>
```

```
    <l>He.</l>
```

```
</sp>
```

Τέλος τα ακόλουθα στοιχεία δηλώνουν την ύπαρξη λίστας:

list: δήλωση λίστας,

item: στοιχείο λίστας

Παράδειγμα 7.4.

<p> I went to the store to buy<list><item>bread,</item> <item>milk,</item>
<item>and bananas</item></p>

7.4.2. Στοιχεία Φράσεων

Τα ακόλουθα στοιχεία χρησιμοποιούνται για την κωδικοποίηση φράσεων:

emph: έμφαση φράσης για γλωσσικό ή ρητορικό σκοπό

foreign: φράση ή λέξη που ανήκει σε άλλη γλώσσα από το τριγύρω κείμενο

term: τεχνικός όρος

title: τίτλος με κυριότερα γνωρίσματα:

- **level** με τιμές: **m** βιβλία, συλλογές, έργα ενός τόμου ή πολύτομα, **s** σειρές, **j** περιοδικό, **u** μη δημοσιευμένο υλικό, **a** αναλυτικός τίτλος που ανήκει σε κάποιο άλλο τεκμήριο (άρθρο, ποίημα κ.λπ.)
- **type** με τιμές: **abbreviated**, **main**, **subordinate** (υπότιτλοι και τίτλοι μερών), **parallel** (παράλληλοι).

Τα στοιχεία **name**, **date**, **time**, **num** διαθέτουν τα γνωρίσματα **type**, **date value**, **time value** και **value** που επεξηγούν τον τύπο και τις τιμές των περιεχομένων τους.

Παράδειγμα 7.5.

<name type="person"> Walter de la Mare</name> was born at <name type="place">Charlton</name>, in <name type="county">Kent</name>, in <date value="1873-02-21">21 Feb 1980</date>
<l>especially when it's nine below zero</l> <l>and <time value="15:00">three o'clock in the afternoon</time></l>
<num value="33">xxxiii</num> <num type="cardinal" value="21">twenty-one</num> <num type="percentage" value="10">ten percent</num> <num type="percentage" value="10">10%</num> <num type="ordinal" value="5">5th</num>

7.4.3. Γραμμές και σελίδες

Το στοιχείο **lb** δηλώνει την αλλαγή γραμμής και έχει γνώρισμα τον αριθμό της γραμμής **n**

Παράδειγμα 7.6.

```
<p><lb n="25"/> Fie, that you'll say so! He plays o' th'
  <lb n="26"/> viol-de-gamboys, and speaks three or four languages
  <lb n="27"/> word for word without book, and hath all the good
  <lb n="28"/> gifts of nature.</p>
```

Το στοιχείο **pb** δηλώνει το σημείο αλλαγής σελίδας και έχει γνωρίσματα τον αριθμό της σελίδας **n** και την έκδοση **ed** στην οποία απαντάται η συγκεκριμένη σελιδοποίηση. Στο ακόλουθο παράδειγμα υπάρχουν δύο σελιδοποιήσεις του ίδιου κειμένου από ισάριθμες εκδόσεις **ED1** και **ED2**.

Παράδειγμα 7.7.

```
<p>I wrote to Moor House and to Cambridge immediately, to say what I had
done: fully explaining also why I had thus acted. Diana and <pb ed="ED1"
n="475"/> Mary approved the step unreservedly. Diana announced that she
would <pb ed="ED2" n="485"/>just give me time to get over the honeymoon,
and then she would come and see me. </p>
```

7.4.4. Αναφορές και σημειώσεις

Για αναφορές από ένα σημείο σε άλλο σημείο του ίδιου κειμένου, χρησιμοποιούνται τα στοιχεία **ptr** και **ref**. Η βασική διαφορά τους είναι ότι το **ptr** είναι ένα κενό στοιχείο, ενώ το **ref** χρησιμοποιείται όταν η αναφορά συνοδεύεται από άλλο κείμενο.

Βασικό γνώρισμα των παραπάνω στοιχείων είναι το **target** το οποίο χρησιμεύει για τη δήλωση του σημείου του κειμένου στο οποίο γίνεται η αναφορά. Το σημείο αυτό είναι η τιμή του γνωρίσματος **id** μιας της υποδιαίρεσης του κειμένου που ορίζεται από στοιχείο **div**.

Παράδειγμα 7.8.

See especially <ref target="SEC12">section 12 on page 34</ref>.

See especially <ptr target="SEC12"/>.

Απαραίτητη προϋπόθεση για την εγκυρότητα της αναφοράς είναι η ύπαρξη της δήλωσης:

```
<div id=" SEC12">
  <head> section 12</head>
</div1>
```

Για αναφορές από ένα κείμενο σε κάποιο άλλο κείμενο χρησιμοποιούνται τα στοιχεία:

- **xprt** το οποίο δηλώνει ένα δείκτη σε ένα σημείο του ίδιου ή άλλου κειμένου

- **xref** το οποίο δηλώνει ένα δείκτη σε ένα σημείο του ίδιου ή άλλου κειμένου που πιθανώς είτε έχει τροποποιηθεί (το κείμενο) ή ακολουθούν πρόσθετα σχόλια.

Σημαντικά γνωρίσματα αυτών των στοιχείων είναι τα:

- **doc** που προσδιορίζει το κείμενο στο οποίο γίνεται η αναφορά
- **from** που προσδιορίζει το σημείο που αρχίζει η αναφορά στο κείμενο που ορίζεται στο γνώρισμα **doc**.
- **to** που προσδιορίζει το πέρας της αναφοράς. Χρησιμοποιείται αν και μόνο αν υπάρχει δήλωση γνωρίσματος **from**.

Τιμές των γνωρισμάτων **doc**, **form** και **to** είναι οι τιμές των **id** γνωρισμάτων των υποδιαίρέσεων (**div**) στις οποίες γίνεται η αναφορά.

Παράδειγμα 7.9.

For full details, see `<xptr doc="P1" from="id (xyz)" to="id (abc)"/>`

Οι σημειώσεις που υπάρχουν σε ένα κείμενο περικλείονται στις ετικέτες του στοιχείου **note**. Βασικά γνωρίσματα του στοιχείου αυτού είναι τα:

- **resp** που καθορίζει αν η σημείωση είναι του εκδότη, ή του συγγραφέα, ή του μεταφραστή κ.λπ, παρέχοντας αντίστοιχα τις τιμές **editorial**, **author**, **translator**.
- **place** που καθορίζει το σημείο της σημείωσης, αν βρίσκεται μέσα στο στίχο μαζί με το κείμενο, μεταξύ στίχων, στο αριστερό ή δεξιό περιθώριο, στο υποσέλιδο ή στο τέλος του τόμου ή του κεφαλαίου ,με αντίστοιχες τιμές **inline**, **interlinear**, **left**, **right**, **foot** και **end**.

Παράδειγμα 7.10.

`<p>IN WITNESS WHEREOF, I have hereunto set my hand and caused the seal of the United States to be affixed.</p>`

`<note resp='ed' place=inline><p>The 72d Congress, which convened following the 1932 election, passed the Twenty-first Amendment to the Constitution to repeal the Eighteenth Amendment.</p> </note>`

Μια μορφή σημειώσεων είναι και οι σκόπιμες παραλήψεις εικόνων ή πινάκων ή άλλες εκδοτικές παρεμβάσεις που γίνονται σε μέρη του κωδικοποιημένου κειμένου. Οι παραλείψεις αυτές κωδικοποιούνται με τη χρήση του στοιχείου **gap** με βασικά γνωρίσματα τα:

- **desc**, που καθορίζει το τι έχει παραληφθεί **reason** που καθορίζει το λόγο της παράληψης **extent** που καθορίζει την έκταση ή το μέγεθος του αντικειμένου που παραλήφθηκε.

Παράδειγμα 7.11.

<p>They are, we will suppose, a Merchant and a Physician, or in other words, an Equilateral Triangle and a Pentagon: how am I to distinguish them?</p>
<p><gap desc='geometric figure' reason='editorial policy' extent='ca. 14 lines'></p>
<p>It will be obvious ... </p>

7.4.5. Βιβλιογραφικές Αναφορές

Οι βιβλιογραφικές αναφορές αποτελούν ένα από τα βασικά τμήματα ενός κειμένου. Μια βιβλιογραφική αναφορά περικλείεται σε ένα στοιχείο **bibl** που δηλώνει μια «χαλαρά» δομημένη βιβλιογραφική αναφορά ή σε ένα στοιχείο **biblFull** το οποίο δηλώνει μια «πλήρη» βιβλιογραφική αναφορά και έχει τη δομή του στοιχείου **fileDesc**. Το στοιχείο **bibl** περιλαμβάνει τα υποστοιχεία **title**, **author**, **publisher**, **date**, **editor**, **biblScope** κ.ά. Το στοιχείο **biblScope** δηλώνει συγκεκριμένο τμήμα της αναφοράς. Επίσης το στοιχείο **listBibl** χρησιμοποιείται όταν υπάρχει μια σειρά από βιβλιογραφικές αναφορές, και περικλείει περισσότερα του ενός στοιχεία **bibl**.

Σημαντικά γνωρίσματα του στοιχείου **title** είναι τα

- **type** που καθορίζει αν ο τίτλος είναι κύριος ή υπότιτλος, ή παράλληλος κ.λπ.
- **level** που καθορίζει το είδος του τεκμηρίου και παίρνει τις τιμές **A**, αν το τεκμήριο είναι άρθρο, **M** αν είναι βιβλίο ή μονογραφία, **J** αν είναι περιοδικό, **S** αν είναι σειρά και **U** αν είναι αδημοσίευτο υλικό, όπως π.χ. μια διατριβή.

Παράδειγμα 7.12.

<bibl>
<title level="M">The Riverside Shakespeare</title>
<author>Shakespeare, William</author>
<publisher>Boston: Houghton Mifflin, 1974</publisher>
<biblScope>5. 88ff</biblScope>
</bibl>

7.5. Εξώφυλλο, οπισθόφυλλο και παραρτήματα

Όπως προαναφέρθηκε το εξώφυλλο ενός κειμένου κωδικοποιείται από το στοιχείο **front**, ενώ το οπισθόφυλλο και τα παραρτήματά του κωδικοποιούνται από το στοιχείο **back**.

Το στοιχείο **front** περιλαμβάνει τουλάχιστον ένα στοιχείο **titlePage** το οποίο περιλαμβάνει τα υποστοιχεία **doctitle**, που υποδιαιρείται σε στοιχεία **titlePart**, και κωδικοποιούν τα μέρη του τίτλου χρησιμοποιώντας το γνώρισμα **type** με τιμές **main**, **subtitle** κ.λπ. Άλλα στοιχεία που περιέχονται στο **titlePage** είναι τα:

- **byline**, το οποίο περικλείει τον δημιουργό του έργου και περιλαμβάνει το υποστοιχείο **docAuthor**
- **docImprint** που κωδικοποιεί πληροφορίες για τον τόπο, την ημερομηνία έκδοσης και τον εκδότη του κειμένου.

Εκτός από το **titlePage** το στοιχείο **front** δομείται από υποδιαιρέσεις που κωδικοποιούνται με το στοιχείο **div**, και το γνώρισμά του **type** που δηλώνει το είδος της υποδιάρθρωσης με τιμές **foreword**, **preface**, **dedication**, **abstract**, **contents**, **ack** (acknowledgements, ευχαριστίες) και **frontispiece**.

Παράδειγμα 7.13.

```
<titlePage>
  <docTitle><titlePart type="main">
    Lives of the Queens of England, from the Norman
    Conquest;</titlePart>
    <titlePart type="sub">with anecdotes of their courts.
  </titlePart></docTitle>
  <titlePart>Now first published from Official Records
    and other authentic documents private as well as
    public.</titlePart>
  <docEdition>New edition, with corrections and
    additions</docEdition>
  <byline>By <docAuthor>Agnes Strickland</docAuthor></byline>
  <docImprint>Philadelphia: Blanchard and Lea</docImprint>
  <docDate>1860.</docDate>
</titlePage>
<div type="dedication">
  <head>To the Right Honourable <name>JOHN Lord Viscount
  BRACLY</name>, Son and Heir apparent to the Earl of
  Bridgewater, &c.</head>
  <p>THis <hi>Poem</hi>, which receiv'd its first occasion of
  Birth from your Self, and others of your Noble Family ....
  and as in this representation your attendant
  <name>Thyrsis</name>, so now in all reall expression
```

</div>

Το στοιχείο **back** δομείται από υποδιαίρεσεις που κωδικοποιούνται με το στοιχείο **div**, και το γνώρισμά του **type** που δηλώνει το είδος της υποδιαίρεσης με τιμές **appendix**, **glossary**, **notes**, **bibliography**, **index**, **colophon**.

7.6. Άλλα στοιχεία

Εκτός από τα προαναφερθέντα στοιχεία το TEI-DTD περιέχει πληθώρα άλλων στοιχείων τα οποία χρησιμεύουν για την κωδικοποίηση: χαρακτήρων που δεν ανήκουν στον ASCII κώδικα (π.χ. τονούμενων λατινικών χαρακτήρων), πινάκων και σχημάτων, εκδοτικών παραμβάσεων (διορθώσεις, παρεμβάσεις σχολιαστών κ.λπ.), υποσημειώσεων και άλλων χαρακτηριστικών ενός κειμένου. Μια σύντομη περιγραφή όλων των στοιχείων του TEI Lite DTD δίνεται στη σελίδα <http://www.tei-c.org/Lite/U5-taglist.html>.

Ολοκληρώνοντας την σύντομη παρουσίαση του TEI-DTD, πρέπει να αναφερθεί ότι γίνονται προσπάθειες βελτίωσης του προτύπου. Συγκεκριμένα οι προσπάθειες στρέφονται στα αδύνατά του σημεία που είναι η έλλειψη ενός πλήρους και λεπτομερούς συνόλου στοιχείων για την κωδικοποίηση πινάκων, μαθηματικών συμβόλων και σχέσεων και πολυμέσων. Επίσης απαιτούνται νέα στοιχεία για την κωδικοποίηση εξωφύλλων σύγχρονων εκδόσεων.

7.7. Ένα ολοκληρωμένο παράδειγμα

Το παράδειγμα που ακολουθεί είναι ένα απόσπασμα από το θεατρικό έργο «Άμλετ» του Σαίξπηρ.

Document type declaration	<?xml version="1.0"?> <!DOCTYPE TEI.2 SYSTEM "teixbaby.dtd">
The root element	<TEI.2>
The header	<teiHeader> <fileDesc> <titleStmnt> <title>Hamlet, Prince of Denmark: an electronic edition</title> <author>Shakespeare, William</author> </titleStmnt>

	<pre> <publicationStmnt> <publisher>Houghton Mifflin</publisher> <pubPlace>Boston MA</pubPlace> <date>1997</date> </publicationStmnt> <sourceDesc> <bibl> <title>The Riverside Shakespeare</title> <author>Shakespeare, William</author> <publisher>Boston: Houghton Mifflin, 1974</publisher> </bibl> </sourceDesc> </fileDesc> </teiHeader> </pre>
Text	<pre> <text> </pre>
Front	<pre> <front><div type="castlist"> <list><item id="Oph">OPHELIA, daughter to Polonius</item> <item id="King">CLAUDIUS, King of Denmark</item> <item id="Queen">GERTRUDE, Queen of Denmark</item> </list> </div></front> </pre>
Body with divs	<pre> <body><div type="act" n="4"><div n="4.5" type="scene"> </pre>
Queen speaks	<pre> <stage><hi rend="i">Enter</hi>KING.</stage> <sp who="Queen"><speaker> <hi rend="i">Queen.</hi></speaker> <l n="37" part="Y"> Alas, look here, my lord.</l> </sp> </pre>
Ophelia speaks	<pre> <sp who="Oph"><speaker> <hi rend="i">Oph.</hi></speaker><stage> <hi rend="i">Song.</hi></stage> </pre>

	<pre><lg part="M" type="song"><l n="38"> "Larded all with sweet flowers,</l> <l n="39"> Which bewept to the ground did not go</l> <l n="40"> With true-love showers."</l></lg> </sp></pre>
King speaks	<pre><sp who="King"><speaker> <hi rend="i">King.</hi></speaker> <l n="41" part="Y"> How do you, pretty lady?</l> </sp></pre>
Ophelia speaks	<pre><sp who="Oph"><speaker> <hi rend="i">Oph.</hi></speaker> <p><lb n="42"/> Well, <rs key="God">God</rs> dild you! They say the owl was a <lb n="43"/> baker's daughter. Lord, we know what we are, but <lb n="44"/> know not what we may be. <rs key="God">God</rs> be at your table!</p> </sp></pre>
King speaks	<pre><sp who="King"><speaker> <hi rend="i">King.</hi></speaker> <l n="45" part="Y"> Conceit upon her father. </l> </sp></pre>
Ophelia speaks	<pre><sp who="Oph"><speaker><hi rend="i">Oph.</hi></speaker> <p><lb n="46"/> Pray let's have no words of this, but when <lb n="47"/> they ask you what it means, say you this:</p> <stage> <hi rend="i">Song.</hi></stage> <lg part="M" type="song"> <l n="48">"To-morrow is <rs key="StValentine">Saint Valentine's</rs> day,</l> <l n="49">All in the morning betime,</l> <l n="50"> And I a maid at your window, </l> <l n="51"> To be your <rs</pre>

	<pre> key="StValentine">Valentine</rs>.</l> <l n="52"> "Then up he rose and donn'd his clo'es,</l> <l n="53"> And dupp'd the chamber-door,</l> <l n="54"> Let in the maid, that out a maid</l> <l n="55"> Never departed more."</l> </lg> </sp> </pre>
Body with divs	<pre></div> </div></body></pre>
Text	<pre></text></pre>
Root element	<pre></TEI.2></pre>

ΚΕΦΑΛΑΙΟ 8: MARC XML

8.1. Εισαγωγή

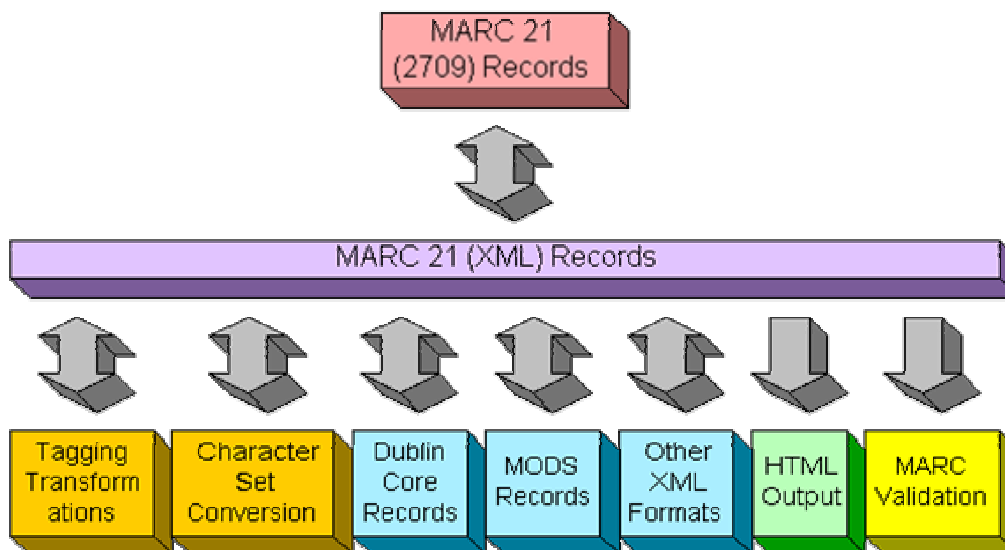
Η «προσαρμογή» του προτύπου MARC (ISO 2709) στο περιβάλλον της XML, η οποία γίνεται από τη Βιβλιοθήκη του Κογκρέσου, στοχεύει στο να δώσει την ευελιξία στους χρήστες του προτύπου να το χρησιμοποιούν με διαφορετικούς τρόπους, έτσι ώστε να καλύπτονται οι ανάγκες τους. Συγκεκριμένα, στόχος της όλης προσπάθειας είναι να δοθεί η δυνατότητα μετασχηματισμού MARC εγγραφών σε μορφές άλλων προτύπων μεταδεδομένων όπως τα Dublin Core, MODS κ.λπ., χωρίς απώλεια πληροφορίας.

Το πλαίσιο MARC XML αποτελείται από πολλές ενότητες που περιλαμβάνουν σχήματα, εργαλεία λογισμικού και stylesheets. Ο πυρήνας αυτού του πλαισίου είναι ένα απλό XML schema που περιγράφει τη *δομή* των MARC εγγραφών (δηλ. του προτύπου ISO2709 και όχι τη σημασιολογία τους με την οποία ασχολούνται τα πρότυπα MARC21, UNIMARC κ.λπ.) και αποτελεί το μέσο για το μετασχηματισμό τους (βλ. παρακάτω σχήμα) σε άλλα πρότυπα. Σε αυτό το XML schema όλα τα αρχικά πεδία (leader fields) και τα πεδία ελέγχου (control fields) κάθε MARC εγγραφής, αντιστοιχούν σε σειρές δεδομένων (data strings). Τα υπόλοιπα πεδία αντιστοιχούν σε στοιχεία (elements), ενώ τα σύμβολά τους (tag, π.χ. πεδίο 245, 350 κ.λπ.) και οι δείκτες (indicators) αντιστοιχούν σε γνωρίσματα (attributes). Τα υπο-πεδία αντιστοιχούν σε υπο-στοιχεία και τα χαρακτηριστικά τους και η στίξη τους σε γνωρίσματα. Τα δομικά στοιχεία του MARC, όπως το μήκος ενός πεδίου και η θέση που αρχίζουν τα περιεχόμενά του δεν είναι απαραίτητα στην XML schema εγγραφή.

Με αυτό τον τρόπο όλα τα στοιχεία δεδομένων μιας MARC εγγραφής μετατρέπονται σε XML και αντιστρόφως, χωρίς καμία απώλεια πληροφορίας. Η μετατροπή των MARC δεδομένων γίνεται από προγράμματα μετατροπής σε XML, ενώ για πολύπλοκους μετασχηματισμούς, καθώς επίσης και για τον έλεγχο επαλήθευσης (εγκυρότητας) της μετατροπής χρησιμοποιούνται ειδικά εργαλεία λογισμικού που μπορούν και διαβάζουν το MARC XML schema.

Μετά τα μέσα της δεκαετίας του 1990, η Βιβλιοθήκη του Κογκρέσου ανέπτυξε δύο XML DTDs για την μετατροπή της MARC21 δομής σε XML και αντιστρόφως, χωρίς απώλεια πληροφοριών. Αυτά τα DTDs διαφέρουν από το αντίστοιχο XML schema γιατί ορίζουν για κάθε MARC στοιχείο δεδομένων, ένα XML στοιχείο (XML element). Το κύριο μειονέκτημα αυτής της προσέγγισης

είναι ότι παράγει πολύ μεγάλα DTDs. Για αυτό το σκοπό σε αυτές τις σημειώσεις εξετάζεται μόνο η περίπτωση του XML schema.



8.2. Χαρακτηριστικά, πλεονεκτήματα - μειονεκτήματα

Η χρήση του MARC XML δίνει τις ακόλουθες δυνατότητες:

1. αναπαράστασης της δομής των MARC εγγραφών σε XML
2. περιγραφής πληροφοριακών πόρων από τη γλώσσα XML
3. συλλογής (harvesting) μεταδεδομένων από το πρωτόκολλο Open Archives Initiative (OAI)
4. επέκτασης και συμβατότητας των MARC εγγραφών στο πρότυπο METS (Metadata Encoding and Transmission Standard)

Τα σημαντικότερα πλεονεκτήματα από τη χρήση του MARC XML είναι ότι

- υποστηρίζει ανεξαρτήτως μορφής (format) όλα τα δεδομένα που έχουν κωδικοποιηθεί από το πρότυπο MARC
- αποτελεί μια αρχιτεκτονική δεδομένων που επιτρέπει την ανάπτυξη διάφορων εφαρμογών λογισμικού που καλύπτουν συγκεκριμένες ανάγκες και υπηρεσίες.

Ωστόσο, το σημαντικότερο μειονέκτημα αυτού του πλαισίου είναι ότι για τον έλεγχο επαλήθευσης (εγκυρότητας) της μετατροπής MARC εγγραφών, χρησιμοποιείται λογισμικό που δεν είναι ενσωματωμένο στο πλαίσιο (δηλ. ειδικά «εξωτερικά» εργαλεία λογισμικού).

Κώδικας	<pre><xsd:annotation> <xsd:documentation>collection is a top level container element for 0 or many records</xsd:documentation> </xsd:annotation> </xsd:element></pre>
---------	---

Ειδικότερα ο σύνθετος τύπος **collectionType** ορίζεται ως εξής:

complexType collectionType

Γνωρίσματα	Name id	Type IdDataType	Use optional	Default	Fixed	Annotation
XML Κώδικας	<pre><xsd:complexType name="collectionType" id="collection.ct"> <xsd:sequence minOccurs="0" maxOccurs="unbounded"> <xsd:element ref="record"/> </xsd:sequence> <xsd:attribute name="id" type="idDataType" use="optional"/> </xsd:complexType></pre>					

Το στοιχείο **record** περιγράφει τη δομή μιας MARC εγγραφής με τα πεδία της της. Μια MARC εγγραφή αποτελείται από ένα αρχικό πεδίο (**leader field**, οι πρώτοι 24 χαρακτήρες), κανένα ή περισσότερα πεδία ελέγχου (**control fields** που αντιστοιχούν στα MARC21 πεδία 001-009) και τέλος κανένα ή περισσότερα πεδία δεδομένων (**data fields** που αντιστοιχούν στα MARC21 πεδία 010-999). Ο ακόλουθος πίνακας περιγράφει τη δομή του στοιχείου **record**.

element record

Τύπος	recordType					
Περιλαμβάνει	leader controlfield datafield					
Χρησιμοποιείται Από	complexType collectionType					
Γνωρίσματα	Name type id	Type recordType idDataType	Use optional optional	Default	Fixed	Annotation
XML Κώδικας	<pre><xsd:element name="record" type="recordType" nillable="true" id="record.e"> <xsd:annotation> <xsd:documentation>record is a top level container element for all of the field elements which compose the record</xsd:documentation> </xsd:annotation> </xsd:element></pre>					

Ο ακόλουθος πίνακας περιγράφει αναλυτικά τον σύνθετο τύπο **recordType**. Στον XML κώδικα παρουσιάζεται η δήλωση των εναλλακτικών πεδίων που περιλαμβάνει μια MARC εγγραφή (leader, control ή data) ως στοιχεία.

complexType recordType

Γνωρίσματα	Name Type	Type recordTypeType	Use optional	Default	Fixed	Annotation
	id	idDataType	optional			
XML Κώδικας	<pre> <xsd:complexType name="recordType" id="record.ct"> <xsd:sequence minOccurs="0"> <xsd:element name="leader" type="leaderFieldType"/> <xsd:element name="controlfield" type="controlFieldType" minOccurs="0" maxOccurs="unbounded"/> <xsd:element name="datafield" type="dataFieldType" minOccurs="0" maxOccurs="unbounded"/> </xsd:sequence> <xsd:attribute name="type" type="recordTypeType" use="optional"/> <xsd:attribute name="id" type="idDataType" use="optional"/> </xsd:complexType> </pre>					

Οι MARC εγγραφές χωρίζονται στις ακόλουθες κατηγορίες: Bibliographic, Authority, Holdings, Classification και Community. Στο MARC XML schema οι κατηγορίες αυτές περιγράφονται από τον απλό τύπο **recordType** ως εξής:

simpleType recordType

XML Κώδικας	<pre> <xsd:simpleType name="recordTypeType" id="type.st"> <xsd:restriction base="xsd:string"> <xsd:enumeration value="Bibliographic"/> <xsd:enumeration value="Authority"/> <xsd:enumeration value="Holdings"/> <xsd:enumeration value="Classification"/> <xsd:enumeration value="Community"/> </xsd:restriction> </xsd:simpleType> </pre>
----------------	--

8.4. Πεδία και υπο-πεδία

Όπως προαναφέρθηκε, κάθε εγγραφή αποτελείται ένα αρχικό πεδίο (**leader field**), κανένα ή περισσότερα πεδία ελέγχου (**control fields** που αντιστοιχούν στα MARC πεδία 001-009) και τέλος κανένα ή περισσότερα πεδία δεδομένων (**data fields** που αντιστοιχούν στα MARC πεδία 010-999).

Το **leader field** είναι σταθερού μήκους και επομένως το γνώρισμα **xml:space** έχει την τιμή **preserve**. Συγκεκριμένα το πεδίο αυτό αποτελείται από τους πρώτους 24 χαρακτήρες ενός MARC record και περιλαμβάνει πληροφορίες για το μήκος της εγγραφής και άλλες πληροφορίες του συστήματος. Ο ακόλουθος πίνακας περιγράφει τον σύνθετο τύπο **leaderFieldType**

complexType leaderFieldType

Τύπος	extension of leaderDataType
Χρησιμοποιείται	element recordType/leader

από						
Γνωρίσματα	Name id	Type idDataType	Use optional	Default	Fixed	Annotation
	xml:space	xsd:NCName			preserve	
XML Κώδικας	<pre> <xsd:complexType name="leaderFieldType" id="leader.ct"> <xsd:annotation> <xsd:documentation>MARC21 Leader, 24 bytes</xsd:documentation> </xsd:annotation> <xsd:simpleContent> <xsd:extension base="leaderDataType"> <xsd:attribute name="id" type="idDataType" use="optional"/> <xsd:attribute ref="xml:space" fixed="preserve"/> </xsd:extension> </xsd:simpleContent> </xsd:complexType> </pre>					

Τα πεδία ελέγχου διαθέτουν απαραίτητως ετικέτα (tag από 001-009) και είναι σταθερού μήκους. Αυτά τα δύο χαρακτηριστικά ορίζονται ως γνωρίσματα στην παρακάτω περιγραφή του σύνθετου τύπου **controlFieldType**.

complexType controlFieldType

Τύπος	extension of controlDataType					
Χρησιμοποιείται Από	element recordType/controlfield					
Γνωρίσματα	Name Id	Type idDataType	Use optional	Default	Fixed	Annotation
	tag	controltagDataType	required			
	xml:space	Xsd:NCName			preserve	
XML Κώδικας	<pre> <xsd:complexType name="controlFieldType" id="controlfield.ct"> <xsd:annotation> <xsd:documentation>MARC21 Fields 001-009</xsd:documentation> </xsd:annotation> <xsd:simpleContent> <xsd:extension base="controlDataType"> <xsd:attribute name="id" type="idDataType" use="optional"/> <xsd:attribute name="tag" type="controltagDataType" use="required"/> <xsd:attribute ref="xml:space" fixed="preserve"/> </xsd:extension> </xsd:simpleContent> </xsd:complexType> </pre>					

Οι ετικέτες των πεδίων δεδομένων ανήκουν στον απλό τύπο **controltagDataType**, που περιγράφεται από τον ακόλουθο πίνακα.

simpleType controltagDataType

Τύπος	restriction of xsd:string
Χρησιμοποιείται	attribute controlFieldType/@tag

Από	
XML Κώδικας	<pre><xsd:simpleType name="controltagDataType" id="controltag.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> <xsd:pattern value="00[1-9A-Za-z]{1}" /> </xsd:restriction> </xsd:simpleType></pre>

Τα πεδία δεδομένων (data fields, έχουν ετικέτες από 010-999), διαθέτουν απαραίτητως ετικέτα (tag) και δείκτες (indicators) - τα οποία δηλώνονται ως γνωρίσματα - και ένα ή περισσότερα υποπεδία (subfields), τα οποία δηλώνονται ως στοιχεία. Ο ακόλουθος πίνακας περιγράφει το σύνθετο τύπο **dataFieldType**.

complexType dataFieldType

Περιλαμβάνει	subfield					
Χρησιμοποιείται από	element recordType/datafield					
Γνωρίσματα	Name Id	Type idDataType	Use optional	Default	Fixed	Annotation
	tag	tagDataType	required			
	ind1	indicatorDataType	required			
	ind2	indicatorDataType	required			
XML Κώδικας	<pre><xsd:complexType name="dataFieldType" id="datafield.ct"> <xsd:annotation> <xsd:documentation>MARC21 Variable Data Fields 010-999</xsd:documentation> </xsd:annotation> <xsd:sequence maxOccurs="unbounded"> <xsd:element name="subfield" type="subfielddataFieldType"/> </xsd:sequence> <xsd:attribute name="id" type="idDataType" use="optional"/> <xsd:attribute name="tag" type="tagDataType" use="required"/> <xsd:attribute name="ind1" type="indicatorDataType" use="required"/> <xsd:attribute name="ind2" type="indicatorDataType" use="required"/> </xsd:complexType></pre>					

Όπως και στην περίπτωση των ετικετών των πεδίων ελέγχου, έτσι και οι ετικέτες των πεδίων δεδομένων ανήκουν στον απλό τύπο **tagDataType** που περιγράφεται από τον ακόλουθο πίνακα.

simpleType tagDataType

Τύπος	restriction of xsd:string
Χρησιμοποιείται Από	attribute dataFieldType/@tag
XML Κώδικας	<pre><xsd:simpleType name="tagDataType" id="tag.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> </xsd:restriction> </xsd:simpleType></pre>

	<pre><xsd:pattern value="0([1-9A-Z] [1-9a-z]){1}[\dA-Z]{1})((([1-9A-Z] [1-9a-z]){1}([\dA-Z] [\da-z]){2})"/> </xsd:restriction> </xsd:simpleType></pre>
--	--

Οι δείκτες (indicators) των πεδίων δεδομένων ανήκουν στον απλό τύπο **indicatorDataType**, του οποίου η περιγραφή δίνεται από τον ακόλουθο πίνακα.

simpleType indicatorDataType

Τύπος	restriction of xsd:string
Χρησιμοποιείται Από	attributes dataFieldType/@ind1 dataFieldType/@ind2
XML Κώδικας	<pre><xsd:simpleType name="indicatorDataType" id="ind.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> <xsd:pattern value="[\dA-Za-z]{1}"/> </xsd:restriction> </xsd:simpleType></pre>

Τα υποπεδία των πεδίων δεδομένων ανήκουν στον σύνθετο τύπο **subfielddataType**, που περιγράφεται από τον ακόλουθο πίνακα.

complexType subfielddataType

Τύπος	extension of subfieldDataType					
Χρησιμοποιείται Από	element dataFieldType/subfield					
Γνωρίσματα	Name id code	Type idDataType subfieldcodeDataType	Use optional required	Default	Fixed	Annotation
XML Κώδικας	<pre><xsd:complexType name="subfielddataType" id="subfield.ct"> <xsd:simpleContent> <xsd:extension base="subfieldDataType"> <xsd:attribute name="id" type="idDataType" use="optional"/> <xsd:attribute name="code" type="subfieldcodeDataType" use="required"/> </xsd:extension> </xsd:simpleContent> </xsd:complexType></pre>					

Ο σύνθετος τύπος **subfielddataType** χρησιμοποιεί τον απλό τύπο **subfieldDataType** του οποίου η περιγραφή δίνεται από τον ακόλουθο πίνακα.

simpleType subfieldDataType

Τύπος	restriction of xsd:string
Χρησιμοποιείται	complexType subfielddataType

Από	
XML Κώδικας	<pre><xsd:simpleType name="subfieldDataType" id="subfield.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> </xsd:restriction> </xsd:simpleType></pre>

Για την ολοκλήρωση της παρουσίασης του MARC XML schema απομένει ο ορισμός του απλού τύπου **subfieldcodeDataType**, για τους κωδικούς των υποπεδίων των πεδίων δεδομένων.

simpleType subfieldcodeDataType

Τύπος	restriction of xsd:string
Χρησιμοποιείται Από	attribute subfielddataType/@code
XML Κώδικας	<pre><xsd:simpleType name="subfieldcodeDataType" id="code.st"> <xsd:restriction base="xsd:string"> <xsd:whiteSpace value="preserve"/> <xsd:pattern value="\dA-Za-z!#\$%&'()*+,-./:;<=>?@_~\[\]\{1}"/> </xsd:restriction> </xsd:simpleType></pre>

Τέλος, τα καθολικά γνωρίσματα του MARC XML schema δίνονται από την ακόλουθη περιγραφή.

attributeGroup xml:xmlGlobalAttributes

Γνωρίσματα	Name	Type	Use	Default	Fixed	Annotation
	X:lang	xsd:language				
	X:space	xsd:NCName		preserve		
	X:base	xsd:anyURI				
XML Κώδικας	<pre><xsd:attributeGroup name="xmlGlobalAttributes"> <xsd:attribute ref="x:lang"/> <xsd:attribute ref="x:space"/> <xsd:attribute ref="x:base"/> </xsd:attributeGroup></pre>					

8.5. Ένα ολοκληρωμένο Παράδειγμα

Η κωδικοποίηση της ακόλουθης MARC21 εγγραφής σε XML schema έχει ως εξής:

```
01142cam 2200301 a 4500 001 92005291
          003 DLC
          005 19930521155141.9
```


008 920219s1993 caua j 000 0 eng
 010 |a 92005291
 020 |a0152038655 :|c\$15.95
 040 |aDLC|cDLC|dDLC
 042 |a|cac
 050 00|aPS3537.A618|bA88 1993
 082 00|a811/.52|220
 100 1 |aSandburg, Carl,|d1878-1967.
 245 10|aArithmetic /|cCarl Sandburg ;
 illustrated as an anamorphic adventure by
 Ted Rand.
 250 |a1st ed.
 260 |aSan Diego :|bHarcourt Brace
 Jovanovich,|cc1993.
 300 |a1 v. (unpaged) :|bill. (some col.) ;|c26 cm.
 500 |aOne Mylar sheet included in pocket.
 520 |aA poem about numbers and their
 characteristics. Features anamorphic, or
 distorted, drawings which can be restored to
 normal by viewing from a particular angle or
 by viewing the image's reflection in the
 provided Mylar cone.
 650 0|aArithmetic|xJuvenile poetry.
 650 0|aChildren's poetry, American.
 650 1|aArithmetic|xPoetry.
 650 1|aAmerican poetry.
 650 1|aVisual perception.
 700 1 |aRand, Ted,|eill.

```

<?xml version="1.0" encoding="UTF-8" ?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>01142cam 2200301 a 4500</leader>
    <controlfield tag="001">92005291</controlfield>
    <controlfield tag="003">DLC</controlfield>
  
```

```
<controlfield tag="005">19930521155141.9</controlfield>
<controlfield tag="008">920219s1993 caua j 000 0
    eng</controlfield>
<datafield tag="010" ind1="" ind2="">
    <subfield code="a">92005291</subfield>
</datafield>
<datafield tag="020" ind1="" ind2="">
    <subfield code="a">0152038655 :</subfield>
    <subfield code="c">$15.95</subfield>
</datafield>
<datafield tag="040" ind1="" ind2="">
    <subfield code="a">DLC</subfield>
    <subfield code="c">DLC</subfield>
    <subfield code="d">DLC</subfield>
</datafield>
<datafield tag="042" ind1="" ind2="">
    <subfield code="a">lcac</subfield>
</datafield>
<datafield tag="050" ind1="0" ind2="0">
    <subfield code="a">PS3537.A618</subfield>
    <subfield code="b">A88 1993</subfield>
</datafield>
<datafield tag="082" ind1="0" ind2="0">
    <subfield code="a">811/.52</subfield>
    <subfield code="2">20</subfield>
</datafield>
<datafield tag="100" ind1="1" ind2="">
    <subfield code="a">Sandburg, Carl,</subfield>
    <subfield code="d">1878-1967.</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">Arithmetic /</subfield>
    <subfield code="c">Carl Sandburg ; illustrated as an
        anamorphic adventure by Ted Rand.</subfield>
</datafield>
<datafield tag="250" ind1="" ind2="">
    <subfield code="a">1st ed.</subfield>
</datafield>
```

```
<datafield tag="260" ind1="" ind2="">
  <subfield code="a">San Diego :</subfield>
  <subfield code="b">Harcourt Brace
    Jovanovich,</subfield>
  <subfield code="c">c1993.</subfield>
</datafield>
<datafield tag="300" ind1="" ind2="">
  <subfield code="a">1 v. (unpaged) :</subfield>
  <subfield code="b">ill. (some col.) ;</subfield>
  <subfield code="c">26 cm.</subfield>
</datafield>
<datafield tag="500" ind1="" ind2="">
  <subfield code="a">One Mylar sheet included in
    pocket.</subfield>
</datafield>
<datafield tag="520" ind1="" ind2="">
  <subfield code="a">A poem about numbers and their
    characteristics. Features anamorphic, or distorted,
    drawings which can be restored to normal by
    viewing from a particular angle or by viewing the
    image's reflection in the provided Mylar
    cone.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="0">
  <subfield code="a">Arithmetic</subfield>
  <subfield code="x">Juvenile poetry.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="0">
  <subfield code="a">Children's poetry,
    American.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="1">
  <subfield code="a">Arithmetic</subfield>
  <subfield code="x">Poetry.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="1">
  <subfield code="a">American poetry.</subfield>
</datafield>
```

```

<datafield tag="650" ind1="" ind2="1">
  <subfield code="a">Visual perception.</subfield>
</datafield>
<datafield tag="700" ind1="1" ind2="">
  <subfield code="a">Rand, Ted,</subfield>
  <subfield code="e">ill.</subfield>
</datafield>
</record>
</collection>

```

8.6. Άλλες υλοποιήσεις

Στις παραπάνω σελίδες περιγράφηκε το XML schema που αναπτύχθηκε από τη Βιβλιοθήκη του Κογκρέσου για την κωδικοποίηση και περιγραφή της δομής του προτύπου MARC (ISO2709). Όπως είναι γνωστό, η XML schema δίνει τη δυνατότητα περιγραφής της ίδιας δομής δεδομένων με πολλούς διαφορετικούς τρόπους. Αποτέλεσμα αυτής της εκφραστικής δύναμης που παρέχει η γλώσσα XML είναι η ανάπτυξη και άλλων σχημάτων για την περιγραφή του MARC προτύπου. Χαρακτηριστικό δείγμα αυτών των εναλλακτικών προσεγγίσεων αποτελεί το XML schema που αναπτύχθηκε από την πρωτοβουλία Open Archives Initiative (OAI). Πρόκειται για μια διαφορετική αντίληψη για την αναπαράσταση της δομής MARC εγγραφών που βασίζεται σε δηλώσεις πεδίων σταθερού και μεταβλητού μήκους (http://www.openarchives.org/OAI/oai_marc.xsd). Σκοπός αυτής της υλοποίησης είναι η δυνατότητα αξιοποίησης μεταδεδομένων σε MARC μορφή, από το πρωτόκολλο συλλογής μεταδεδομένων (harvesting) του OAI.

Όπως προαναφέρθηκε στην εισαγωγή, το πρότυπο MARC έχει υλοποιηθεί εκτός από XML schema και από XML DTD. Η σημαντικότερη διαφορά των XML DTD έναντι των XML schema είναι ότι τα DTD είναι μεν πιο χρήσιμα σε διαδικασίες ελέγχου εγκυρότητας των MARC εγγραφών γιατί εκτός από τη δομή απεικονίζουν και τη σημασιολογία των εγγραφών και των πεδίων, αλλά είναι λιγότερα εύχρηστα γιατί παράγουν πολύ μεγάλα XML τεκμήρια. Εκτός από την υλοποίηση του MARC DTD από τη Βιβλιοθήκη του Κογκρέσου, υπάρχουν δύο ακόμα σημαντικές πρωτοβουλίες οι οποίες στοχεύουν στην κωδικοποίηση MARC εγγραφών. Η πρώτη από αυτές ονομάζεται TVS (Transport, Validation and Services) (Carvalho, Cordeiro, 2002), προέρχεται από την Πορτογαλία και εστιάζει στη δομή του ISO 2709 προτύπου, ενώ η δεύτερη ονομάζεται BiblioML (<http://www.culture.fr/BiblioML/>), προέρχεται από το γαλλικό Υπουργείο Πολιτισμού και αποτελεί τη μοναδική προσπάθεια κωδικοποίησης του UNIMARC σε XML.

Βιβλιογραφία

1. <http://www.tei-c.org/Lite/> (ημ/νία πρόσβασης 27.3.2003).
2. <http://gutenberg.hwg.org/teidtds.html> (ημ/νία πρόσβασης 27.3.2003).
3. M. Mueller. A very gentle introduction to the TEI markup language. <http://bistro.northwestern.edu/AnaServer?tei+0+frame.anv> (ημ/νία πρόσβασης 27.3.2003).
4. <http://www.tei-c.org/Lite/DTD/teixlite.dtd> (ημ/νία πρόσβασης 27.3.2003).
5. <http://www.tei-c.org/Lite/U5-taglist.html> (ημ/νία πρόσβασης 27.3.2003).
6. <http://www.loc.gov/standards/marcxml/> (ημ/νία πρόσβασης 25.5.2003).
7. <http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd> (ημ/νία πρόσβασης 25.5.2003).
8. <http://www.dlib.vt.edu/projects/OAi/marcxml/marcxml.html> (ημ/νία πρόσβασης 25.5.2003).
9. http://www.openarchives.org/OAI/oai_marc.xsd (ημ/νία πρόσβασης 25.5.2003).
10. J. Carvalho, M. Ines Cordeiro "XML and bibliographic data: the TVS (Transport, Validation and Services) model", IFLA conference, Glasgow, 2002.
11. <http://www.culture.fr/BiblioML/> (ημ/νία πρόσβασης 25.5.2003).

Ευρετήριο όρων

address.....	90	head.....	91, 92
author.....	91, 97	hi 91	
authority.....	90	HTML.....	85
availability.....	90	id 92	
back.....	86, 87, 92, 97, 99	idno.....	90
bibl.....	91, 97	indicatorDataType.....	110
biblFull.....	97	ISO 2709.....	103
biblScope.....	97	item.....	91, 94
body.....	86, 87, 91, 92	l 91, 92, 93	
cell.....	91	lb 91, 94	
cit.....	91	leader.....	106, 107
collection.....	105	lg 91, 92, 93	
collectionType.....	105, 106	list.....	91, 93
control field.....	106, 107	listBibl.....	97
controlFieldType.....	108	MARC.....	103, 104, 105, 111
controltagDataType.....	108	MARC21.....	103
data.....	92	METS.....	104
data field.....	106, 107	milestone.....	91
dataFieldType.....	109	MODS.....	103
date.....	90, 91, 94, 97	n 92	
distributor.....	90	name.....	91, 92, 94
div.....	91, 92, 95, 96, 98, 99	note.....	96
DTD.....	85, 86, 87, 88, 92, 99, 103, 104	num.....	91, 94
Dublin Core.....	103	p 91, 92	
editor.....	91, 97	pb.....	91, 92, 95
emph.....	91, 94	profileDesc.....	86, 87
encodingDesc.....	86, 87	ptr.....	91, 95
field.....	106	publicationStmt.....	89, 90
figDesc.....	91	publisher.....	90, 91, 97
figure.....	91	pubPlace.....	90, 91
fileDesc.....	86, 87, 90	q 91, 92	
foreign.....	91, 94	record.....	105, 106
front.....	86, 87, 92, 97	recordType.....	105, 106
gap.....	96	ref.....	91, 95
group.....	87, 88	resp.....	91
		respStmt.....	91

revisionDesc	86, 87	time	94
row	91	title	91, 94, 97
rs91		titlePage	98
SGML	85	titlePart	98
soCalled	91	titleStmt	89
sourceDesc	90	type	92
sp	91, 92	UNIMARC	103
speaker	92, 93	XML	103
stage	93	XML schema	103, 105, 111
subfielddataType	110	xprt	95
subfieldcodeDataType	111	xptr	91
subfieldDataType	110	xref	91, 96
table	91	γνώρισμα	
tagDataType	109	attribute	87, 92, 94, 95, 96, 97, 103
teiHeader	86, 87, 91	δείκτες	
term	91, 94	indicators	110
text	86, 87	κωδικοποίηση κειμένου	85
Text encoding initiative		στοιχείο	85
TEI	86, 87, 88, 89, 92, 99	element	85, 86, 88, 89, 90, 91, 92, 94, 95, 96, 97
Text Encoding Initiative			
TEI	85, 86		