



# TEI Lite

Λίνα Μπουντούρη

[[boudouri@ionio.gr](mailto:boudouri@ionio.gr)]

Μάρτιος 2006

Ιόνιο Πανεπιστήμιο

Έργο «Αναμόρφωση του ΠΠΣ του Τμήματος  
Αρχειονομίας και Βιβλιοθηκονομίας». ΕΠΕΑΕΚ ΙΙ



Ανάπτυξη πανετού. Ανάπτυξη για όλους.

ΥΠΟΥΡΓΕΙΟ ΕΘΝΙΚΗΣ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΕΑΕΚ  
ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ  
ΣΥΓΧΡΗΜΑΤΟΔΟΤΗΣΗ  
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ  
ΕΥΡΩΠΑΪΚΟ ΤΑΜΕΙΟ ΠΕΡΙΦΕΡΕΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ



Η ΠΑΙΔΕΙΑ ΣΤΗΝ ΚΟΡΥΦΗ  
Επιχειρησιακό Πρόγραμμα  
Εκπαίδευσης και Αρχικής  
Επαγγελματικής Κατάρτισης

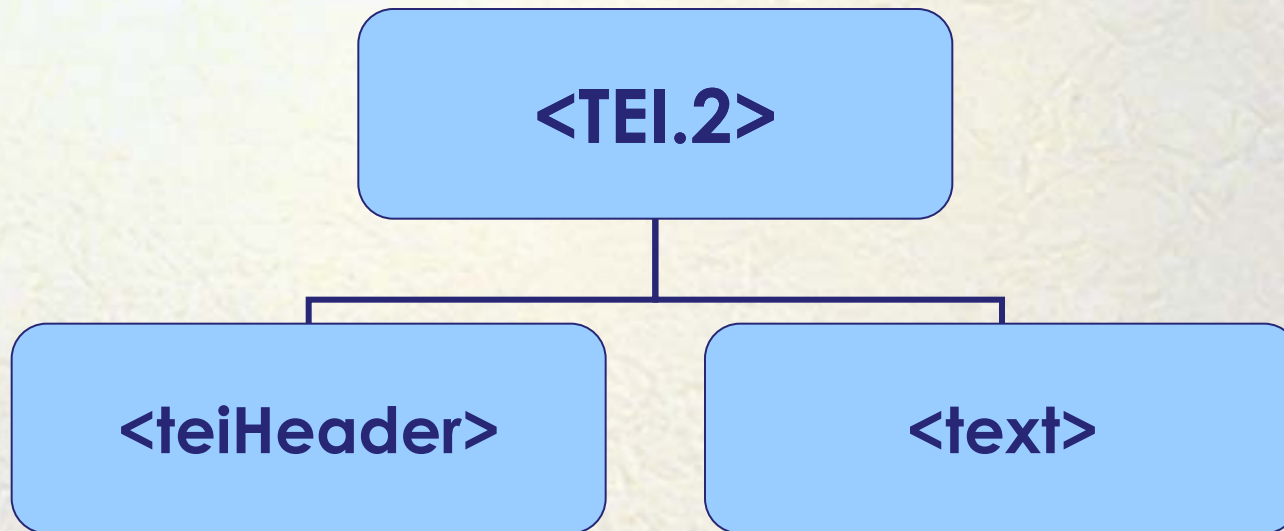
# TEI Lite: γενικά...

- TEI Lite: Σχήμα μεταδεδομένων
  - Κωδικοποίηση πληροφορίας σε μορφή κειμένου (έμφαση στις Ανθρωπιστικές Επιστήμες)
    - Μεταδεδομένα για το τεκμήριο (όπως τα MARC, EAD, MODS)
    - Το περιεχόμενο του τεκμηρίου (για παράδειγμα το κείμενο ενός βιβλίου)
- Υπό-σύνολο του TEI
- Ανάγκη για **μη** υπέρ-αναλυτική κωδικοποίηση
- TEI Lite DTD [<http://www.tei-c.org/Lite/DTD/>]
  - Συνοπτικό σε σχέση με το TEI
  - Περιλαμβάνει τα περισσότερα βασικά σύνολα στοιχείων του TEI
  - Διαχειρίζεται ποικίλα είδη κειμένων

# Πριν την κωδικοποίηση

- Θεωρητικό πλαίσιο
  - Τι θέλουμε να κωδικοποιήσουμε
    - Φύση, μορφή και ιδιαιτερότητες κειμένου
    - Ανάγκες χρηστών
- Οι κόμβοι (στοιχεία, γνωρίσματα) που επιλέγουμε για κωδικοποίηση πρέπει να καλύπτουν τις ανάγκες του κειμένου που θέλουμε να κωδικοποιήσουμε
  - Ανάγκη χάραξης πολιτικής κωδικοποίησης
- Έλεγχος των ιδιομορφιών του κειμένου. Για παράδειγμα
  - Δεν υπάρχει διάκριση ανάμεσα στα μονά εισαγωγικά και στην απόστροφο [μπορεί να είναι δύσκολο να διακρίνουμε τον ευθύ λόγο]

# Η δομή ενός TEI Lite εγγράφου



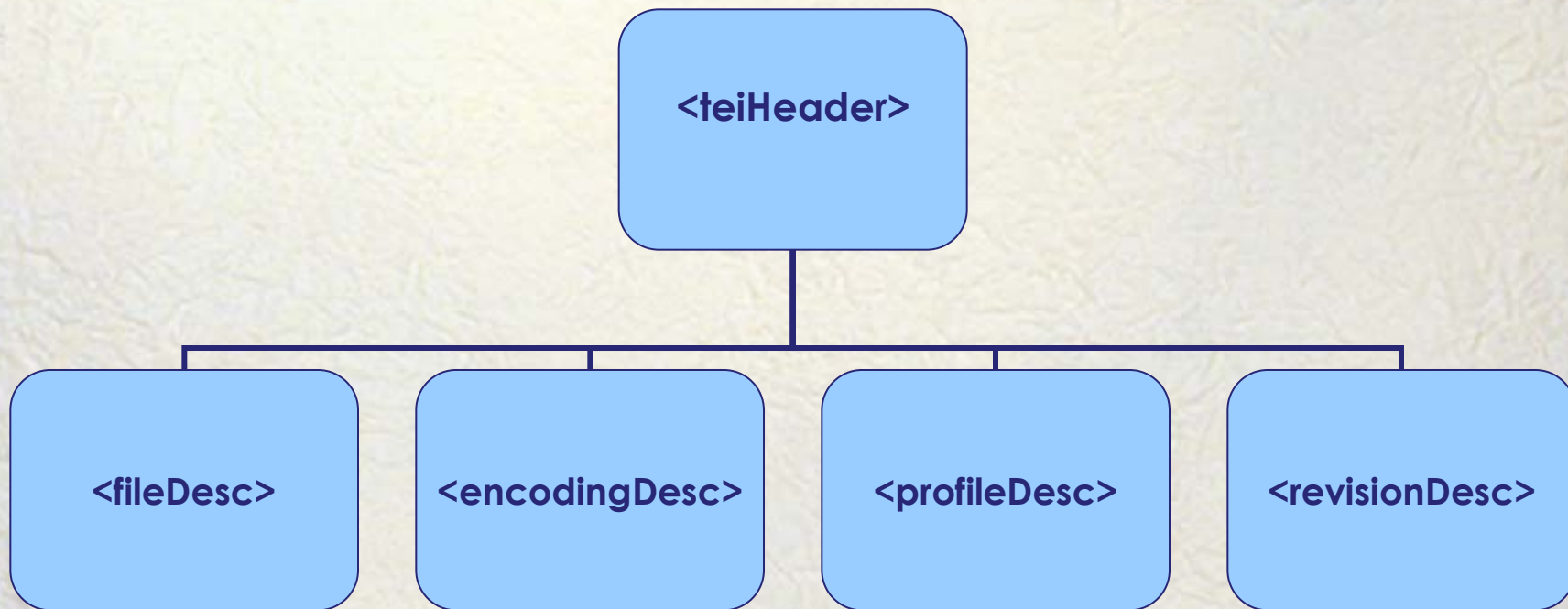
**<TEI.2>**

**<teiHeader> </teiHeader>**

**<text></text>**

**</TEI.2>**

# TEI header - <teiHeader>



# TEI header - <teiHeader>

- Μερικά στοιχεία του TEI header ομαδοποιούνται ως εξής:
  - Στοιχεία που το όνομά τους λήγει σε *Stmt* (statement) συνήθως περιέχουν ομάδες στοιχείων που κωδικοποιούν δομημένη πληροφορία
  - Στοιχεία που το όνομά τους λήγει σε *Decl* (declaration) περιέχουν πληροφορίες για συγκεκριμένες πρακτικές κωδικοποίησης
  - Στοιχεία που το όνομά τους λήγει σε *Desc* (description) περιέχουν περιγραφή σε πεζό λόγο

# TEI header - <teiHeader>

- Πληροφορίες ανάλογες με αυτές που παρέχονται στη σελίδα τίτλου ενός έντυπου κειμένου
  - **<fileDesc>**: Βιβλιογραφική περιγραφή του ηλεκτρονικού εγγράφου
  - **<encodingDesc>**: Τεκμηρίωση για τη σχέση ανάμεσα στο ηλεκτρονικό έγγραφο και την πηγή ή τις πηγές από τις οποίες προέρχεται
  - **<profileDesc>**: Περιγραφή των μη-βιβλιογραφικών «όψεων» του κειμένου, π.χ. οι μετέχοντες στη δημιουργία του ηλεκτρονικού εγγράφου
  - **<revisionDesc>**: Πληροφορίες για τυχόν αναθεωρήσεις στο ηλεκτρονικό έγγραφο, π.χ. αλλαγές στην κωδικοποίηση

# <fileDesc>

- Υποχρεωτικό πεδίο
- Περιλαμβάνει
  - **<titleStmt>**: Πληροφορίες για τον τίτλο του έργου και των πνευματικά υπεύθυνών του
  - **<editionStmt>**: Πληροφορίες σχετικές με την έκδοση του κειμένου
  - **<extent>**: Καταγράφει το «κατά προσέγγιση» μέγεθος του ηλεκτρονικού κειμένου, όπως αυτό αποθηκεύεται σε ένα μέσο, π.χ. 4565 bytes
  - **<publicationStmt>**: Πληροφορίες για την έκδοση ή διάθεση του ηλεκτρονικού εγγράφου
  - **<seriesStmt>**: Πληροφορίες για τη σειρά / ες – αν υπάρχουν – στις οποίες ανήκει η έκδοση
  - **<notesStmt>**: Ομαδοποιεί σημειώσεις που παρέχουν πληροφορίες για ένα κείμενο επιπρόσθετο του κειμένου που κωδικοποιήθηκε σε άλλα σημεία της βιβλιογραφικής περιγραφής
  - **<sourceDesc>**: Περιλαμβάνει βιβλιογραφική περιγραφή των κειμένων από τα οποία προήλθε το ηλεκτρονικό κείμενο



# <titleStmt>

- **<title>**: Ο τίτλος του έργου, συμπεριλαμβανομένων εναλλακτικών τίτλων ή υπότιτλων
- **<author>**: Οι κύρια πνευματικοί υπεύθυνοι ενός έργου, πρόσωπο ή συλλογικό όργανο, π.χ. συγγραφέας
- **<sponsor>**: Σπόνσορας, π.χ. Ένα ινστιτούτο
- **<funder>**: Το όνομα ενός προσώπου ή συλλογικού οργάνου υπεύθυνου για τη χρηματοδότηση του έργου
- **<principal>**: Το όνομα του κύριου ερευνητή που είναι υπεύθυνος για τη δημιουργία του ηλεκτρονικού κειμένου
- **<respStmt>**: Δήλωση πνευματικής υπευθυνότητας όπου δεν μπορεί να καλυφθεί από άλλα στοιχεία
  - **<resp>**: Φράση που περιγράφει τη φύση της πνευματικής υπευθυνότητας, π.χ. *Επιμέλεια του*
  - **<name>**: Περιλαμβάνει το όνομα της πνευματικής υπευθυνότητας

# <titleStmt>

## Παράδειγμα

**<teiHeader>**

**<fileDesc>**

**<titleStmt>**

**<title>**Όνομα ταυρομάχου: μία μηχαναγνώσιμη μορφή του έργου**</title>**

**<author>**Luis Sepulveda**</author>**

**<respStmt>**

**<resp>**Μετάφραση**</resp>**

**<name>**Αχιλλέας Κυριακίδης**</name>**

**</respStmt>**

**</titleStmt>**

**</fileDesc>**

**</teiHeader>**

# <publicationStmt>

- Υποχρεωτικό
- **<publisher>**: Όνομα οργανισμού υπεύθυνου για την έκδοση ή διάθεση του βιβλιογραφικού αντικειμένου
- **<distributor>**: Όνομα προσώπου ή συλλογικού οργάνου υπεύθυνου για τη διάθεση του κειμένου
- **<authority>**: Όνομα προσώπου ή συλλογικού οργάνου υπεύθυνου για τη διάθεση του ηλεκτρονικού εγγράφου, άλλο από *publisher* ή *distributor*.
- **<pubPlace>**: Τόπος έκδοσης βιβλιογραφικού αντικειμένου
- **<address>**: Διεύθυνση, π.χ. του εκδότη
- **<idno>**: Αριθμός που χρησιμοποιείται για την ταυτοποίηση ενός βιβλιογραφικού αντικειμένου
  - **type**: κατηγοριοποιεί τον αριθμό, π.χ. ISBN
- **<availability>**: Πληροφορίες για τη διαθεσιμότητα του κειμένου
  - **status**: Κατάσταση διαθεσιμότητας, π.χ. restricted, unknown, free
- **<date>**: Ημερομηνία, π.χ. `έκδοσης

# <publicationStmt>

## Παράδειγμα

```
<teiHeader>  
  <fileDesc>  
    <titleStmt>...</titleStmt>  
    <publicationStmt>  
      <publisher>Βιβλιοθήκη TAB, Ιόνιο Πανεπιστήμιο</publisher>  
      <pubPlace>Κέρκυρα</pubPlace>  
      <date>20060523</date>  
      <availability status="restricted">  
        <p>Τα πνευματικά δικαιώματα του πρωτότυπου έργου δεν  
        ανήκουν στο φορέα που κωδικοποιεί το έργο. Επιτρέπεται η  
        χρήση του ηλεκτρονικού κειμένου (download) μόνο για μη  
        κερδοσκοπικούς εκπαιδευτικούς σκοπούς.</p>  
      </availability>  
    </publicationStmt>  
  </fileDesc>  
</teiHeader>
```

# <sourceDesc>

- Υποχρεωτικό
- <bibl>: «Χαλαρά» δομημένη βιβλιογραφική αναφορά
- <biblFull>: Πλήρως δομημένη βιβλιογραφική αναφορά
- <listBibl>: Λίστα βιβλιογραφικών αναφορών

## Παράδειγμα

```
<teiHeader>
  <fileDesc>
    <titleStmt>...</titleStmt>
    <publicationStmt>...</publicationStmt>
    <sourceDesc>
      <biblFull>
        <titleStmt>
          <title>Όνομα ταυρομάχου</title>
          <author>Luis Sepulveda</author>
        </titleStmt>
        <publicationStmt>
          <publisher>Εκδόσεις Opera</publisher>
          <pubPlace>Αθήνα</pubPlace>
          <date>1996</date>
        </publicationStmt>
      </biblFull>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

# <text>

- Μεταγραφή του κειμένου
- Ένα TEI κείμενο μπορεί να αποτελεί ένα μοναδικό έργο ή ένα σύνθετο έργο (π.χ. Μία ανθολογία ή μία συλλογή από πολλά έργα).
- Το <text> περιλαμβάνει τα στοιχεία:
  - **<front>**: Προαιρετικό. Απεικονίζει τη σελίδα τίτλου ενός βιβλίου και/ή και άλλες εισαγωγικές πληροφορίες (Title Page και Prefatory Matter)
  - **<body>**: Υποχρεωτικό. Περιλαμβάνει το κείμενο του έργου που κωδικοποιείται.
  - **<back>**: Προαιρετικό. Περιλαμβάνει πληροφορίες όπως ευρετήρια, γλωσσάρια, βιβλιογραφία κτλ.

# <body>

- Στοιχεία υποδιαίρεσης κειμένου (Text Division Elements)

- **<div>** ή **<div1>**, **<div2>**...: Υποδιαίρεση κειμένου στο **<body>**, αλλά και στο **<front>** και **<back>**
  - **type**: Κατηγοριοποιεί την υποδιαίρεση του κειμένου, π.χ. `type="chapter"`
  - **id**: μοναδικός αριθμός για την υποδιαίρεση που χρησιμοποιείται για δημιουργία συνδέσμων, αναφορών κτλ.
  - **n**: όνομα ή αριθμός της υποδιαίρεσης
- **<p>**: Παράγραφος σε πεζό λόγο

## Παράδειγμα

```
<div1 id="WN1" n="1" type="book">
```

```
<div2 id="WN101" n="1.1" type="chapter">
```

```
<p>Με το που φάνηκε η σιλουέτα του καβαλάρη στην άκρη του δρόμου, ο οδηγός του "Lucero de la Pampa" γούρλωσε τα μάτια του...</p>
```

```
</div2>
```

```
</div1>
```

# <body>

- Επικεφαλίδες και «Κλείσιμο» (Headings and Closings)
  - **<head>**: Επικεφαλίδα, π.χ. Επικεφαλίδα κεφαλαίου, μέρους κτλ.
  - **<trailer>**: Τίτλος κλεισίματος ή footer στο τέλος ενός κειμένου

## Παραδείγματα

```
<body>  
  <div1 type="chapter">  
    <head>Γή του Πυρός: Γύπες στον ουρανό</head>  
    <p>...</p>  
  </div1>  
</body>
```

```
<body>  
  <div1 type="chapter">  
    <head>  
      <num>1</num>  
      <title>Γή του Πυρός: Γύπες στον ουρανό</title>  
    </head>  
    <p>...</p>  
  </div1>  
</body>
```



# <body>

- Πεζός λόγος, έμμετρος λόγος και θεατρικό έργο (Prose, Verse and Drama)
  - **<I>**: Περιλαμβάνει μία γραμμή – μπορεί και ημιτελή – έμμετρου λόγου (*δεν περιλαμβάνει τυπογραφικές γραμμές, για αυτό το σκοπό χρησιμοποιείται το <lb> στοιχείο*)
    - **part**: ορίζει αν η γραμμή είναι ολοκληρωμένη ως προς το μέτρο με συγκεκριμένες τιμές (F, Y, N, I, M)
  - **<lg>**: Περιλαμβάνει μία ομάδα γραμμών έμμετρου λόγου, π.χ. μία στροφή, ένα ρεφρέν κτλ.
    - **type**: Ορίζει το είδος της ομάδας γραμμών, π.χ. *verse, stanza* κτλ.

# <body>

- Πεζός λόγος, έμμετρος λόγος και θεατρικό έργο (Prose, Verse and Drama)
  - **<sp>**: Περιλαμβάνει το λόγο ενός ατόμου (*individual speech*)
    - **who**: Ορίζει το όνομα του ομιλητή δίνοντας ένα ID
  - **<speaker>**: Περιλαμβάνει σε μορφή επικεφαλίδας το όνομα ενός ή περισσότερων ομιλητών
  - **<stage>**: Περιλαμβάνει σκηνικές οδηγίες (*stage directions*)
    - **type**: Ορίζει το είδος της οδηγίας, π.χ. *entrance, exit, setting, delivery* κτλ.

# <body>

Παράδειγμα [<http://docsouth.unc.edu/imls/hewitt/hewitt.xml>]

```
<body>
```

```
  <div1>
```

```
    <lg type="verse" n="1">
```

```
      <head>CANTO I.</head>
```

```
        <lg type="stanza" n="1">
```

```
          <l part="N">A sullen murmur, like the moan of  
            waves</l>
```

```
          <l part="N">That circle round old Ocean's hidden  
            caves,</l>
```

```
          <l part="N">And live in story or in graceful song,</l>
```

```
          <l part="N">Your grave the Mecca of th' admiring  
            throng.</l>
```

```
        </lg>
```

```
      </lg>
```

```
    </div1>
```

```
</body>
```

# <body>

## Παράδειγμα

```
<body>
```

```
  <div1 type="Act" n="I">
```

```
    <div2 type="Scene" n="I">
```

```
      <stage type="entrance">Είσοδος του Barnardo και του  
      Francisco στη σκηνή</stage>
```

```
      <sp><speaker>Barn</speaker><l>Ποιος είναι  
      εκεί;</l></sp>
```

```
      <sp><speaker>Fran</speaker><l>Είμαι ο  
      Francisco.</l></sp>
```

```
    </div2>
```

```
  </div1>
```

```
</body>
```

# <body>

- Σελίδες και αριθμοί γραμμών (Page and Line numbers)
  - <pb>: Ορίζει την αρίθμηση των σελίδων (κενό στοιχείο)
  - <lb>: Ορίζει την αρχή μίας νέας τυπογραφικής γραμμής (κενό στοιχείο)

## Παράδειγμα

```
<body>
  <div1 type="chapter">
    <pb n="15"/>
    <head>
      <num>2</num>
      <title>Βερολίνο: Aufwiedersehen (Adios, pampa mia)</title>
    </head>
    <p><lb n="1"/>Ξέρω πως αυτό το γράμμα έχει πολλά σκοτεινά σημεία, πρέπει
    όμως να καταλάβετε πως η μνήμη δεν είναι <lb n="2"/>πάντα αλάθητη και
    πως καμία εξομολόγηση δεν είναι διαυγής όταν τη σκιάζει το βάρος μίας
    προδοσίας...να ξεμουδιάσουν ή να πιουν<pb n="16"/><lb n="11"/> ένα
    φλιτζάνι καφέ...</p>
  </div1>
</body>
```

# <body>

- «Επιφωτισμένες» φράσεις (Highlighted phrases)
  - Για αλλαγές στο “typeface”
    - **rend**: μπορεί να τοποθετηθεί σχεδόν σε όλα τα στοιχεία
      - π.χ. `rend="bold"`, `rend="italic"`
    - **<hi>**, **<emph>**, **<mentioned>**, **<term>**, **<title>**
  - Εισαγωγικά (Quotations and related features)
    - **<q>**, **<mentioned>**, **<soCalled>**, **<gloss>**
  - Ξένες λέξεις ή εκφράσεις
    - **lang**, **<foreign>**

# <body>

## Παραδείγματα

<p>

<q who="Νίκος">Ξαφνικά, ο Βασίλης ήρθε στο γραφείο μου και μου είπε: <q who="Βασίλης">Θα έλθεις μαζί μου στο σπίτι της Ειρήνης;</q></q>

</p>

<p>Ο Γιάννης γνωρίζει το <foreign lang="fre">savoir-vivre</foreign></p>

# <body>

- Σημειώσεις (Notes): Όλες οι σημειώσεις (*footnotes, endnotes, marginals* κτλ) κωδικοποιούνται μέσα στο στοιχείο
  - **<note>**
    - **type**: Ορίζει τον τύπο της σημείωσης, π.χ. type="footnote"
    - **resp**: Ορίζει τον υπεύθυνο για τη σημείωση, π.χ. resp="editor"
    - **place**: Ορίζει το που εμφανίζεται η σημείωση στο κείμενο της πηγής. (Ενδεικτικές τιμές: inline, left, right, margin, foot, end κτλ)

## Παράδειγμα

**<div>**

**<note place="margin" resp="editor">**Εδώ ο συγγραφέας παραθέτει μαρτυρίες για τα Δεκεμβριανά**</note>**

**<p>...</p>**

**<p>...</p>**

**</div>**



# <body>

- Σύνδεσμοι (Cross References και Links)

- Σύνδεσμος από ένα σημείο σε ένα άλλο μέσα στο ίδιο έγγραφο
  - **<ref>**: πιθανότατα συνοδευόμενο από κείμενο ή σχόλια
  - **<ptr>**

Και στα δύο στοιχεία εμφανίζεται το γνώρισμα:

- **target**: Ορίζει τον προορισμό του συνδέσμου, το οποίο είναι συνήθως ένα γνώρισμα ID σε ένα άλλο στοιχείο

## Παράδειγμα

...Βλέπε στο **<ref target="CHAP12">**Κεφάλαιο 12**</ref>**

...**<div1 id="CHAP12">**...

...Βλέπε στο **<ptr target="CHAP12"/>**...

...**<div1 id="CHAP12">**...

# <body>

- Σύνδεσμοι (Cross References και Links)

- Σύνδεσμος από ένα σημείο σε ένα άλλο μέσα στο ίδιο έγγραφο / προς άλλο έγγραφο

- <xref>: πιθανότατα συνοδευόμενο από κείμενο ή σχόλια
- <xptr>

Και στα δύο στοιχεία εμφανίζεται το γνώρισμα:

- doc: Ορίζει το έγγραφο προς το οποίο γίνεται ο σύνδεσμος

## Παράδειγμα

...Βλέπε **<xref doc="P5">**Οδηγίες υλοποίησης TEI P5**</xref>**...

...Βλέπε **<xptr doc="P5"/>**...

# <teiCorpus>

- Υπάρχει η δυνατότητα να ορίσουμε ένα σύνολο TEI texts, όπου κάθε ένα από αυτά έχει το δικό του TEI header.

## Παράδειγμα

**<teiCorpus>**

**<teiHeader>** Πληροφορίες "header" για το Corpus **</teiHeader>**

**<TEI.2>**

**<teiHeader>** Πληροφορίες "header" για το 1<sup>ο</sup> κείμενο **</teiHeader>**

**<text>** 1<sup>ο</sup> κείμενο του Corpus **</text>**

**</TEI.2>**

**<TEI.2>**

**<teiHeader>** Πληροφορίες "header" για το 2<sup>ο</sup> κείμενο **</teiHeader>**

**<text>** 2<sup>ο</sup> κείμενο του Corpus **</text>**

**</TEI.2>**

**</teiCorpus>**

# Δημιουργία TEI εγγράφων

- Επεξεργαστής κειμένου (Notepad, UltraEdit-32 κτλ)

ή

- XML editor (π.χ. XMetal, XMLSPY κτλ)
  - TEI Lite DTD
  - “Well-formed” και “Valid”
  - Οδηγός (template)
  - CSS / XSLT (Cascading Style Sheets / eXtensible Stylesheet Language Transformations)

# Σύνδεσμοι

- TEI Lite (<http://www.tei-c.org/Guidelines2/index.xml.ID=lite>)
- Text Encoding Initiative (TEI) (<http://xml.coverpages.org/tei.html>)
- Projects using the TEI (<http://www.tei-c.org/Applications/>)