TEI and Dublin Core

TEXT ENCODING INITIATIVE (TEI)
• The Text Encoding Initiative (TEI) is an international project to develop guidelines for the preparation and interchange of electronic texts for scholarly research, and to satisfy a broad range of uses by the language industries more generally.
• The TEI is sponsored by the Association for Computers and the Humanities (ACH), the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC).
• Major support for the project has come from the U.S. National Endowment for the Humanities (NEH), Directorate XIII of the Commission of the European Communities (CEC/DG-XIII), the Andrew W. Mellon Foundation, and the Social Science and Humanities Research Council of Canada.

TEXT ENCODING INITIATIVE (TEI)
• There is a pressing need for a common text encoding scheme researchers can use in creating electronic texts.
• Three organizations sponsor TEI: the Association for Computers and the Humanities (ACH), the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC).
• Allows for ability to
  – Search text
  – Select subsets of text
  – “Mix and match” bits of text

TEXT ENCODING INITIATIVE (TEI)
• developed between 1987-1994
• joint industry/educational/govt/non-profit initiative with hundreds of participants
• a subset of SGML developed specifically for humanities applications
• TEI P1 published in 1990
• TEI P2 published between 1992-93
• TEI P3 published in 1994
• TEI P4 published in 1999
• TEI U5 published in 1995 (TEI Lite)
• TEI P5 (XML) published in 2002

TEXT ENCODING INITIATIVE (TEI)
• the Guidelines provide a means of making explicit certain features of a text in such a way as to aid the processing of that text by computer programs running on different machines. This process of making explicit we call markup or encoding.
• The guidelines create a standard and guide for “the preparation and interchange of electronic texts for scholarly research, and to satisfy a broad range of uses by the language industries more generally”

TEXT ENCODING INITIATIVE (TEI)
• Consortium to maintain the TEI formed in 1999
• hosted by:
  – University of Virginia
  – Brown University
  – Oxford University
  – University of Bergen
TEXT ENCODING INITIATIVE (TEI)

- Intended use
  - guidance for individual or local practice in text creation and data capture
  - support for data interchange
  - support of application-independent processing

Goals of the TEI (1987):

- suffice to represent the textual features needed for research;
- be simple, clear, and concrete;
- be easy for researchers to use without special purpose software;
- allow the rigorous definition and efficient processing of texts;
- provide for user-defined extensions;
- conform to existing and emergent standards.

To facilitate the widest possible usership, it was important to ensure:

- the common core of textual features be easily shared;
- additional specialist features be easy to add to (or remove from) a text;
- multiple parallel encodings of the same feature should be possible;
- the richness of markup should be user-defined, with a very small minimal requirement;
- adequate documentation of the text and its encoding should be provided.

TEI has become the de facto standard

- most major humanities computing projects utilize it
- in theory, allows for the exchange of texts across projects and archives
- helps to ensure uniform encoding of text: extremely important for both humans and parsers

What does TEI facilitate?

- structural divisions within a text
  - title-page, chapter, scene, stanza, line, etc
- typographical elements
  - changes in typeface, special characters, etc
- other textual features
  - grammatical structure, location of illustrations, variant forms, etc

What does TEI facilitate?

- facilitates the long-term preservation of electronic texts
- repositories have one basic DTD to preserve and understand
  - though that DTD may have many "flavors"
TEXT ENCODING INITIATIVE (TEI)

- all TEI documents follow the same essential format
- TEI header
  - documents the electronic edition being created
  - the header is essential for:
    • bibliographic control and identification
    • resource documentation and processing
- TEI body
  - contains the content being created

Structure of a TEI text

- A text may be unitary or composite
- a unitary text contains
  - optional front matter
  - optional back matter
  - a body
- in a composite text, the body is replaced by a group of texts (or nested groups)
- A corpus is a collection of text and header pairs, which also has its own header.

A text usually has divisions

- generic, hierarchic subdivisions, each incomplete
- the type attribute is used to label a particular level e.g. as "part" or "chapter"
- vanilla or numbered tags may be used to identify level explicitly
- the n attribute gives a particular division a name or number
- the xml:id attribute gives a particular division a unique identifier
- associated <head> and <trailer> elements (from the divtop class) may also be supplied

TEI Header

The header:

```xml
<teiHeader>
  <fileDesc>
    <encodingDesc>
      Bibliographic information: Author, title, publisher, etc
    </encodingDesc>
    <profileDesc>
      How the material was modified when digitized
    </profileDesc>
    <revisionDesc>
      Non bibliographic information: subject descriptors, etc
    </revisionDesc>
  </fileDesc>
  <teiHeader>
    <text>
      <front>
        <!-- titlepage, etc here -->
      </front>
      <body>
        <div type='book' n='I' xml:id='JA0100'>
          <head>Book I.</head>
          <div type='chapter' n='1' xml:id='JA0101'>
            <head>Of writing lives in general,...
            </div>
          </div>
          <div n='2' xml:id='JA0102'>
            <!-- chapter 2 here -->
          </div>
        </div>
        <div type='book' n='II' xml:id='JA0200'>
          <!-- book 2 here -->
        </div>
        <!-- remaining books here -->
      </body>
    </text>
</teiHeader>
```
Text components

• What are divisions composed of?
  – prose is mostly paragraphs (<p>)
  – verse is mostly lines (<l>), sometimes in
    hierarchic groups (<lg>)
  – drama is mostly speeches (<sp>) containing
    <p> or <l> elements interspersed with stage
    directions (<stage>)
• These may be mixed, and may also
  appear directly within undivided texts.

Short Example

CHAPTER 38

Reader, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said --

'Mary, I have been married to Mr Rochester this morning.' The housekeeper and her husband were of that decent, phlegmatic...

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...

Short Example

<p>Reader, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said --</p>

'Mary, I have been married to Mr Rochester this morning.' The housekeeper and her husband were of that decent, phlegmatic...

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...

Direct speech

• Use the who attribute to show speakers
• Speeches can be nested in other speeches

<q who="Wilson">Spaulding, he came down into the office just this day eight weeks with this very paper in his hand, and he says:--</q>

<q who="Spaulding">I wish to the Lord, Mr. Wilson, that I was a red-headed man.</q>
Correction and Regularization

- `<corr>` marks a correction
- `<sic>` marks a (deliberate) non-correction
- `<reg>` marks a regularization
- `<orig>` marks something deliberately unnormalized

Other Elements

- `<list>` lists of all kinds
- `<note>` notes (authorial or editorial)
- `<figure>` pictures or figures
- `<table>` tables
- `<bibl>` bibliographic descriptions
- **Lists**
  - use `<list>` for lists of any kind (use type attribute to distinguish)
  - use `<label>` in two-column lists as alternative to n attribute
  - may be nested as necessary

Lists

- `<list type="xmas">`
- `<label>For my true love</label>`
- `<item>three calling birds</item>`
- `<item>two french hens</item>`
- `<item>a partridge in a pear tree</item>`
- `<list>`
- `<label>For Uncle Joe</label>`
- `<item>socks as usual</item>`
- `<list>`

Bibliography

- The `<listBibl>` element lists bibliographic citations
- Individual citations may be represented loosely as `<bibl>` elements, or in a more structured way as `<biblStruct>` elements
- In either case, elements from the tei.biblpart class are used, e.g.
  - `<author>`, `<editor>`, (generic) `<respStmt>` etc.
  - `<title>` with optional level attribute to distinguish monographic, analytic etc.
  - `<imprint>` groups publication info (publisher, date etc.)
  - `<biblScope>` adds page references etc.

Biblioography

- `<div><head>Bibliography</head>`
- `<listBibl>`
- `<bibl xml:id="REG92">`
- `<author>Ed Regis</author>`
- `<title level="m">Great Mambo Chicken and the Trans-Human Experience</title>`
- `<pubPlace>London </pubPlace>`
- `<publisher>Penguin Books</publisher>`
- `<date>1992</date>`
- `<biblScope>pp 144 ff</biblScope>`
- `<bib>`
- `</listBibl>`
- `</div>`

Poem

- `<DIVO TYPE="poem">`
- `<HEAD>Straw in the Street.</HEAD>`
- `<LG TYPE="stanza">`
- `<L><H1>STRAW</H1> in the street where I pass to\&hyphen;day</L>`
- `<L>Dulls the sound of the wheels and feet.</L>`
- `<L>&rsquo;Tis for a failing life they lay</L>`
- `<L REND="indent1">Straw in the street.</L>`
- `</DIVO>`
Shakespeare in TEI

<sp><speaker>Kent</sp><br />
Thou swear'st thy gods in vain.<br />
O vassal! miscreant!<br />
Laying his hand on his sword.<br />
Dear sir, forbear!<br />
Do; Kill thy physician, and the fee bestow<br />
Upon the foul disease. Revoke thy gift.<br />
Or, whilst I can vent clamour from my throat.<br />
I'll tell thee thou dost evil.<br />

Structure of a TEI Document

Unitary Text

<TEI.2>
<teiHeader>
<text>
<front>
<body>
<back>
</text>
</teiHeader>
</TEI.2>

Composite Text

<TEI.2>
<teiHeader>
<text>
<front>
<group>
<text>
<front>
<body>
<back>
</text>
</group>
<!-- more texts or groups of texts here -->
</text>
</teiHeader>
</TEI.2>

TEI Corpus

<teiCorpus>
<teiHeader>
<teiHeader>
<text>
<front>
<body>
<back>
</text>
</teiHeader>
<text>
<front>
<body>
<back>
</text>
</teiHeader>
<!-- more texts here -->
</teiCorpus>
**TEI Lite**

- a subset of TEI designed to meet average encoding needs
- the TEI encoding scheme might be adopted to meet 90% of the needs of 90% of the TEI user community.
- TEI Lite is most suited to
  - printed texts
  - structural encoding
  - light content encoding
  - electronic repositories which want to make lightly encoded texts available to scholars

**Goals of TEI Lite**

- it should include most of the TEI “core” tag set, since this contains elements relevant to virtually all text types and all kinds of text-processing work;
- it should be able to handle adequately a reasonably wide variety of texts, at the level of detail found in existing practice
- it should be useful for the production of new documents as well as encoding of existing ones;

**how does TEI Lite fulfill its goals**

- for encoding poetry, TEI Lite offers:
  - `<lg>` for encoding line groups (stanzas)
  - `<l>` for encoding a line of poetry
- the verse tag set offers:
  - "numbered" `<lg>` elements are provided, by analogy with the "numbered" divn class elements
  - a special purpose `<caesura>` element is provided, to allow for segmentation of the verse line
  - a set of attributes is provided for the encoding of rhyme scheme and metrical information

**Encoding the Body**

- `<front>` contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found before the start of a text proper.
- `<group>` contains a number of unitary texts or groups of texts.
- `<body>` contains the whole body of a single unitary text, excluding any front or back matter.
- `<back>` contains any appendixes, etc., following the main part of a text.

**Text Division Elements**

**Main Elements**

- `<p>` marks paragraphs in prose.
- `<div>` contains a subdivision of the front, body, or back of a text.
- `<div1>` contains a first-level subdivision of the front, body, or back of a text (the largest, if `<div0>` is not used, the second largest if it is).

**Common Attributes**

- `type` This indicates the conventional name for this category of text division.
- `id` This specifies a unique identifier for the division, which may be used for cross references or other links to it.
- `n` This attribute specifies a mnemonic short name or number for the division.
Headings and Closings

Every <div>, <div1>, <div2>, etc., may have a title or heading at its start, and (less commonly) a closing such as ‘End of Chapter 1’. The following elements may be used to transcribe them:

- **<head>**
  contains any heading, for example, the title of a section, or the heading of a list or glossary.

- **<trailer>**
  contains a closing title or footer appearing at the end of a division of a text.

Prose, Verse, and Drama

- **<l>**
  contains a single, possibly incomplete, line of verse. Attributes include:

- **<lg>**
  contains a group of verse lines functioning as a formal unit e.g. a stanza, refrain, verse paragraph, etc.

- **<ap>**
  contains an individual speech in a performance text, or a passage presented as such in a prose or verse text.

- **<speaker>**
  contains a special form of heading or label, giving the name of one or more speakers in a performance text or fragment.

- **<stage>**
  contains any kind of stage direction within a performance text or fragment.

Page and Line Numbers

- **<pb>**
  marks the boundary between one page of a text and the next in a standard reference system.

- **<lb>**
  marks the start of a new (typographic) line in some edition or version of a text.

- **<milestone>**
  marks the boundary between sections of a text, as indicated by changes in a standard reference system. Attributes include:

  - **<ed>** indicates the edition or version to which the milestone applies.
  - **<unit>** indicates what kind of section is changing at this milestone.

Changes of Typeface

- **<hi>**
  marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made.

- **<emph>**
  marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.

- **<foreign>**
  identifies a word or phrase as belonging to some language other than that of the surrounding text.

- **<mentioned>**
  marks words or phrases mentioned, not used.

- **<term>**
  contains a single-word, multi-word or symbolic designation which is regarded as a technical term.

- **<title>**
  contains the title of a work, whether article, book, journal, or series, including any alternative titles or subtitles.

Quotations & Related Features

- **<q>**
  contains a quotation or apparent quotation — a representation of speech or thought marked as being quoted from someone else (whether in fact quoted or not); in narrative, the words are usually those of a character or speaker; in dictionaries, <q> may be used to mark real or contrived examples of usage.

- **<mentioned>**
  marks words or phrases mentioned, not used.

- **<soCalled>**
  contains a word or phrase for which the author or narrator indicates a disclaiming of responsibility, for example by the use of scare quotes or italics.

- **<gloss>**
  marks a word or phrase which provides a gloss or definition for some other word or phrase.

Minimal TEI Header

```xml
<teiHeader>
	<titleStmt>
		<title>Introduction to cataloging and classification</title>
		<respStmt>
			<name>Bohdan S. Wynar</name>
			<resp>8th edition by</resp>
			<name>Arlene G. Taylor</name>
		</respStmt>
	</titleStmt>
	<publisherStmt>
		<distributor>Libraries Unlimited</distributor>
	</publisherStmt>
	<sourceDesc>
	</sourceDesc>
</teiHeader>
```
What Does the TEI Offer for Digital Libraries?
- A framework for encoding text
- A multi-purpose encoding scheme
- A better encoding system for text than any other previous one

TEI and Libraries: Encoding
- Not a mechanical process
- Who does the encoding?
- What level of markup do they use?
- Encoding is interpretation
- Encoding drives what can be done with the document
- To what extent is this the role of a library?

TEI and Libraries
- In the library, some standardization is essential
- But at what level?
- Is TEI Lite the answer here?
- Few people will get an electronic text from a digital library and use their own software on it
- Most will want to use an existing delivery system
- But with what functionality?

TEI and Libraries
- Retrieval – look up words and phrases
- Determined by the encoding, but also by the indexing
- Who decides what to index?
- Browsing: Random movement from one piece of information to another
- Can be done by linking
- But who decides what the links are?

TEI and Libraries
- Can automate some of the searching and linking
- But it does not always make sense
  - Same name appearing several times in one document
  - Importance of links – intellectual decision

TEI Tools
- TEI Software
  A list of links to software for creating, managing, and processing TEI documents in SGML or XML
- Customized Templates for EAD-Encoded Finding Aids
  [http://sunsite.berkeley.edu/FindingAids/uc-ead/templates/](http://sunsite.berkeley.edu/FindingAids/uc-ead/templates/)
  Templates for OAC (The Online Archive of California) Project participants to generate EAD version 1.0 markup.
**Dublin Core**

- Initiative to improve *resource discovery* on Web
  - not for complex resource description
  - simple "document-like objects"

- Interdisciplinary consensus on simple element set
  - 15 elements
  - all optional
  - all repeatable

**Dublin Core**

- Initially proposed at a workshop held in March 1995 at Dublin, Ohio
- Librarians and archivists, researchers, computer and information scientists, software developers, publishers, and members of Internet Engineering Task Force (IETF)
- Developed through a series of workshops
- Standard agreed upon by consensus of attendees: librarians, web system designers, commercial information providers

**Dublin Core**

- All elements optional and repeatable
- Elements display in any order
- Authority control not required
- Simple and Qualified DC
- Extensible
- Flexible
- International

**Dublin Core**

- Simple
  - Lowest common denominator
  - Less rich
  - Discovery role – leads to resource or more complete description of resource
- Qualified
  - More precise
  - Less interoperable

**The Dublin Core Metadata Element Set**

- **TITLE**: The name given to the resource by the CREATOR or PUBLISHER.
- **CREATOR**: The person(s) or organization(s) primarily responsible for the intellectual content of the resource.
- **SUBJECT**: The topic of the resource, or keywords or phrases that describe the subject or content of the resource.
- **DESCRIPTION**: A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources.
- **PUBLISHER**: The entity responsible for making the resource available in its present form, such as a publisher, a university department, or a corporate entity.
- **CONTRIBUTORS**: Person(s) or organization(s) in addition to those specified in the CREATOR element who have made significant intellectual contributions to the resource but whose contribution is secondary to the individuals or entities specified in the CREATOR element.
- **DATE**: The date the resource was made available in its present form.
- **RESOURCE TYPE**: The category of the resource.
The Dublin Core Metadata Element Set

- **FORMAT:** The data representation of the resource.
- **RESOURCE IDENTIFIER:** String or number used to uniquely identify the resource.
- **SOURCE:** The work, either print or electronic, from which this resource is derived, if applicable.
- **LANGUAGE:** Language(s) of the intellectual content of the resource.

---

**Dublin Core**

- **TITLE:** Introduction to cataloging and classification
- **CREATOR:** Taylor, Arlene G.
- **OTHER CONTRIBUTOR:** Wynar, Bohdan S.
- **DATE:** 1992
- **FORMAT:** BOOK
- **LANGUAGE:** ENG
- **PAGES:** 633
- **PUBLISHER:** Libraries Unlimited
- **SUBJECT:** Cataloging.

**RESOURCE TYPE:** text.monograph

**RESOURCE IDENTIFIER:** (ISBN) 0872879674

---

**Qualified Dublin Core**

- Dublin Core can be simple or qualified
- Qualifiers **EXPAND** the element
- 2 kinds of qualifiers, refinements & encoding
- Additional elements
Dublin Core Qualifiers

- **Title**
  The qualifiers below are recommended for the Title element.
- **Qualifiers that refine Title:**
  - Alternative
    - Name: alternative
    - Label: Alternative
    - **Definition:** Any form of the title used as a substitute or alternative to the formal title of the resource.
    - **Comment:** This qualifier can include Title abbreviations as well as translations.

Dublin Core Qualifiers

- The qualifiers below are recommended for the Subject element.
- **Encoding Schemes for Subject:**
  - LCSH
    - Name: LCSH
    - Label: LCSH
    - **Definition:** Library of Congress Subject Headings
  - MeSH
    - Name: MESH
    - Label: MeSH
    - **Definition:** Medical Subject Headings

Qualified Dublin Core: Additional Elements

- audience
  - mediator
  - educationLevel
- provenance
- rightsHolder
- instructionalMethod
- accrualMethod
- accrualPeriodicity
- accrualPolicy

Use of the Dublin Core

- As a header for an HTML document
- As a lowest common denominator into which other metadata formats can be poured to allow searching over disparate databases of descriptive data--the “bucket” metaphor
- As a source for search engines in preference to indexing of entire web documents

Dublin Core in Libraries

- Some libraries and special projects using DC now
  - Many non-US libraries actively using DC, especially in Australia and Scandinavia
- DC built into Connexion
  - Many of us will encounter it there first
  - Includes automated mapping between MARC and DC: may have more uses in future
- DC may be good choice for cataloging Web sites, digitized files, etc.
  - may already be present in some Web sites we want to provide local access to

Dublin Core

- Dublin Core metadata sets reside in the header of web-resources, fueling search engines.
- With a metadata set a web-resource becomes, essentially, self-describing.
- The Dublin Core set is very much a work in progress. That is, it is a dynamic standard, being updated and revised more or less constantly.
- This means that keeping an eye on changes will be critical for staying up to date with the DCMI. Syntax is described in the usage guide.
Dublin Core

– Simplicity of semantics, ease of use
– Provides basic semantic interoperability
  • across domains
  • across language communities
– Allows for extensibility
  • but tension between extending DC and choosing other, richer schema

Dublin Core

– Interoperability requires
  • use of content rules/standards
  • clarity about resource being described
    – e.g. work, expression, manifestation, item
– Real resources more complex than (stable) "document-like object"?
  • characteristics of resources change through time
  • agents perform actions which produce changes

Dublin Core

– Not a replacement for richer descriptive standards
– A “pidgin” language for use by “tourists on the Internet commons”
  • Tom Baker, “A Grammar of Dublin Core”
– Can provide 15 “windows” into richer resource descriptions
  • disclose rich description in simple form
  • semantic cross-walks, mappings

Dublin Core

• A basic Dublin Core set looks like this:
  <dc:creator>Rose Bush</dc:creator>
  <dc:description>Describes process for planting and nurturing different kinds of rose bushes.</dc:description>
  <dc:date>2001-01-20</dc:date>

Dublin Core

• <HTML>
  <HEAD>
    <META NAME="DC.title" CONTENT="Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment">
    <META NAME="DC.creator" CONTENT="Gilliland-Swetland, Anne J." SCHEME="LCNA">
    <META NAME="DC.publisher" CONTENT="Council on Library and Information Resources">
    <META NAME="DC.date" CONTENT="2000">
    <META NAME="DC.type" CONTENT="Technical Report">
    <META NAME="DC.identifier" CONTENT="http://www.clir.org/pubs/abstract/pub89abst.html">
    <META NAME="DC.source" CONTENT="ISBN 1-887334-74-2">
  </HEAD>
  <BODY>
  </BODY>

Extending the Dublin Core

– Does allow for extensibility
  • but tension between extending DC and choosing other, richer schema
  • greater specificity = lesser interoperability?
– Improve semantic precision of DC elements using qualifiers
  • element refinements
    – make the meaning of an element narrower
  • value encoding schemes
    – specify that value is from controlled vocabulary, or formatted in a standard way
<table>
<thead>
<tr>
<th>Qualified DC: Identifier (of the resource)</th>
<th>Qualified DC: Publisher</th>
</tr>
</thead>
</table>
| • <meta name="DC.Identifier" scheme="LCCN" content="18008632">  
• <meta name="DC.Identifier" scheme="URI" content="http://www.gutenberg-e.org"> | • <meta name="DC.Publisher.place" content="London;">  
• <meta name="DC.Publisher" content="H.M. Stationery off. [Harrison and Sons, printers]">  
• Note: Qualifiers not approved by DCMI |

<table>
<thead>
<tr>
<th>Qualified DC: Description (textual description or abstract)</th>
<th>Qualified DC: Type (category or genre)</th>
</tr>
</thead>
</table>
| • <meta name="DC.Description.TableofContents" lang="en" content="The Author gives some Account of Himself and Family -- His First Inducements to Travel -- He is Shipwrecked, and Swims for his Life -- Gets safe on Shore in the Country of Lilliput -- Is made a Prisoner, and carried up the Country">  
• <meta name="DC.Description.Abstract" content="The kinematics of the jaws and hyolingual apparatus in Caiman crocodilus were examined by cineradiography and electromyography. After catching, caimans position their prey between the teeth by a series of inertial bites and then kill and crush it by a forceful bite.""> | • <meta name="DC.Type" scheme="DCMIType" content="Software">  
• <meta name="DC.Type" scheme="DCMIType" content="Dataset">  
• <meta name="DC.Type" scheme="DCMIType" content="Event">  
• <meta name="DC.Type" scheme="AACR2-gmg" content="[electronic resource]"> |

<table>
<thead>
<tr>
<th>Qualified DC: Language (of the content of the resource)</th>
<th>Qualified DC: Date (of creation or availability of resource;)</th>
</tr>
</thead>
</table>
| • <meta name="DC.Language" content="en">  
• <meta name="DC.Language" content="en;fr">  
• <meta name="DC.Language" scheme="rfc1766" content="en">  
• <meta name="DC.Language" scheme="ISO639-2" content="eng">  
• <meta name="DC.Language" scheme="rfc1766" content="en-US"> | • <meta name="DC.Date.Created" content="1998-05-14">  
• <meta name="DC.Date.Available" content="1998-05-21">  
• <meta name="DC.Date.Valid" content="1998-05-28">  
• <meta name="DC.Date.issued" scheme="MARC21-Date" content="2002-9999"> |
Qualified DC: Relation (relationship to a second resource plus its identifier)

- `<meta name="DC.Relation.IsBasedOn" content="Shakespeare's Romeo and Juliet">`

The Dublin Core in context

- In practice, metadata implementers
  - combine elements from different sources (e.g. DC plus elements from other schemas, "local" elements)
  - refine definitions of elements
  - constrain use of elements
- Application profiles
  - if simple DC is a "pidgin", an application profile is a "regional idiom or creole" (Baker)
  - element set plus policies, guidelines

The Dublin Core in context

- DC provides **simple** element set for cross-domain resource discovery
- Supported by open community of practitioners and theorists
- Widely adopted - but not a complete, off-the-shelf solution

Dublin Core

- **Drawbacks:**
  - Too Flexible and Simple for complex, sophisticated collections;
  - Elements lack standardized use and precision.
  - Different Communities are developing extensions to specify and categorize the elements. Approved extensions are available but slow to appear.
  - Some elements (rights, coverage) are ambiguous in their application
  - Intended for web objects that are textual or primarily textual. Does not provide for:
    - Media asset components (video sequences, scenes, shots, frames, objects);
    - sequential media (audio and video, slide shows);
    - synchronized media (video, audio, caption file or transcription; slide shows).

Dublin Core Templates

- Nordic Dublin Core record creation template
  [http://www.lub.lu.se/cgi-bin/nmdc.pl](http://www.lub.lu.se/cgi-bin/nmdc.pl)
  (Note: This tool will generate a record in HTML4 format, not a valid XML format. The template has included very useful pull-down lists for the metadata values.)

- DC-Dot's Dublin Core metadata editor
  [http://www.ukoln.ac.uk/metadata/dcdot/](http://www.ukoln.ac.uk/metadata/dcdot/)
  (Note: submit any webpage's URL and get a suggested metadata record in XHTML format. Then use the template to edit the record.)

Dublin Core Tools

- [http://dublincore.org/tools](http://dublincore.org/tools)
  - Utilities
  - Creating Metadata (Templates)
  - Tools for the Creation/Change of Templates
  - Automatic Extraction/Gathering of Metadata
  - Automatic Production of Metadata
  - Conversion Between Metadata Formats
  - Integrated (Tool) Environments
  - Commercially Available Software
<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Simple</td>
<td>• Can be too simple</td>
</tr>
<tr>
<td>• Dublin Core (DC) has 15 elements, all</td>
<td>• Doesn’t facilitate serendipitous discovery</td>
</tr>
<tr>
<td>repeatable,</td>
<td>very well</td>
</tr>
<tr>
<td>• Qualified Dublin Core (qDC) provides</td>
<td>• Ambiguous input standards</td>
</tr>
<tr>
<td>refinements on the basic elements</td>
<td>• Implementations vary</td>
</tr>
<tr>
<td>• Compatible with large variety of digital</td>
<td></td>
</tr>
<tr>
<td>objects</td>
<td></td>
</tr>
<tr>
<td>• Intuitive element names</td>
<td></td>
</tr>
<tr>
<td>• Clear documentation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>