

Comp 336/436 - Markup Languages

Fall Semester 2018 - Week 14

Dr Nick Hayward

Text Encoding Initiative - textual phenomena

title pages

- `<titlePage>` used for transcription and encoding of physical title page
- within `<front>` or `<back>` using `<titlePage>`
- `<titlePage>` commonly includes `<docTitle>`
- `<docTitle>` may include subsections, `<titlePart>`
- `<docAuthor>` for name of the author
- `<byline>` for primary statement of responsibility for a work
- other elements can include
 - `<docEdition>`
 - `<docDate>`
 - `<docImprint>` *may contain*
 - `<pubPlace>`
 - `<docDate>`
 - `<publisher>`
 - ...
 - `<epigraph>`
 - `<imprimatur>`
 - `<graphic/>`

Text Encoding Initiative - textual phenomena

title pages - examples

- <docTitle> can consist of several subsections
- <titlePart>, which may itself use a @type attribute documenting their role.

```
<docTitle>
  <titlePart type="main">main title...</titlePart>
  <titlePart type="sub">sub-title</titlePart>
</docTitle>
```

- an inline graphic, illustration, or figure <graphic/>

```
<figure>
  <graphic url="fig1.png"/>
  <head>.</head>
  <figDesc>.</figDesc>
</figure>
```

Text Encoding Initiative - general usage

Exercise - part 2 - Front & Title Page with Image

Use the image on the next slide to encode a title page in the front part of a text.

Encode the following:

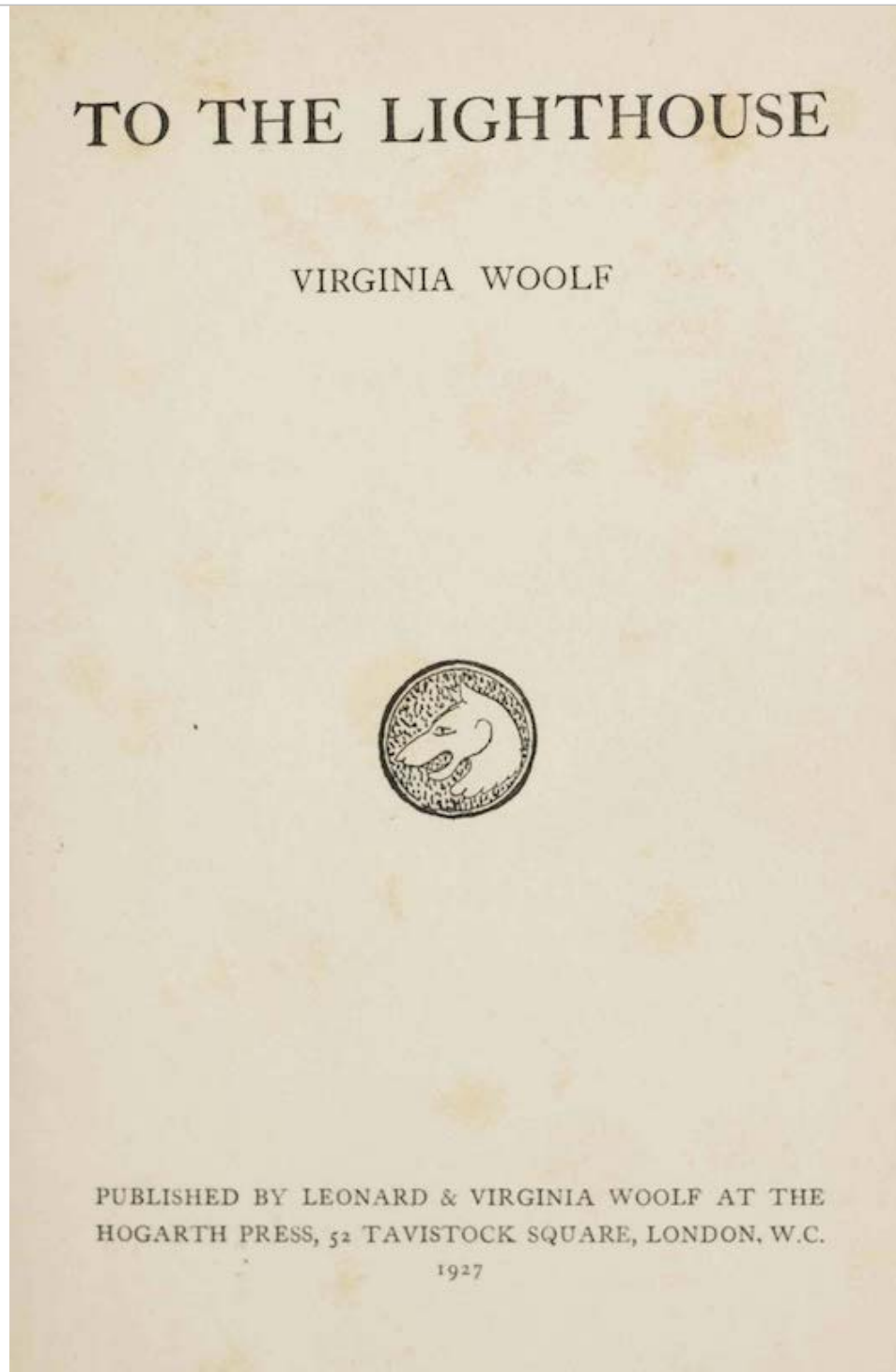
1. title and author information
2. any stylistic considerations in the font, size, spacing...
3. the graphic
4. any other appropriate information such as edition, publication information...

Please feel free to consult the TEI guidelines where necessary.

- ~ 15 minutes

Image - To The Lighthouse

front page



To the Lighthouse - GB 1st Edition

Text Encoding Initiative - general usage

front & title page example

A simple example of encoding a front and title page

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title/>
      </titleStmt>
      <publicationStmt>
        <publisher/>
      </publicationStmt>
      <sourceDesc>
        <p/>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <front>
      <titlePage>
        <docTitle>
          <titlePart type="main">TO THE LIGHTHOUSE</titlePart>
        </docTitle>
        <byline>
          <docAuthor>Virginia Woolf</docAuthor>
        </byline>
        <figure>
          <graphic url="hogarth-logo.jpg"/>
          <figDesc>Black and white rendition of Hogarth Logo</figDesc>
        </figure>
        <docImprint>
          <publisher>PUBLISHED BY LEONARD & VIRGINIA WOOLF AT THE HOGARTH
          <pubPlace>52 TAVISTOCK SQUARE, LONDON. W.C.</pubPlace>,
          <docDate>1927</docDate>
        </docImprint>
      </titlePage>
    </front>
  </text>
</TEI>
```

```
        </titlePage>
    </front>
    <body>
    </body>
</text>
</TEI>
```

n.b. in the <docImprint> element

- for the <publisher> element - we could be more specific and use <name> elements
- e.g. for Leonard and Virginia Woolf
- for the <pubPlace> element - we could enclose this within an <address> element
- this might help to add further structure and metadata
- for the <docDate> element - we could add a @when attribute with value set to the year 1927
- again, it helps to add further metadata

Text Encoding Initiative - textual phenomena

logical and semantic features

- encodes underlying semantics with specific elements other than `<hi>`
- highlighting may be used to show

```
<emph> <foreign> <distinct>
```

- uses of quotation

```
<said> <quote> <cit> <mentioned> <soCalled>
```

- other examples include

```
<term> <gloss>
```

- generic elements may also carry semantic and logical information

```
<title> <name> <num> <measure> <date> <address> <abbr> <expansion>
```


Text Encoding Initiative - textual phenomena

logical and semantic features - referring strings

- in TEI
 - *places, people, objects can use the `<name>` elements with `@type` attribute*
 - e.g.

```
<name type="person">Jack</name>
```

- proper nouns can also be referred to using the `<rs>` element plus `@type` attribute
 - e.g.

```
<rs type="person">he</rs>
```

Text Encoding Initiative - general usage

Exercise - part 3 - common structure and elements

Use the excerpt on the next slide to encode a structured grouping of text, e.g. a division, segment, page...

Encode the following:

1. this text as a paragraph of prose
2. indicate that the language of this paragraph is English
3. indicate that the paragraph is number 7
4. encode any abbreviations with their correct expansions
5. encode names with the appropriate element and attribute

Please feel free to consult the TEI guidelines where necessary.

- ~ 15 minutes

Text Encoding Initiative - general usage

Exercise - part 3 - text excerpt

- encode the following text excerpt

"Good-evening, Mrs. McNab," she would say.
She had a pleasant way with her. The girls
all liked her. But dear, many things had changed
since then (she shut the drawer); many families
had lost their dearest. So she was dead; and
Mr. Andrew killed; and Dr. Prue dead too,
they said, with her first baby; but every one had
lost some one these years. Prices had gone up
shamefully, and didn't come down again neither.
She could well remember her in her grey cloak.

Text Encoding Initiative - general usage

text excerpt - element encoding of local language

A simple example of a text excerpt with element defined local language encoding.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title/>
      </titleStmt>
      <publicationStmt>
        <publisher/>
      </publicationStmt>
      <sourceDesc>
        <p/>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p xml:lang="en" n="7">Good-evening,
        <name type="person">
          <choice>
            <abbr type="title">Mrs.</abbr>
            <expan>Mistress</expan>
          </choice>
          McNab</name>
          , " she would say.</p>
    </body>
  </text>
</TEI>
```

n.b.

- en-GB - the value must conform to BCP 47 (tags for identifying languages)
- the <rs> element with a @type attribute can also be used instead of <name>
- <name> is solely for proper nouns
- the rest of the paragraph in the example follows the same pattern

Text Encoding Initiative - general usage

text excerpt - header encoding of global language

A simple example of a text excerpt with header defined global language encoding.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title/>
      </titleStmt>
      <publicationStmt>
        <publisher/>
      </publicationStmt>
      <sourceDesc>
        <p/>
      </sourceDesc>
    </fileDesc>
    <profileDesc>
      <langUsage>
        <language ident="en-GB">British English</language>
      </langUsage>
    </profileDesc>
  </teiHeader>
  <text>
    <body>
      <p n="7" [5]>"Good-evening,
        <name type="person">
          <choice>
            <abbr type="title">Mrs.</abbr>
            <expan>Mistress</expan>
          </choice>
          McNab</name>
          ," she would say.</p>
    </body>
  </text>
</TEI>
```

n.b.

- header encoding of global language this helps specify the languages used throughout the text being encoded
- we could specify the language used, as in the above example, but now it would only make sense if the language was different from the specified default in the TEIHeader.

Text Encoding Initiative - textual phenomena

logical and semantic features - dates and time

- `<date>` and `<time>`
- system or calendar used may be documented using `@calendar` attribute
- value of date or time supplied using `@when` attribute
- normalised representation of `<date>` should conform to valid W3C datatype

```
<date when="2003-12-22" calendar="Gregorian">22 Nov 2003</date>
```

- `<date>` can also be used to mark a period of time using the attributes
 - *@from*
 - *@to*
 - *@notBefore*
 - *@notAfter*

Text Encoding Initiative - textual phenomena

logical and semantic features - numbers and measures

- <num>

```
<num type="percentage" value="23">23%</num>
```

- <measure>

```
<measure type="volume" quantity="1.5" unit="litre" commodity="wine">1.5L k
```

Text Encoding Initiative - textual phenomena

Exercise - part 4

If **20,000 Leagues Under the Sea** was an actual recording of the distance travelled by Captain Nemo in the Nautilus, how would you specify this using TEI?

n.b. league = ~ 3 nautical miles or 3.5 standard English miles

- create a TEI encoded XML document
 - *add some details for the novel **20,000 Leagues Under the Sea** by **Jules Verne***
 - *correctly encode **20,000 leagues***
 - *write a simple XSL stylesheet to render the content and metadata to HTML*
 - *test in a browser*
- ~ 15 minutes

Text Encoding Initiative - textual phenomena

logical and semantic features - addresses

- physical and digital addresses can be encoded
 - *<address>*, *<email>*
- *<email>* can also use a @type attribute
- *<addrLine>* within *<address>*
 - *<name>*, *<street>*, *<postCode>*, *<postBox>*

Text Encoding Initiative - textual phenomena

addresses - examples

```
<address>  
  <addrLine></addrLine>  
</address>
```

- you can also be specific, and use semantically rich elements

```
<name>  
<street>  
<postCode>  
<postBox>
```

Text Encoding Initiative - textual phenomena

Exercise - part 5

- how would you encode **Alice's** address in **Wonderland** using TEI?
- create example TEI XML and encode this **fictional** address
- write a simple XSL stylesheet to render the content and metadata to HTML
- test in a browser
- ~ 10 minutes

Text Encoding Initiative - textual phenomena

logical and semantic features - abbreviations and expansions

- explicitly encode
 - `<abbr>`
 - `<expn>`
- @type attribute may be used with `<abbr>`
 - e.g.

```
<abbr type="title">Dr</abbr>
```

- `<expn>` used with `<abbr>` within `<choice>`

Text Encoding Initiative - textual phenomena

abbreviations and expansions - example

```
<choice>  
  <abbr type="title">Dr</abbr>  
  <expan>Doctor</expan>  
</choice>
```

Text Encoding Initiative - textual phenomena

Exercise - part 6

- how might you use the `choice` element to encode **PhD**?
- use TEI encoded XML
- ~ 10 minutes

Text Encoding Initiative - textual phenomena

analytical features - notes and annotations

- `<note>` can be used to record a textual annotation
- `@type` attribute used to differentiate notes
- `@resp` attribute used to assign responsibility
- position of the note can be referenced using the `@place` attribute

e.g.

```
<note n="1" place="foot" type="editorial" resp="NJH">a new note...</note>
```

Text Encoding Initiative - textual phenomena

analytical features - index entries

- pre-existing in <front> or <back>
 - *using <list> inside <div>*
- new index using <term> inside <index>
 - *add at the location of the index item*
 - e.g.

```
<index>  
  <term>new index term...</term>  
</index>
```

Text Encoding Initiative - textual phenomena

analytical features - errors

- indicated using `<sic>`
- corrected using `<corr>`
- combine `<sic>` and `<corr>` within `<choice>`
- use `@certand` `@resp` attributes to encode degree of certainty and editor responsible
- e.g.

```
<corr cert="high" resp="#NJH">correction...</corr>
```

- **n.b.** *hash* in the `@resp` attribute value
 - a pointer to a name element in the `<teiHeader>`
 - e.g.

```
<name xml:id="NJH">...</name>
```

Text Encoding Initiative - textual phenomena

Exercise - part 7

- how would you add a pointer to a name element in the `<teiHeader>`?
- use TEI encoded XML
- ~ 10 minutes

Text Encoding Initiative - textual phenomena

<respStmt> - example

- place this name within the <respStmt> element as follows,

```
<respStmt>
  <resp>editor</resp>
  <name xml:id="NJH">Nicholas J Hayward</name>
</respStmt>
```

- then reference this editor as necessary in the encoded document

Text Encoding Initiative - textual phenomena

analytical features - regularisation

- `<reg>` element for regularisation
- `<orig>` for original, non-normalised form
- use `<reg>` in isolation or combined with `<orig>` within `<choice>`
- e.g.

```
<choice>
  <orig>thou</orig>
  <reg resp="#NJH">you</reg>
</choice>
```

Text Encoding Initiative - textual phenomena

analytical features - additions, deletions & omissions

- `<gap>` element used for omission, both material and editorial
- `@reason` attribute used to indicate reason for omission
- `@extent` and `@unit` attributes can be used to record extent of omission
- editorial omissions should be recorded using `<editorialDecl>`
 - *add inside `<editorialDesc>` in `<teiHeader>`*
- `<gap>` may be empty or include a `<desc>` of the material omitted
- `<add>` and `` may also be used for words and phrases
- `@rend` attribute may also be used with `<add>` and ``
- `<addSpan/>` and `<delSpan/>` for longer passages
- `<subst>` to contain `<add>` and `` with causal relationship
- `<unclear>` with `@reason` attribute for difficult to read deletions in the text

Text Encoding Initiative - non-textual phenomena

graphics &c.

- graphics such as illustrations, diagrams, drawings, artwork...
- anchor in a text using `<graphic/>` and optional `@url` attribute
- e.g.

```
<graphic url="http://www.somewhere.com/image.jpg"/>
```

- use `<figure>` element as parent to create a full listing for a graphic

Text Encoding Initiative - Attributes

global Attributes

Global Attributes are currently as follows:

- @cert
 - *provided by responsibility subclass*
- @n
- @rend
 - *provided by rendition subclass*
- @rendition
 - *provided by rendition subclass*
- @resp
 - *provided by responsibility subclass*
- @source
 - *provided by source subclass*
- @style
 - *provided by rendition subclass*
- @xml:base
- @xml:id
- @xml:lang
- @xml:space

Text Encoding Initiative - Further Examples

bibliography listing

- Sample from Verne Digital Corpus

Text Encoding Initiative - Further Examples

regularise

- regularise with soundex and metaphone
- regularise with soundex and metaphone - basic compare
- regularise with soundex and metaphone - array compare

Text Encoding Initiative - Further Examples

working with images

- working with image and transcription
 - *representation of primary sources*
- TEI
 - *Facsimile - guidelines & examples*
 - *Surface - guidelines & examples*
 - *Zone - guidelines & examples*

XML & Semantic Web

- semantic web as a broad grouping of organised data
- different XML-based recommendations
- RDF (Resource Description Framework)
 - *XML based text format*
 - *supports resource description and metadata apps*
- GRDDL (Gleaning Resource Descriptions from Dialects of Languages)
 - *introduces markup based on existing standards*
 - *declare XML document with data compatible with RDF*
 - *add linking to algorithms - usually as XSLT*
 - *used to extract data from document...*
- OWL (Web Ontology Language)
 - *semantic markup language for publishing and sharing ontologies*
 - *OWL is a vocabulary extension of RDF*
 - *derived from DAML+OIL Web Ontology Language*
- SPARQL
 - *query language for RDF*
 - *used to express queries across diverse data sources*

Metadata typology and uses - intro

the main goal of metadata, whether considered as a theory or implemented in a functional system, is to make data useful...

- Garoufallou, E., Greenberg, J., *Metadata and Semantics Research: 7th International Conference*. MSTR. 2013.
- scholars and practitioners need to ascribe types to things
- classify kinds of *things*
 - *enables better understanding of variety and complexity...*
- need to typify metadata
 - *also metadata schemas, metadata uses, metadata elements, metadata values...*
- **types of metadata** includes different concepts
 - *types of metadata standards*
 - *types of metadata values*
 - *types of metadata functions*
 - *and types of metadata elements*

Metadata typology and uses - which types?

- common metadata usage is *Dublin Core*
 - *often perceived as a catch-all solution...*
- many different options available for metadata, e.g.
 - *metadata for preservation*
 - *metadata for intellectual property rights*
 - *METS*
 - *metadata for geospatial information*
 - *AACR*
 - *metadata for describing music*
 - *FRBR, LOM, RDA, RDF...*
- often use a mixture of models, domains, uses...
 - *disparate parts forming our understanding of metadata*

Metadata typology and uses - typify

- need *criteria* to typify something
 - *and potentially everything in a given domain...*
- simple term *metadata* and simple definition, *data about data*
 - *encompass a complex technical and intellectual infrastructure*
 - *used to manage and retrieve digital objects in different digital contexts*
 - *within different digital information systems and services*
- criteria often matches sheer diversity of digital information management systems
- could be applied to different metadata levels, e.g.
 - *records, agents, elements, schemas...*

Metadata typology and uses - classify metadata

- metadata may describe different attributes or properties of information objects
 - *giving them meaning, context and organisation in a standardised way*
- many uses and dimensions of metadata
 - *has now led to construction of a very broad typology*
- use obvious criteria to typify metadata
- might classify metadata, mixing different levels, e.g.
 - *systems, practices, models, schemas, elements, records and trends*
- mix different levels with seven criteria, e.g.
 - *criterion 1: way of creation...*
 - *criterion 2: moment of creation...*
 - *criterion 3: way of storage...*
 - *criterion 4: level of structure...*
 - *criterion 5: purpose of the metadata...*
 - domain independent metadata
 - domain dependent metadata
 - *criterion 6: application...*
 - i.e. what's metadata used for?
 - *criterion 7: level of standardisation*

Metadata typology and uses - types and systems

- often consider metadata types to help understand associated metadata systems
- need to distinguish different types of metadata
 - *to help categorise metadata*
- classification can mix different levels to categorise metadata, e.g.
 - *record level, element level or schema/element set level*
- we often consider the following
 - *general versus specialist*
 - *minimalist versus rich*
 - *hierarchical versus linear*
 - *structured versus unstructured*
 - *machine generated versus human authored*
 - *embedded versus detached*
- **surface** metadata
 - *information that can be gathered by machines and converted into metadata*
 - *process of gathering it is often known as **screen scraping***
- *other types of metadata*
 - *including keywords, Google, tags, user created metadata, &c.*
 - *self-description mechanisms for social tagging and new trends*
 - *avoid keyword stacking*
 - *Kartus, E. 2006.*

Metadata typology and uses - types and vocabularies

- purpose of any metadata, whether element or model
 - *describe a property or set of useful properties*
 - *for a given information resource or object*
- many current metadata systems include an inherent multidimensional nature
 - *such structures often called metadata vocabularies*
- a few useful definitions
 - *metadata models/formats help express properties of a resource, e.g. a subject*
 - *metadata schemas = metadata formats encoded in a standardised machine readable markup language*
 - *metadata scheme = a set of rules or terms for encoding the value of a particular metadata term*
 - *both schemas and schemes involve metadata elements*

Metadata typology and uses - formats and schemas

- in the 1990's, common to speak about *metadata formats* or *metadata models*
 - *to refer to a set of properties*
 - *expressed and defined in a standardised way*
 - *served to describe a digital information object*
- each description constituted a metadata record
 - *applying that metadata format to a particular object*
- grouped metadata formats along a continuum of growing structural and semantic richness
 - *used to identify shared characteristics of each grouping*
 - *patterns from paths of the given metadata formats*
 - *Dempsey, L. & Heery, R. 1998.*

Metadata typology and uses - schema/scheme

- DCMI Glossary definition for *schema/scheme*

any organization, coding, outline or plan of concepts. In terms of metadata, a systematic, orderly combination of elements or terms.

- when a declaration of metadata terms is represented in XML or RDF schema language
 - *it might be better considered strictly a schema*
- schemas are machine-processable specifications
 - *define structure and syntax of metadata specifications in a formal way*
- difficult to think of a single metadata element or a single metadata value
 - *in isolation, without broader context*
- current Semantic Web approach
 - *considers schemas and schemes*
 - *assumes they are formalised in a machine-understandable way*
- schemas are sets of metadata elements and rules
 - *previously stated for a particular purpose*
- schemes are a set of rules for encoding information
 - *supports a specific information domain*
- schemas and schemes are commonly expressed in XML/RDF vocabularies
 - *easier to read and use*
- also form a perceived complete metadata infrastructure

- *for current digital information services*
 - *becoming Metadata Vocabularies*
- metadata classification starting from metadata vocabularies
- *metadata schemas, metadata schemes, and metadata elements*

Metadata typology and uses - types of schemas

- current metadata schemas include
 - *not only the semantics of an entire element set*
 - *but also encoding of elements and structure with a markup language*
- two crucial criteria to take into account
 - *to categorise the complexity and variety of current metadata*
 - **granularity** and the **information domain**

Metadata typology and uses - types of schemas - granularity

- draw up a typology of metadata schemas
 - *based on hierarchical relations between different sets of metadata*
- extensible nature of metadata, e.g.
 - *global metadata schemas:*
 - e.g. Dublin Core, which is the best known example
 - *local metadata schemas:*
 - e.g. SEPIADES, Europeana Semantic Elements...
 - *container metadata schemas:*
 - container architectures
 - e.g. RDF, METS, ONIX, MARC, &c.
 - *conceptual metadata schemas*
 - e.g. CIDOC-CRM

Metadata typology and uses - types of schemas - domain

- an example of domain specific is *Web-based information*
 - *sectioned in a vertical way*
- there are global search engines, and very large digital libraries like Europeana
- increasingly, there are also different portals
 - *digital libraries*
 - *and other digital information systems and services*
- locations where the objects you can find belong to a category
 - *either subject-oriented or type-oriented*
- a few domain specific examples
 - *metadata for Cultural Heritage*
 - *metadata for geographic and geospatial information systems*
 - *metadata for educational information systems*
 - *metadata for digital preservation/curation*

Metadata typology and uses - types of schemas - domain examples

metadata for Cultural Heritage

- often applied to cultural objects and visual resources
- in this huge information context, many early developed metadata schemas
- schemas range from those influenced by the librarian's domain
 - *to those built from archives and museums*
 - *or arts and architecture domains...*
- for example,
 - *traditional cataloging standards converted into schemas - MARCXML, MODS...*
 - *metadata standards for Finding Aids - Encoding Archival Description (EAD)*
 - *representation of digital texts - Text Encoding Initiative (TEI)*
 - *metadata schemas/standards for digital visual arts - CDWA...*

Metadata typology and uses - types of schemas - domain examples

metadata for geographic and geospatial information systems

- metadata in this particular domain has a clear protagonist, and almost standard
- *Content Standard for Digital Geospatial Metadata (CSDGM)*
 - originally adopted for the Federal Geographic Data Committee in 1994
 - and then revised in 1998
- its main elements were embraced by the international community
- adopted through the ISO metadata standard *ISO 19115*

Metadata typology and uses - types of schemas - domain examples

metadata for educational information systems

- another big domain to analyse specific metadata schemas and standards
 - *the educational and learning environment*
- *Learning Objects Metadata (LOM)*
 - *developed by the IEEE Learning Technology Standards Committee*
 - *the main metadata schema in this field*
- learning objects used and reused for educational purposes
 - *a particular type of digital information object*
 - *requiring particular types of metadata schemas*
 - *helps describe their education-specific properties*
- *Learning Objects Metadata* has a more complicated structure than many other schemas
 - *integrates different kinds of metadata element*
 - *e.g. descriptive, administrative, and technical elements*

Metadata typology and uses - types of schemas - domain examples

metadata for digital preservation/curation

- digital preservation and curation is a broad domain and concept
- implies much more than metadata
 - *a broad range of systems and infrastructures*
 - *designed to preserve the usable life of digital objects*
- preservation metadata is a domain
 - *supports all processes associated with digital preservation*
- primary examples for preservation metadata as a specific domain
 - *Open Archival Information Systems (OAIS (ISO 14721))*
 - *PREMIS (Preservation Metadata: Implementation Strategies)*

Metadata typology and uses - types of schemas - domain

- domain approach to digital information
 - *increasingly manifest throughout digital information services*
- besides big metadata schemas based upon a specific domain
 - *there are many application specific profiles*
- these are data elements drawn from other metadata schemas
 - *then combined together*
 - *and used for more specific or even local application domains*
 - *e.g. for public sector information*
- also find application profiles based on one schema
 - *but tailored for a particular information community*
 - *e.g. Dublin Core Education Application Profile*

Metadata typology and uses - types of schemes

- schemes describe a particular way to encode *metainformation* describing a given resource
 - *similar to Anglo American Cataloging Rules (AACR), Cataloging Cultural Objects (CCO)...*
- schemes are also sets of terms or vocabularies
 - *only concern the possible values for a particular metadata element*
- *metadata schemes* are a range of values
 - *values that might be provided for an assertion about a resource*
 - *e.g. Date-time formats, authority lists, controlled vocabularies...*

Metadata typology and uses - types of schemes - examples

- *authority list* is a metadata scheme
 - *could be applied to encode the values of a metadata element*
 - *or term in a metadata schema*
- e.g. if we're dealing with a resource's authorship
 - *the element `DC.Creator` in Dublin Core*
 - *or `Author` in TEI Header*
 - ...
- large-scale metadata content standards such as *Cataloging Cultural Objects (CCO)*
 - *include general instructions applicable to a metadata model or a set of its elements*
- metadata schemes are specific vocabularies
 - *devoted to the values of a particular metadata element*
- encoding schemes provide contextual information
 - *or parsing rules that aid in the interpretation of an element value*
- such contextual information may take different forms, e.g.
 - *controlled vocabularies, formal notations, or parsing rules*
- schemes might also be called *value spaces*
 - *forming the set of possible values for a given type of data*

Metadata typology and uses - types of schemes - domain

- as with schemas, schemes may be classified as well
- *general purpose schemes*
 - e.g. *universal classifications or generic subject heading lists...*
- *specific purpose schemes*
 - e.g. *traditional metadata schemes*
 - *thesauri, other vocabularies applied to a specific information domain...*
- many, many vocabularies encoded as metadata schemes...
 - e.g. *metadata vocabularies for cultural heritage*
 - *Union List of Artist Names (ULAN)*
 - *Arts and Architecture Thesaurus (AAT)*
 - ...
- schemes could be
 - *traditional thesauri*
 - *classification schemes*
 - *and other knowledge organisation systems...*
- schemes could also be
 - *subject-based metadata like ontologies and folksonomies*
 - *traditional vocabularies encoded for the Semantic Web*
 - e.g. *following a formal schema like Simple Knowledge Organization System (SKOS)*

Metadata typology and uses - types of schemes - ontologies

- most of the ontologies we find, for example, are also domain oriented
- they divide the realm of knowledge that they represent into:
 - *individuals*
 - *classes*
 - *attributes*
 - *relations*
 - *and events*

Metadata typology and uses - types of schemes - folksonomies

- semantically weaker *folksonomies* does not usually have a domain orientation
 - *they do not have any kind of control*
- a *folksonomy* is a record with labels, tags or keywords
 - *used by many people on the Web*
 - *usually without a particular purpose or initial structure*
- a *folksonomy* will be more general
 - *dynamic in its growth and application*
- some experiences using social tagging within a specific domain
 - *e.g. Steve Museum - a systematic research project*
 - *considers how social tagging can best serve the museum community and its visitors*
 - *Steve Museum Project*
- in a distributed networked environment - rapid scheme changes
 - *managing semantics of changes vital to functioning and utility of schemes*

Metadata typology and uses - types of metadata - part I

- many traditional classifications of metadata types are based on metadata elements
 - *assuming that every metadata schema has elements of similar types*
- allows fair comparison with known points of comparison
- almost every metadata handbook distinguish the following types of metadata
 - **Descriptive** metadata elements
 - **Structural** metadata elements
 - **Administrative** metadata elements
- types and functions of metadata elements classified into broad categories
 - *descriptive, structural, and administrative*
 - *they do not have well-defined boundaries and often overlap*
- different types blend into one another when using a specific metadata schema

Metadata typology and uses - types of metadata - part 2

- in general, metadata schemas and standards include these types of metadata elements
- e.g., **METS** packages structural, descriptive, administrative, and other metadata
 - *with an information object or digital surrogate*
 - *indicates types of relationships*
 - *e.g amongst different parts of the current complex information object*
- so, any metadata schema can classify their elements as different types of metadata
- a traditional categorisation using this pattern
 - *Dublin Core elements division*
 - *distinguishes three groups for its metadata elements*
- *Dublin Core groupings*
 - *Content elements - Title, Subject, Description, Source, Language, Relation, and Coverage*
 - *Intellectual Property elements - Creator, Publisher, Contributor, Rights*
 - *Instantiation elements - Date, Type, Format, and Identifier*

Metadata typology and uses - shaping standards

- several factors
 - *such as the type of resource and the application domain*
- help shape and influence the formation of metadata standards
- as digital information systems and services are increasingly vertical
 - *e.g. subject-oriented or community-oriented*
 - *metadata matches this trend as well...*
- identifying and monitoring actual use of metadata - not a straightforward process

Metadata typology and uses - user needs - part I

Setting metadata loose, through the internet and their widening user base, has in some cases resulted in new user contexts for existing metadata. The biodiversity domain provides an excellent illustration of how metadata which previously might have had a strong local focus and user community can attract a wider interest on the internet. International collaborative efforts, such as the European Network for Biodiversity Information (ENBI), are currently building large metadata repositories by aggregating metadata of local institutions through portals. Existing metadata are thus repurposed in the context of the international research on global warming and its impact on biodiversity.

- Dempsey, L and R Heery. *Metadata: A current view of practice and issues*. Journal of Documentation, 54(2). PP.145–172. 1998.

Metadata typology and uses - user needs - part 2

- user needs regarding cultural heritage usually defined in general and vague terms
- museums, libraries, and archives have struggled to identify user needs
 - *lots of digitised material to access, not sure how it's being accessed...*
- user needs have attracted attention on a research level and within individual projects
- application of research outcomes and recommendations in the field remains problematic
- metadata practitioners rarely have chance to start from scratch
 - *issues with reviews, surveys, legacy requirements and standards...*
- metadata services rarely built from scratch...

Metadata typology and uses - use case evaluation

- no clear, standard set of practices or methodologies to monitor the use of metadata
- might monitor metadata using more traditional and resource intensive methods
 - *such as user surveys and interviews*
- there are a couple of examples for better monitoring usage of metadata
 - *dynamic search interfaces - last.fm, Facebook...*
 - *use case evaluation - collection management...*

Metadata typology and uses - use case evaluation - example

monitoring metadata usage with dynamic search interfaces

- marketers and HCI specialists often deduce user information
 - *regarding preferences of users*
 - *by analysing their behavior and actions*
 - *analyse through logfiles and tools such as heatmaps*
- Facebook and last.fm rely heavily on the analysis of use data
 - *helps deliver a more personalised service*
- last.fm tracks, for example, all of the music files played by the user
 - *allows them to fine-tune the user's profile*
 - *also allows them to offer extra features*
 - *features such as concert notification*
- Facebook is able to offer customised features, similar to last.fm

Metadata typology and uses - use case evaluation

- metadata fields are effectively used within a collection management database
- crucial to offer direct added value to users - encourages interaction with an application
- user is confronted with a default search interface
 - *may be intuitively configured to a user's individual needs*
 - *e.g. by adding, deleting, re-arranging metadata fields/elements...*
- possibility to monitor metadata fields usage within the database by a collections manager
- use recorded outcomes to manage relevance and usage of metadata fields
- provides statistical support for modifications and updates to metadata
- metadata providers need to start experimenting with tools and methodologies
 - *to allow them to monitor the effective use of the metadata*

References

- Content Standard for Digital Geospatial Metadata
- Dempsey, L and R Heery. *Metadata: A current view of practice and issues*. Journal of Documentation, 54(2). PP.145–172. 1998.
- Dublin Core
- Europeana Semantic Elements
- Garoufallou, E., Greenberg, J., *Metadata and Semantics Research: 7th International Conference*. MSTR. 2013.
- Kartus, E. *Types of metadata*. University of Melbourne. 2006. http://www.infodiv.unimelb.edu.au/metadata/add_info.html
- Learning Objects Metadata
- Open Archival Information Systems
- PREMIS
- Simple Knowledge Organization System
- Steve Museum Project
- XPath Version 1.0 functions
- XPath Version 2 functions