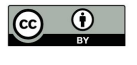# Data modelling

Helena Bermúdez Sabel

Goals

- **Understand the importance of modelling in Digital Humanities**
- Understand the importance of standards in the context of the FAIR principles
- Gain a basic understanding of linked open data
- Be able to encode a text in TEI

Model and modelling: definitions

According to McCarty (2005: 24):

- **Model**: either a representation of something for purposes of study, or a design for realising something new.
  - Simplified, idealised representation of real objects
- **Modeling**: the heuristic process of constructing and manipulating models.

Types of modelling

Classification in [Flanders and Jannidis (2015)](#):

- **Research-driven modelling**: to express specific research ideas in cases where data is being created to support the creator's own research needs
- **Curation-driven modelling**: to serve as an interchange format for some types of users and user communities where data is typically being created and modeled with someone else's needs in mind

Other classifications

- **Model of** vs. **model for** ([McCarty 2005](#))
- **Modeling for production** vs. **modeling for understanding** ([Eide 2014](#))

Why modelling?	Using a model	Creating a model
Modelling for production	Using TEI to encode and publish a document	Creating a module in the TEI Guidelines
Modelling for understanding	Using TEI encoding as a research method	Creating a module in the TEI Guidelines

The importance of modelling

- Modelling as the key methodological structure of Digital Humanities
- As computers depend on models to operate, within corpora development and exploitation, we need models for:
  - Automatisation of tasks during development
  - Hypothesis testing
  - Analysis

Modelling and digital editing ([Pierazzo 2015](#))

1. Organisation of entities such as texts, documents and works along with their relationships and how the were produced.
2. Definition of the type and purpose for the production of a new edition, its implied community of users and the features best represent their various needs.

Corpus methods

From data to theory: Wallis and Nelson (2011) and the 3A perspective:

- **Annotation** consists of the application of a scheme to texts. Annotations may include structural markup, part-of-speech tagging, parsing, and numerous other representations.
- **Abstraction** consists of the translation (mapping) of terms in the scheme to terms in a theoretically motivated model or dataset. Abstraction typically includes linguist-directed search but may include e.g., rule-learning for parsers.
- **Analysis** consists of statistically probing, manipulating and generalising from the dataset. Analysis might include statistical evaluations, optimisation of rule-bases or knowledge discovery methods.

Goals

- Understand the importance of modelling in Digital Humanities
- **Understand the importance of standards in the context of the FAIR principles**
- Gain a basic understanding of linked open data
- Be able to encode a text in TEI

The importance of standards

- Data standardisation vs data normalisation
- Standard: codified rules and and guidelines for the creation, description and management of digital resources ([Gill and Miller, 2002](#))
- Advantages:
  - Reusability
  - Interoperability
  - Long-term preservation (format/schema update)

FAIR principles

Findable

1. (Meta)data are assigned a globally unique and persistent identifier
2. Data are described with rich metadata
3. Metadata clearly and explicitly include the identifier of the data they describe
4. (Meta)data are registered or indexed in a searchable resource

FAIR principles

Accessible

1. (Meta)data are retrievable by their identifier using a standardised communications protocol
   a. The protocol is open, free, and universally implementable
   b. The protocol allows for an authentication and authorisation procedure, where necessary
2. Metadata are accessible, even when the data are no longer available

FAIR principles

Interoperable

1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
2. (Meta)data use vocabularies that follow FAIR principles
3. (Meta)data include qualified references to other (meta)data

FAIR principles

Reusable

1. (Meta)data are richly described with a plurality of accurate and relevant attributes
   a. (Meta)data are released with a clear and accessible data usage license
   b. (Meta)data are associated with detailed provenance
   c. (Meta)data meet domain-relevant community standards

Which standard?

- Extremely domain-specific and conditioned by the research questions
- Tips:
  - Instead of querying the Standards Developing Organisations (SDO) websites, like ISO, look at what your peers are doing
  - Look for papers/reports aiming at clarifying the state-of-the-art in your field. E.g.:
    - Stührenberg, Maik, Antonina Werthmann, and Andreas Witt. 2012. "Guidance through the Standards Jungle for Linguistic Resources". In Proceedings of the LREC-12 Workshop on Collaborative Resource Development and Delivery. Istanbul, Turkey, May 2012, 47-50. European Language Resources Association (ELRA). https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/4494.

Goals

- Understand the importance of modelling in Digital Humanities
- Understand the importance of standards in the context of the FAIR principles
- **Gain a basic understanding of linked open data**
- Be able to encode a text in TEI

Linked Data

- An approach to structure data representation that manifests as a Web of Data
- It enables:
  - Transparent access to data
  - Data integration across disparate data silos
  - Platform independent data connectivity

Linked Open Data

- A huge database (worldwide) on the Web of Data
- Since the data is open, the community can rely also on other sources of data
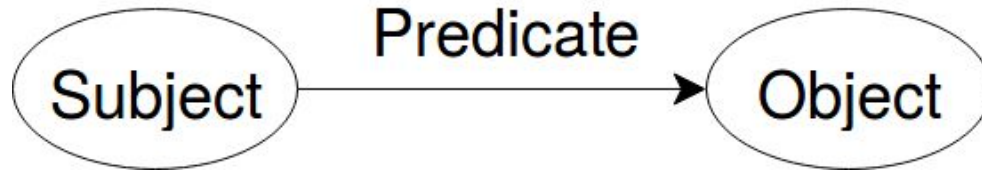

https://lod-cloud.net/
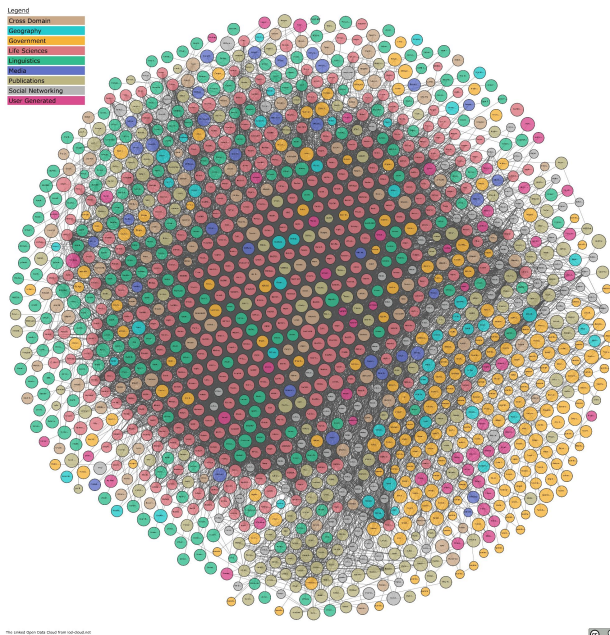
Web of Documents vs Web of Data

- Web of Documents: the Web we use everyday
- Web of Data: no Documents, only DATA (links between data)
- Web of Data and Web of Documents live together in parallel

RDF

- The Resource Description Framework (RDF) is a general framework for representing interconnected data on the web
- RDF data model:

RDF

- It is a data model
- It facilitates data merging and data exchange
- It extends the linking structure of the Web to use URIs to name both the relationship between things and the nodes

Example in TURTLE

```turtle
@prefix dc: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix schema: <http://schema.org/> .

<http://www.viaf.org/viaf/190446859> a dc:Text,
    schema:CreativeWork;
    dc:title "(Estudio)",
    "Rumbo",
    "Sin rumbo";
    schema:creator <http://www.viaf.org/viaf/24681042>;
    schema:version
<https://raw.githubusercontent.com/cligs/textbox/master/spanish/novela-hispanoamericana/tei/nh0011.xml> .

<http://www.viaf.org/viaf/24681042> a foaf:Person;
    foaf:name "Cambaceres",
    "Cambaceres, Eugenio" .
```
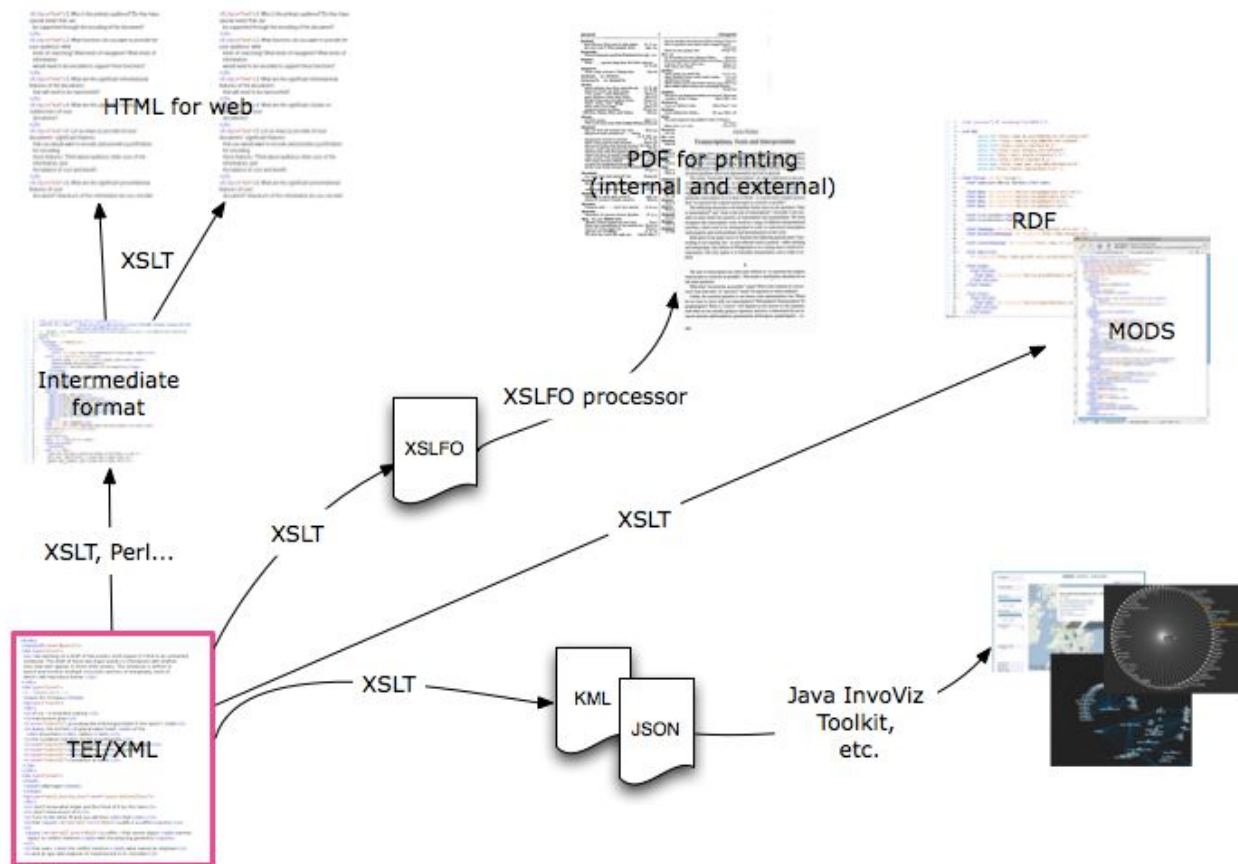
Goals

- Understand the importance of modelling in Digital Humanities
- Understand the importance of standards in the context of the FAIR principles
- Gain a basic understanding of linked open data
- **Be able to encode a text in TEI**

Brief recap on XML

- Modelling texts as **ordered hierarchies of content objects** ([Renear et al. 1993](#))
  - Limitation: coexistence of multiple logical hierarchies (the problem of overlapping elements)
  - Modelling advantages:
    - Analytical practises frequently determine hierarchies of objects (or decomposition into hierarchies is often possible)
    - Human and machine readable
    - Easy and simple syntax

Brief recap on XML

- Technical advantages of XML:
    - Control over input and output
    - Software and hardware independent
    - Supported by a wide range of software (open + proprietary)

HTML for web

PDF for printing
(internal and external)

RDF

MODS

XSLT

Intermediate
format

XSLFO processor

XSLFO

XSLT, Perl...

XSLT

XSLT

TEI/XML

XSLT

KML

JSON

Java InvoViz
Toolkit,
etc.

Flanders (2018)

```xml
<front>
    <div>
        <listWit>
            <witness corresp="#A">Cancioneiro da Ajuda</witness>
            <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
        </listWit>
    </div>
</front>
<body>
    <div type="poem" corresp="#A82B186b">
        <head>
            <title>
                <app>
                    <rdg wit="#A" hand="#a">
                        <idno>A82</idno>
                        <locus from="21r" to="21r"/></rdg>
                    <rdg wit="#B" hand="#d">
                        <idno>B186</idno>
                        <locus from="48r" to="48v"/></rdg>
                </app>
            </title>
            <name role="author" ref="#PGarBu"/>
        </head>
        <lg type="cobra">
            <l n="1">
                <app>
                    <rdg wit="#A #B">De</rdg>
                </app>
                <app>
                    <rdg wit="#A">quantos</rdg>
                    <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">mui</rdg>
                    <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">coitados</rdg>
                    <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
                </app>
                <app>
                    <rdg wit="#A #B">son</rdg>
                </app>
            </l>
            <l n="2">
                <app>
                    <rdg wit="#A #B">a que</rdg>
```

```xml
<front>
    <div>
        <listWit>
            <witness corresp="#A">Cancioneiro da Ajuda</witness>
            <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
        </listWit>
    </div>
</front>
<body>
    <div type="poem" corresp="#A82B186b">
        <head>
            <title>
                <app>
                    <rdg wit="#A" hand="#α">
                        <idno>A82</idno>
                        <locus from="21r" to="21r"/></rdg>
                    <rdg wit="#B" hand="#d">
                        <idno>B186</idno>
                        <locus from="48r" to="48v"/></rdg>
                </app>
            </title>
            <name role="author" ref="#PGarBu"/>
        </head>
        <lg type="cobra">
            <l n="1">
                <app>
                    <rdg wit="#A #B">De</rdg>
                </app>
                <app>
                    <rdg wit="#A">quantos</rdg>
                    <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>ǫ</am><ex>os</ex></rdg>
                </app>
                <app>
                    <rdg wit="#A">mui</rdg>
                    <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">coitados</rdg>
                    <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rd
                </app>
                <app>
                    <rdg wit="#A #B">son</rdg>
                </app>
            </l>
            <l n="2">
                <app>
                    <rdg wit="#A #B">a que</rdg>
```

Testemunho: A	Testemunho: B
Texto: A2	Texto: B92
Localização: 1r	Localização: 24v
Copista: α	Copista: c

1. Sennor fremosa grand enueia ei	1. Senhor fremosa muy grãd enueia ey
2. eu a tod ome que ueio morrer	2. eu a tod ome que ueio morrer
3. e segud ora o meu coñoçer	3. e segund ora o meu conhocer
4. enquant este faço mui gran razon	4. enquant est faço muy gram razon
5. ca ei por uos e no meu coraçon	5. ca ey por uos e no meu coraçon
6. tan gran cuita que mil uezes meten	6. tan gram coyta que mil uezes etem
7. sennor sen falla et sen tod o sen	7. senhor sen fala essem tod o sem
8. e nõ uus queredes de min doer	8. e non uǫ queredes d mj doer
9. Pero sēnor ūa ren uǫ direi	9. Pero senhor hūa rē uǫ direy
10. con tod est ora nõ ei eu poder	10. cõ tod est ora nõ ey eu poder

```xml
<front>
    <div>
        <listWit>
            <witness corresp="#A">Cancioneiro da Ajuda</witness>
            <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
        </listWit>
    </div>
</front>
<body>
    <div type="poem" corresp="#A82B186b">
        <head>
            <title>
                <app>
                    <rdg wit="#A" hand="#a">
                        <idno>A82</idno>
                        <locus from="21r" to="21r"/></rdg>
                    <rdg wit="#B" hand="#d">
                        <idno>B186</idno>
                        <locus from="48r" to="48v"/></rdg>
                </app>
            </title>
            <name role="author" ref="#PGarBu"/>
        </head>
        <lg type="cobra">
            <l n="1">
                <app>
                    <rdg wit="#A #B">De</rdg>
                </app>
                <app>
                    <rdg wit="#A">quantos</rdg>
                    <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>ǫ</am><ex>os</ex></seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">mui</rdg>
                    <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">coitados</rdg>
                    <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
                </app>
                <app>
                    <rdg wit="#A #B">son</rdg>
                </app>
            </l>
            <l n="2">
                <app>
                    <rdg wit="#A #B">a que</rdg>
```

Cantiga: A224, B39S, VS

Autor: Afonso Lopez de Baian

Período: 1246-1280 (período 3)

1.
A	Sennor que grau oi a mi e
B	Senhor que grau oi a mj e
V	Senhor que grau oi a mi e

2.
A	de me auer de uos a partir
B	de m auer de nos a partir
V	de m auer de uos a partir

3.
A	ca sei de pran pois m eu partir
B	ca sey de pram poys m eu partir
V	ca sey de pram poys m eu partir

4.
A	que mi a uerra per bõa fe
B	que mh a ueira per boa fe
V	que mh a uerra per boa fe

```xml
<front>
   <div>
      <listWit>
         <witness corresp="#A">Cancioneiro da Ajuda</witness>
         <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
      </listWit>
   </div>
</front>
<body>
   <div type="poem" corresp="#A82B186b">
      <head>
         <title>
            <app>
               <rdg wit="#A" hand="#α">
                  <idno>A82</idno>
                  <locus from="21r" to="21r"/></rdg>
               <rdg wit="#B" hand="#d">
                  <idno>B186</idno>
                  <locus from="48r" to="48v"/></rdg>
            </app>
         </title>
         <name role="author" ref="#PGarBu"/>
      </head>
      <lg type="cobra">
         <l n="1">
            <app>
               <rdg wit="#A #B">De</rdg>
            </app>
            <app>
               <rdg wit="#A">quantos</rdg>
               <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></seg></
            </app>
            <app>
               <rdg wit="#A">mui</rdg>
               <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
            </app>
            <app>
               <rdg wit="#A">coitados</rdg>
               <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
            </app>
            <app>
               <rdg wit="#A #B">son</rdg>
            </app>
         </l>
         <l n="2">
            <app>
               <rdg wit="#A #B">a que</rdg>
```


Nasalização progressiva

Testemunho

Ms. B — c, d, e, a, b (49.25%)
Ms. A — α (47.76%)
Ms. V — v (10.45%)

Período

Per. 3 (79.1%)
Per. 2 (20.9%)

```xml
<front>
    <div>
        <listWit>
            <witness corresp="#A">Cancioneiro da Ajuda</witness>
            <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
        </listWit>
    </div>
</front>
<body>
    <div type="poem" corresp="#A82B186b">
        <head>
            <title>
                <app>
                    <rdg wit="#A" hand="#α">
                        <idno>A82</idno>
                        <locus from="21r" to="21r"/></rdg>
                    <rdg wit="#B" hand="#d">
                        <idno>B186</idno>
                        <locus from="48r" to="48v"/></rdg>
                </app>
            </title>
            <name role="author" ref="#PGarBu"/>
        </head>
        <lg type="cobra">
            <l n="1">
                <app>
                    <rdg wit="#A #B">De</rdg>
                </app>
                <app>
                    <rdg wit="#A">quantos</rdg>
                    <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>ǫ</am><ex>os</ex></seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">mui</rdg>
                    <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">coitados</rdg>
                    <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
                </app>
                <app>
                    <rdg wit="#A #B">son</rdg>
                </app>
            </l>
            <l n="2">
                <app>
                    <rdg wit="#A #B">a que</rdg>
```
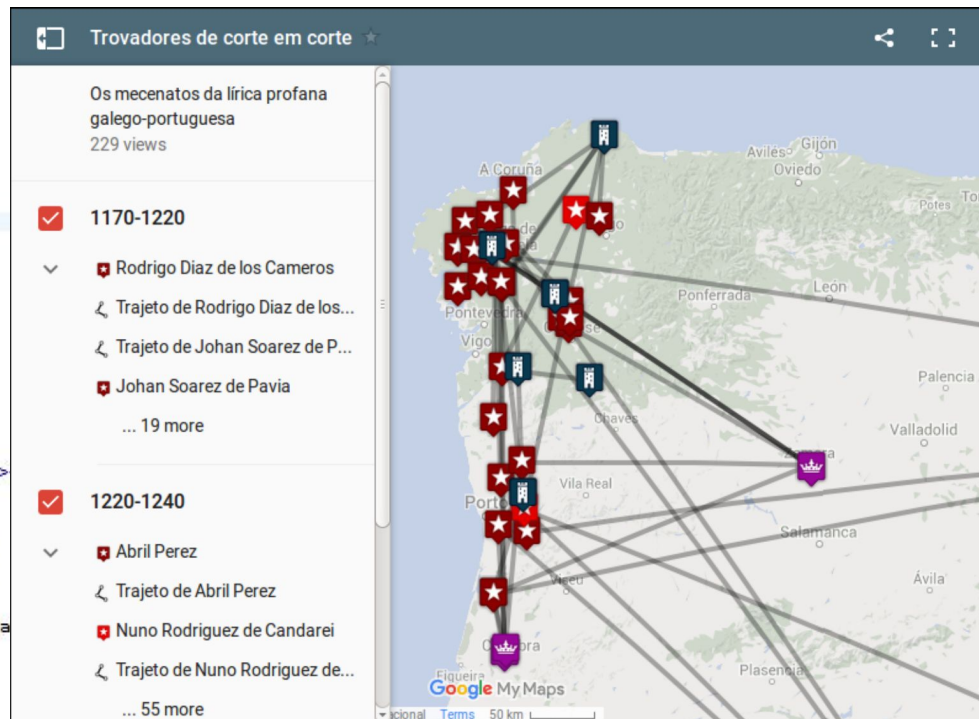
Fenómeno	Testemunho	Localização	Copista	Autor	Período	Cantiga	V.
quix] q̇s	A	11v-12r	α	Martin Soarez	1230-1270 (3)	A48, B160	18
quige] q̇s	A	22v-23r	α	Pero Garcia de Burgalês	1240-1270 (3)	A89, B193	13

```xml
<front>
    <div>
        <listWit>
            <witness corresp="#A">Cancioneiro da Ajuda</witness>
            <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
        </listWit>
    </div>
</front>
<body>
    <div type="poem" corresp="#A82B186b">
        <head>
            <title>
                <app>
                    <rdg wit="#A" hand="#a">
                        <idno>A82</idno>
                        <locus from="21r" to="21r"/></rdg>
                    <rdg wit="#B" hand="#d">
                        <idno>B186</idno>
                        <locus from="48r" to="48v"/></rdg>
                </app>
            </title>
            <name role="author" ref="#PGarBu"/>
        </head>
        <lg type="cobra">
            <l n="1">
                <app>
                    <rdg wit="#A #B">De</rdg>
                </app>
                <app>
                    <rdg wit="#A">quantos</rdg>
                    <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>o</am></seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">mui</rdg>
                    <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
                </app>
                <app>
                    <rdg wit="#A">coitados</rdg>
                    <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>ta
                </app>
                <app>
                    <rdg wit="#A #B">son</rdg>
                </app>
            </l>
            <l n="2">
                <app>
                    <rdg wit="#A #B">a que</rdg>
                </app>
```
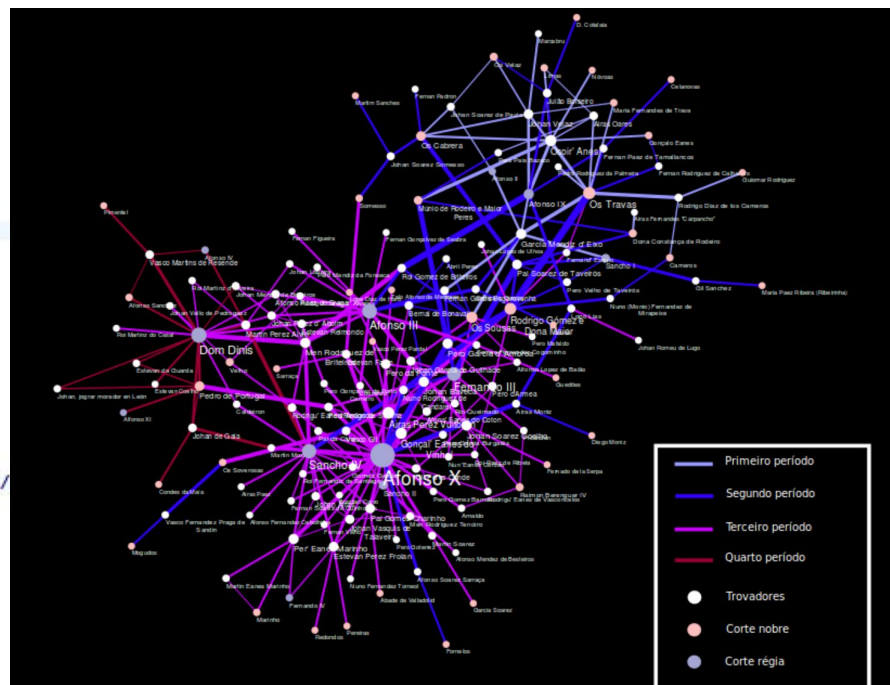
Trovadores de corte em corte

Os mecenatos da lírica profana galego-portuguesa
229 views

☑ **1170-1220**

- ⚑ Rodrigo Diaz de los Cameros
- ⚲ Trajeto de Rodrigo Diaz de los...
- ⚲ Trajeto de Johan Soarez de P...
- ⚑ Johan Soarez de Pavia
- ... 19 more

☑ **1220-1240**

- ⚑ Abril Perez
- ⚲ Trajeto de Abril Perez
- ⚑ Nuno Rodriguez de Candarei
- ⚲ Trajeto de Nuno Rodriguez de...
- ... 55 more

Google My Maps

```xml
<front>
  <div>
    <listWit>
      <witness corresp="#A">Cancioneiro da Ajuda</witness>
      <witness corresp="#B">Cancioneiro da Biblioteca Nacional</witness>
    </listWit>
  </div>
</front>
<body>
  <div type="poem" corresp="#A82B186b">
    <head>
      <title>
        <app>
          <rdg wit="#A" hand="#a">
            <idno>A82</idno>
            <locus from="21r" to="21r"/></rdg>
          <rdg wit="#B" hand="#d">
            <idno>B186</idno>
            <locus from="48r" to="48v"/></rdg>
        </app>
      </title>
      <name role="author" ref="#PGarBu"/>
    </head>
    <lg type="cobra">
      <l n="1">
        <app>
          <rdg wit="#A #B">De</rdg>
        </app>
        <app>
          <rdg wit="#A">quantos</rdg>
          <rdg wit="#B" ana="#abb">quant<seg corresp="#abb"><am>q</am><ex>os</ex></
        </app>
        <app>
          <rdg wit="#A">mui</rdg>
          <rdg wit="#B" ana="#uj">mu<seg corresp="#uj">j</seg></rdg>
        </app>
        <app>
          <rdg wit="#A">coitados</rdg>
          <rdg wit="#B" ana="#y-dip">co<seg corresp="#y-dip">y</seg>tados</rdg>
        </app>
        <app>
          <rdg wit="#A #B">son</rdg>
        </app>
      </l>
      <l n="2">
        <app>
          <rdg wit="#A #B">a que</rdg>
```


Legend: Primeiro período; Segundo período; Terceiro período; Quarto período; Trovadores; Corte nobre; Corte régia

Valid vs. well-formed

- Well-formed: compliant with XML rules
- Valid:
    - Correct use of the vocabulary: elements and attributes are available in that language
    - Correct use of grammar: elements are used in the proper place following the defined order

TEI as an XML standard

Some key advantages:

- Interoperability
- Exchange
- Technology transfer
- Efficiency

What is TEI

Text Encoding Initiative

- Set of guidelines to encode cultural documents
- International Consortium that maintains and develops said guidelines
- Community of projects and researchers that implement the TEI guidelines

Concepts   **XML** Syntax   **TEI** Language: vocabulary and grammar

paragraph footnote heading

```
<element>
    <element attribute="value">
        content
    </element>
</element>
```

```
<p>

<note type="foot">

<head>
```

Flanders (2018)

Modelling within TEI

- Definition of the vocabulary
- Formalisation of restrictions
- Design of the output

} TEI customisation

The TEI Guidelines

Text Body

https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html

The TEI Guidelines

Divided into two parts:

- Chapters (commonly known as "prose")
- Specs

\<s\>

\<s\> (s-unit) contains a sentence-like division of a text. [17.1 Linguistic Segment Categories 8.4.1 Segmentation]

Module	analysis — Simple Analytic Mechanisms
Attributes	att.global (@xml:id, @n, @xml:lang, @xml:base, @xml:space) (att.global.rendition (@rend, @style, @rendition)) (att.global.linking (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) (att.global.analytic (@ana)) (att.global.facs (@facs)) (att.global.change (@change)) (att.global.responsibility (@cert, @resp)) (att.global.source (@source)) att.segLike (@function) (att.metrical (@met, @real, @rhyme)) (att.datcat (@datcat, @valueDatcat, @targetDatcat)) (att.fragmentable (@part)) att.typed (@type, @subtype) att.notated (@notation)
Member of	model.segLike
Contained by	**analysis:** cl phr s **core:** abbr add addrLine author bibl biblScope citedRange corr date del distinct editor email emph expan foreign gloss head headItem headLabel hi item l label measure mentioned name note num orig p pubPlace publisher q quote rb ref reg rs rt said sic soCalled speaker stage street term textLang time title unclear unit **dictionaries:** case colloc def dictScrap entryFree etym form gen gram gramGrp hyph iType lang lbl mood number orth per pos pron re sense stress subc syll tns usg xr **drama:** actor camera caption castItem role roleDesc sound tech view **figures:** cell **header:** change distributor edition extent geoDecl handNote licence scriptNote **linking:** ab seg **msdescription:** accMat acquisition additions catchwords collation colophon condition custEvent decoNote explicit filiation finalRubric foliation heraldry incipit layout material musicNotation objectType origDate origPlace origin provenance rubric secFol signatures source stamp summary support surrogates typeNote watermark **namesdates:** addName affiliation birth bloc country death district education faith floruit forename genName gender geogFeat geogName nameLink nationality objectName occupation offset orgName persName persPronouns placeName region residence roleName settlement sex socecStatus surname **spoken:** u writing **tagdocs:** eg **textcrit:** lem rdg wit witDetail

Example of the specs of the element `<s>`

Let's see a TEI file

- Task 1: Go to the folder "exemplars" and open the file `tei_example.xml` which contains a novel by George Eliot

Valid vs. well-formed

- Well-formed: compliant with XML rules
- Valid:
  - Correct use of the vocabulary: elements and attributes are available in that language
  - Correct use of grammar: elements are used in the proper place following the defined order
  - Validity is enforced thanks to the formalisation of an schema in a schema languages

Schema languages

- W3C XML Schema (XML Schema or XSD)
- Document Type Definition (DTD)
- **REgular LAnguage for XML Next Generation (RELAx NG)**
  - XML syntax
  - compact syntax
- **Schematron**

TEI ODD: a TEI XML-conformant specification format that allows one to customize TEI P5 in a literate programming fashion

TEI ODD: One Document Does it All

- Select modules
- Delete unnecessary elements
- Add new elements and/or attributes
- Change the name of an element or attribute
- Constraint the values of an attribute
- Constraint the structure
- Manipulate functional groupings of elements
- Internationalisation (i-18)
- **Documentation**

Customisation

Customisation tool [Roma](#)

- Task 1: Create a schema with Roma:
  - Design an ODD for a corpus that includes linguistic annotation. Modify the description of the **@pos** attribute so only the values of the [Universal POS tagset](#) are allowed.
  - Create a new attribute "**syllab**" in the namespace: **http://www.eighteenthcenturypoetry.org/ns** whose prefix is **ecep**
  - Download a Relax NG schema
- Task 2: Associate this schema to the file in exemplars: **tei-annotated_example.xml**

To associate an schema

```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model
href="https://vault.tei-c.org/P5/current/xml/tei/custom/schema/relaxng/tei_all.rng"
schematypens="http://relaxng.org/ns/structure/1.0"
type="application/xml"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" />
```

Bibliography

- Eide, Øyvind. 2014. "Ontologies, Data Modeling, and TEI." Journal of the Text Encoding Initiative, Issue 8 (December). doi:10.4000/jtei.1191.
- Flanders, Julia, and Fotis Jannidis. 2015. "Knowledge Organization and Data Modeling in the Humanities." [Link].
- Gill, Tony, y Paul Miller. 2002. "Re-Inventing the Wheel? Standards, Interoperability and Digital Cultural Content". D-Lib Magazine 8 (1). doi:10.1045/january2002-gill.
- McCarty, Willard. 2005. Humanities Computing. Basingstoke: Palgrave Macmillan. doi:10.1057/9780230504219
- Wallis, S., and G. Nelson. 2001. "Knowledge discovery in grammatically analysed corpora". Data Mining and Knowledge Discovery, 5: 307–340.
- Pierazzo, Elena. 2015. Digital Scholarly Editing: Theories, Models and Methods. London ; New York: Routledge.