

Converting Asturian Notaries Public deeds to Linked Data using TEI and ShExML

Herminio García-González, Elena Albarrán-Fernández, José Emilio Labra-Gayo and Miguel Calleja-Puerta



Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

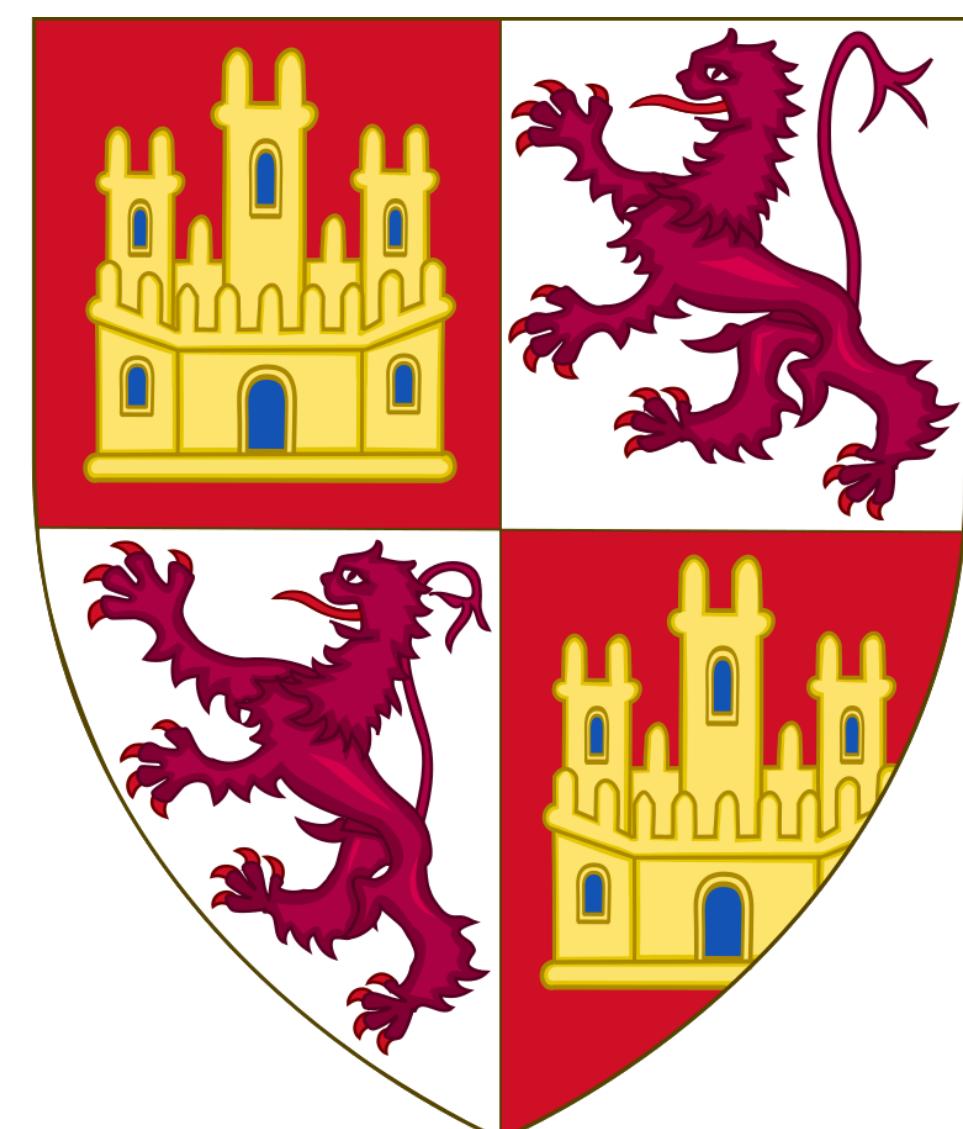


(Photo extracted from Wikimedia Commons)

Where is Asturias?



Asturias is a northern region in Spain and it was part of the Castilian crown.



King Alfonso X (1252-1284)

- New policy for Notaries
 - Written culture no Church exclusivity
 - Services to everyone
 - Defined by the law



Back to the present - Some particularities

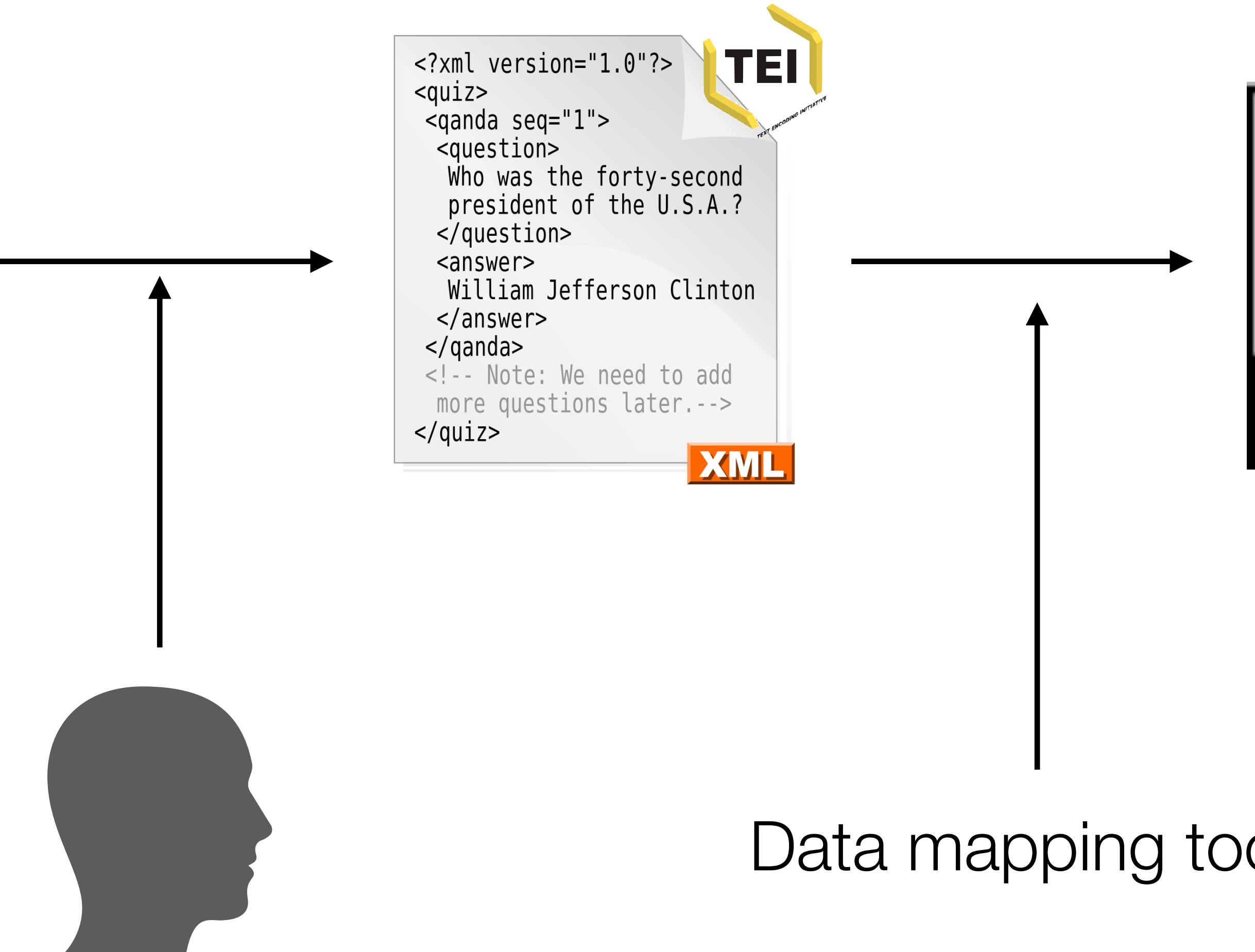
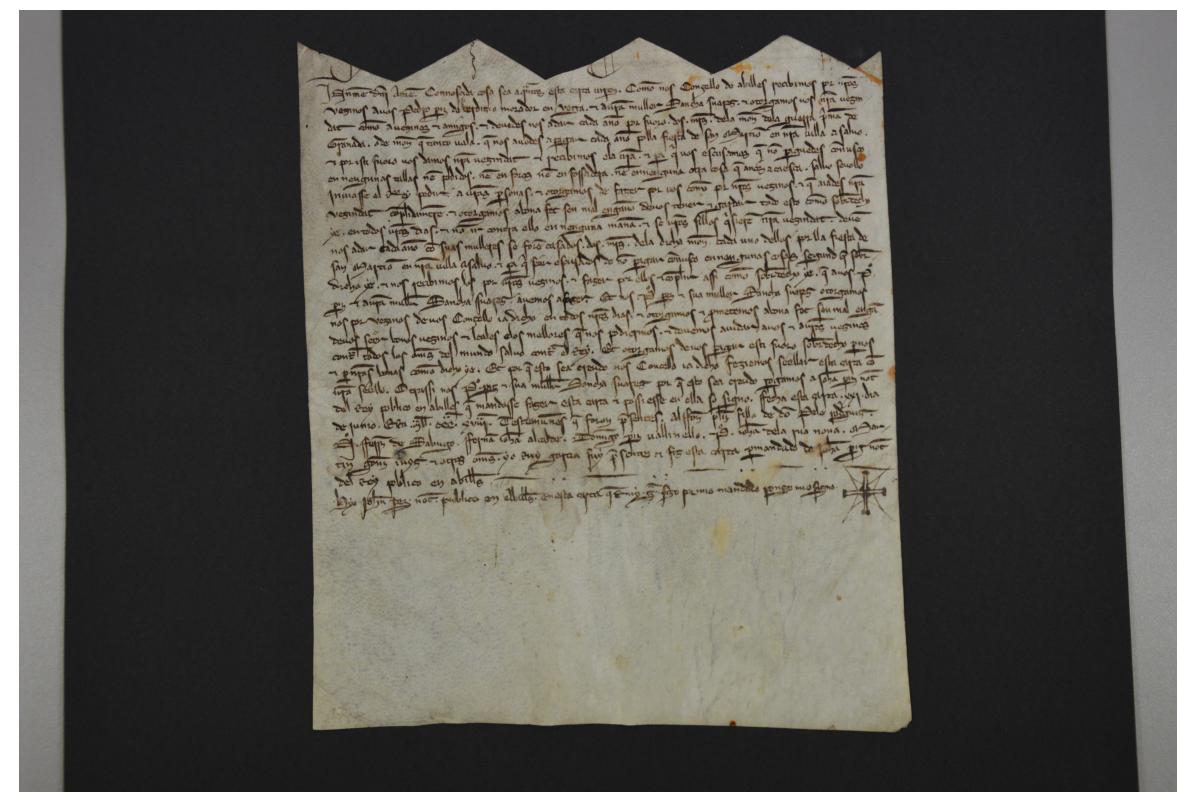
- No notarial registers remain
- Lot of manuscripts guarded by ecclesiastical institutions
- Some of them in public-state archives

Problems derived from this particularity

- Ecclesiastical institutions
 - No digitised versions available
 - A transcriber must go there
 - Limited opening hours
- Public-state archives
 - Digitised versions in some cases
 - Low quality in many cases
 - OCR to be proven

Corollary: A transcriber is a must for this material

Methodology



(XML icon extracted from Wikimedia Commons,
TEI logo from TEI website and
RDF logo from W3C)

Our TEI model

- Using TEI as the basis
- Need for diplomatic features
- Most similar extension CEI but not complete
- Addition of diplomatic elements
- More on [1]

Heterogeneous data mapping tools (the new wave)

- No need for *ad-hoc* solutions
- Same script no matter the data shape
- Easier for non-experts users
- Faster workflow

Why ShExML?

- Heterogeneous data mapping tool
- Simple syntax
- Created by two of the authors
- Can be tuned if necessary
- shexml.herokuapp.com

Script creation

- Using schema.org vocabulary (schema:ArchiveComponent, schema:ArchiveOrganization, schema:Collection, schema:Place, etc.)
- Nesting subtypes (linked shapes)
- Linking with other LOD cloud entities

Script creation (II)

```
78 ▼ :Manuscript :[manuscriptId] {  
79     a schema:ArchiveComponent ;  
80     schema:holdingArchive @:Archive ;  
81     schema:editor @:Person ;  
82     schema:licence [manuscripts.licence] ;  
83     schema:material @:Material ;  
84     schema:abstract [manuscripts.abstract] @es ;  
85     schema:dataCreated [manuscripts.origDate] xs:date ;  
86     schema:locationCreated @:LocationCreated ;  
87     schema:producer @:Office ;  
88     schema:text [manuscripts.text] @ast ;  
89     schema:citation @:Biblio ;  
90 }
```

Nesting

```
72 ▼ :Archive :[manuscripts.institution] {  
73     a schema:ArchiveOrganization ;  
74     schema:name [manuscripts.institution] ;  
75     schema:location @:ArchiveLocation ;  
76 }
```

```
105 ▼ :LocationCreated :[manuscripts.oficinaWhere] {  
106     a schema:Place ;  
107     schema:name [manuscripts.oficinaWhere] @es ;  
108     schema:sameAs wd:[manuscripts.oficinaWhere MATCHING city] ;  
109 }
```

Linking



Demo

shexml.herminiogarcia.com/editor

The case of a diplomatic study

```
<oficina type="notaria" subtype="regia" where="Aviles">
    <persName xml:id="JuanPerez" role="notario">Juan Pérez, notario público
    del rey Avilés</persName>
    <persName xml:id="RuyGarcia" role="amanuense">Ruy García fuy presente e
    fiz esta carta</persName>
</oficina>
```

```
<seg function="suscriptio" type="notaria"><seg type="amanuense">Yo, Ruy
García, fuy presente e fiz esta carta per mandado de Iohan Périz, notario
del rey público en Abillés.</seg> <seg type="notario">Hyo Iohan Périz,
notario público en Abillés, en esta carta que Ruy García hizo por mio
mandado pongo mio signo</seg></seg>.
```

There is a specific relationship which is present in the text but not well reflected here!!!

```
187 ▼ :Office :[officeId] {
    a schema:Organization ;
    schema:name [manuscripts.oficina] ;
    schema:location @:LocationCreated ;
    :person @:Notario ;
    :person @:Excusador ; Not a worker but a
    :person @:Amanuense ; special hierarchy!!!
    :person @:Registrador ;
}
196
197 ▼ :Notario :[notarioId] {
    a :[manuscripts.notarioType] ;
    :text [manuscripts.notario] ; Not a text but a
    _____ diplomatic clause!!!
}
201
202 ▼ :Amanuense :[amanuenseId] {
    a :[manuscripts.amanuenseType] ;
    :text [manuscripts.amanuense] ;
}
205 }
```

Charter Encoding Initiative

- Closest one
- Covers some of the notaries clauses (dispositio, datatio, invocatio, etc)
- Not an ontology (need for transitioning)
- Does not cover the previous issue

Specific ontology required!!!

Other limitations or challenges

```
<oficina type="notaria" subtype="regia" where="Aviles">
  <persName xml:id="JuanPerez" role="notario">Juan Pérez, notario público
  del rey Avilés</persName>
  <persName xml:id="RuyGarcia" role="amanuense">Ruy García fuy presente e
  fiz esta carta</persName>
</oficina>
```

Who is Juan Pérez?
Who is Ruy García?

These names and surnames can
appear for other people!!!

Possibility: Include more variables

- Place (mobility?)
- Co-workers (change of role?)

Need for a personalised entity disambiguation mechanism

Conclusions

- Notaries' manuscripts in Asturias
- Need for a transcriber
- Transcriptions in TEI
- Heterogeneous data mapping languages
- Limitations on existing vocabularies
- Further techniques (e.g., entity disambiguation)

The End