



Regesta Imperii with XML and other cool stuff

N. Bussas A. Güccük Y. Lewash S. Sert A. F. Silva M. Topalova Freie Universität Berlin

Web Data and Interoperability, 2016

### Outline



# Regesta Imperii

Coding da Vinci and Regesta Imperii The Data

#### The Work

Data Preprocessing Web Application Additional Processing

Live Demo



## Coding da Vinci

- Germany's first culture Hackathon
- organized by Deutsche Digitale Bibliothek
- goal: developing digital applications based on cultural data.



# Coding da Vinci

- Germany's first culture Hackathon
- organized by Deutsche Digitale Bibliothek
- goal: developing digital applications based on cultural data.

## Regesta Imperii

- chronologically records documents of
  - Roman-German kings/emperors (751–1519)
  - popes of early and High Middle Ages
- took part in Coding da Vinci 2015
  - to develop new applications/services from open data



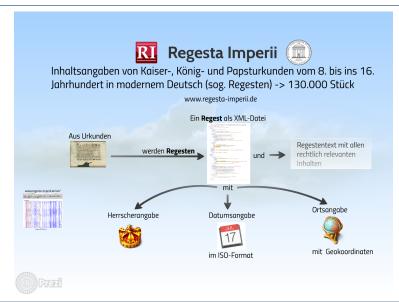
- 130,000 papal regesta
- each as XML file

# something like this:

```
<?xml version="1.0" encoding="UTF-8"?>
<cei xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="http://www.cei.lmu.de/schema/cei060122.xsd">
 <teiHeader>
   <fileDesc>
    <titleStmt>
      <title>Friedrich III. - [RI XIII] H. 14 n. 7</title>
    </titleStmt>
    <editionStmt>
      [RI XIII] H. 14 - Friedrich III., Nürnberg 1 (1440-1449)
      Regesta Imperii Online: <ref type="external" target="http://www.</pre>
        regesta-imperii.de/cei/013-014-000/sources/1440-05-16 1 0 13 14 0 7 7"></
        ref>
    </editionStmt>
    <publicationStmt>
      Deutsche Kommission für die Bearbeitung der Regesta Imperii e.V. bei der Akademie
          der Wissenschaften und der Literatur | Mainz
      <publisher>Akademie der Wissenschaften und der Literatur | Mainz — Digitale
```

Akademie</publisher>







All regesta XML files were transformed into a single XML file.

- using Python? Bash?
- some code?

The new XML file can be validated via XSD?

some code?



A web interface for searching the regesta was build.

- web standards + PHP
- possible to search for
  - title
  - date
  - place
  - İssuer
  - and any combination of them of each regesta
    - using XPath in SimpleXML
- output with additional info/links to resources

```
$xml = simplexml_load_file("index.xml");

// XML: st> <document> <title>... </title>... </document> 
$document = $xml->xpath("/list/document");

foreach ($document as $node) {
    if (stripos ($node->title, $_POST['TitleInput']) > -1) { $elem[] = $pos; }
    $pos += 1;
}
```



# Something for XQuery

- ► Foobar Foobar
- some code?

## Result?

something else



# Something for XQuery

- ► Foobar Foobar
- some code?

### Result?

something else



# Something about Web Services

- ► Foobar Foobar
- some code?

# DBpedia?

something else



### Microdata was used

to arrange results into pages for CSS

```
foreach ($elem as $num) {
    if ($pos % 5 == 0) { echo '<section id="', ($pos / 5), '">'; }
    // found regesta
    if ($pos % 5 == 4) { echo '</section>'; }
    $pos += 1;
}
```

to insert vCards with very sensitive personal data into the website

```
<address class="vcard">
  <span class="givenName">Numerius</span>
  <span class="familyName">Negidius</span><br/>
  <a class="familyName">Negidius</span><br/>
  <a class="email" href="mailto:spam@hotmail.com">n.nescio@nn.gov</a><br/>
  <span class="category">Management</span>,
  <span class="org">Freie Universität Bärlin</span>
</address></address>
```



Let's have a look...



- ► The first main message of your talk in one or two lines.
- ► The second main message of your talk in one or two lines.
- ▶ Perhaps a third message, but not more than that.

- Outlook
  - Something you haven't solved.
  - Something else you haven't solved.