From Document to Data:
Prosopography and Topography in the Tax Rolls of Medieval Paris

Nathan A. Daniels

JOHNS HOPKINS UNIVERSITY

INTRODUCTION

- A TEI-based Digital Edition of the tax rolls levied by King Philip IV on the city of Paris between 1292-1313
- Data includes:
  People; Occupations; Names; Taxes (financial); Relationships (familial, personal, occupational); Geography & Topography (streets, landmarks, buildings; cities, regions, countries); Institutions
- Preserve text of manuscripts, but extract data for tabulating, mapping, exporting, and Linked Open Data

TRANSCRIPTION & MARKUP

- Semi-Diplomatic transcription notes and marks up abbreviations, additions & erasures, and marginalia
- But focus is on historical data (esp. named entities)
- People / places all have unique XML Identifiers
  - References are disambiguated by these IDs
    e.g. "are these two *Jehans* the same person?"
- Occupations, offices, names, goods, payments, etc. are normalized in the *metadata* for extraction and analysis

- Example:

  Bertaut de compiegne poullaillier   ii.s.vi.d p.

```
<item>
    <persName ref="#BC01">
        <forename>Bertaut</forename>
        <nameLink>de</nameLink>
        <surname>compiegne</surname>
    </persName>
    <roleName ref="#BC01" role="poulailler"
        type="occupation">poullaillier</roleName>.
    <measureGrp>
        <measure quantity="2" unit="sous">ii.s.
            </measure>
        <measure quantity="6" unit="deniers">vi.d
            </measure>
    </measureGrp>
    <seg type="status">p.</seg>
</item>
```
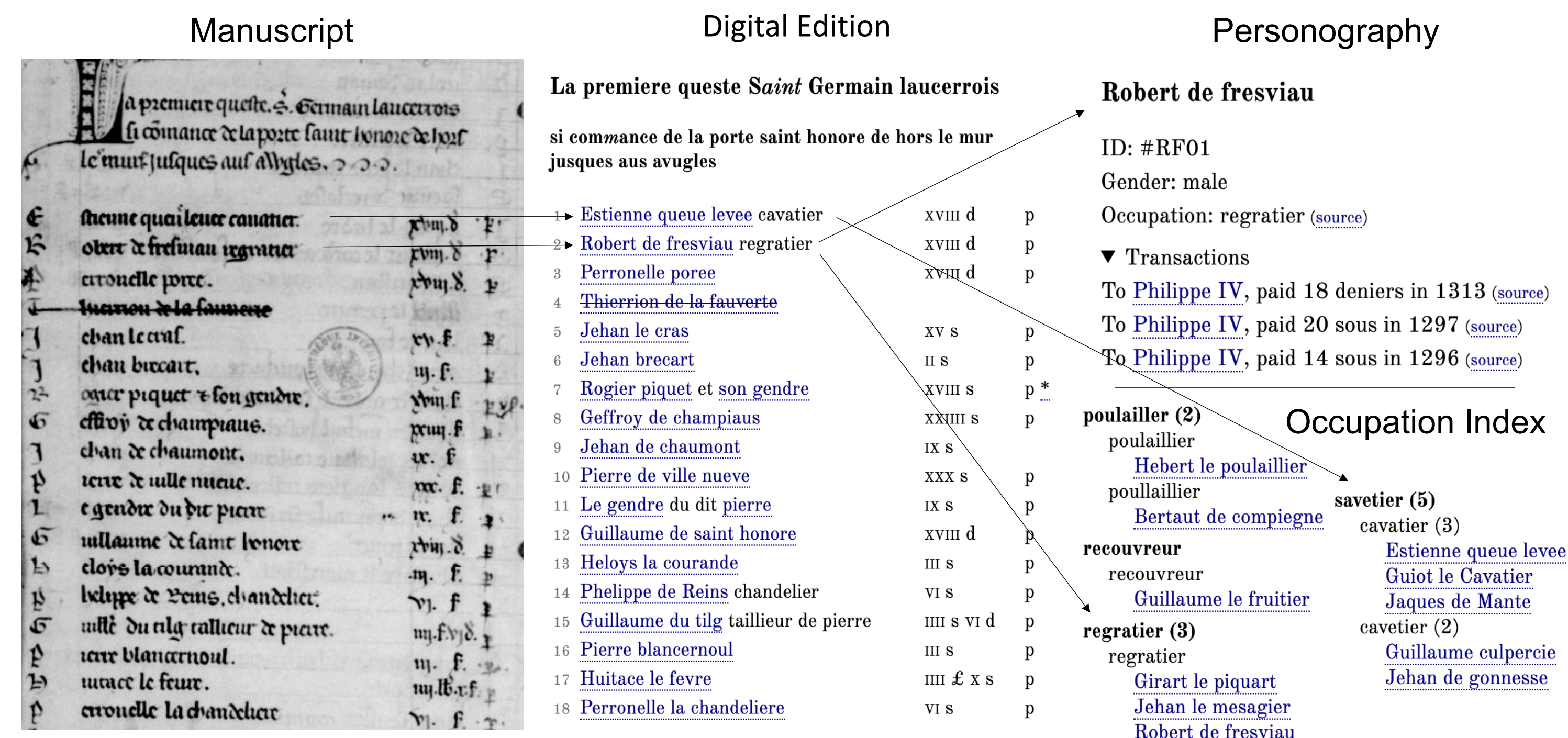
✉ nathan.daniels@jhu.edu   🐦 @nathanadaniels

QUESTION
How do we get from document to data?

Manuscript   Digital Edition   Personography
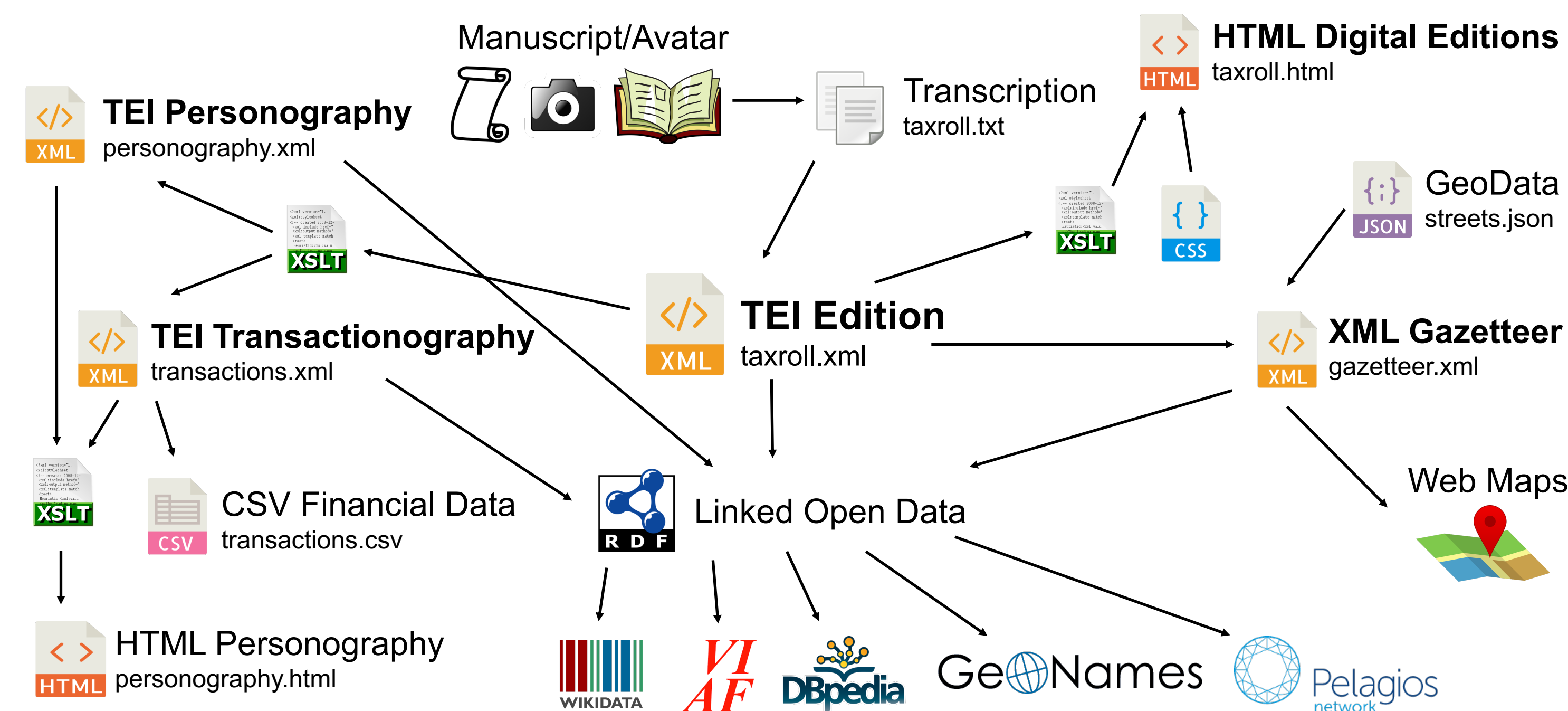

WORKFLOW

TRANSACTION MARKUP

- Bookkeeping Ontology by Digital Edition Publishing Cooperative for Historical Accounts (DEPCHA) (bit.ly/30Wr9bd)
- Uses @ana to provide markup for RDF

```
<rs ana="#bk:to">Le roi de France</rs>
<date ana="#bk:when">1313</date>
...
<item ana="#bk:entry">
    <persName ana="#bk:from">
        Perronelle poree</persName>
    <measure ana="#bk:money" quantity="2"
        unit="sous">ii.s.</measure>
    <seg ana="#bk:status">p.</seg>
</item>
```

- Result: From *Perronelle poree* to *Philip IV*, paid *2 sous* in *1313*

FUNCTION & SUSTAINABILITY

- Better to use flat files or databases?
- Open source or proprietary software?

- Complex relational databases and Content Management Systems (CMS) can provide additional functionality for searching, exporting, Linked Open Data, and API Management. Proprietary software packages can also be easier to use.

  BUT...

- They often require dedicated developers to set up, and need significant maintenance over the long term (updates, security, etc). Open file formats are more likely to last and keep working.
- Currently using: XML/TEI; XSLT; HTML/CSS; JavaScript/jQuery; JSON/GeoJSON

taxrolls.github.io