

歴史研究におけるRDFの活用推進に向けて

小川潤 (ROIS-DS人文学オープンデータ共同利用センター・特任研究員)

どんな研究？

- テキストや画像、3Dといった多様な歴史資料、そこに含まれる歴史情報をRDFデータとして構造化する手法を検討
- 情報学に精通していなくてもデータ構築に携わることができるような環境・システムの開発
- RDF, Linked Data関連技術を用いて、具体的な歴史研究の課題に取り組む

何がわかる？

- 多様な歴史知識をLinked Dataとして整備し、機械的な検索・解析・可視化を可能にする
- 歴史研究者を巻き込んだ広範な協働に基づく大規模なデータ構築環境の実現
- Linked Dataに基づく具体的な歴史研究成果の発信、歴史学との協働による学際的かつデータ駆動型の歴史研究の実践

研究内容

歴史情報の構造化

歴史マイクロナレッジ

マイクロナレッジとは？
資料における歴史的な出来事、関係性、状況、環境への「言及」をデータの基本単位として歴史知識をRDFで構造化するためのモデル

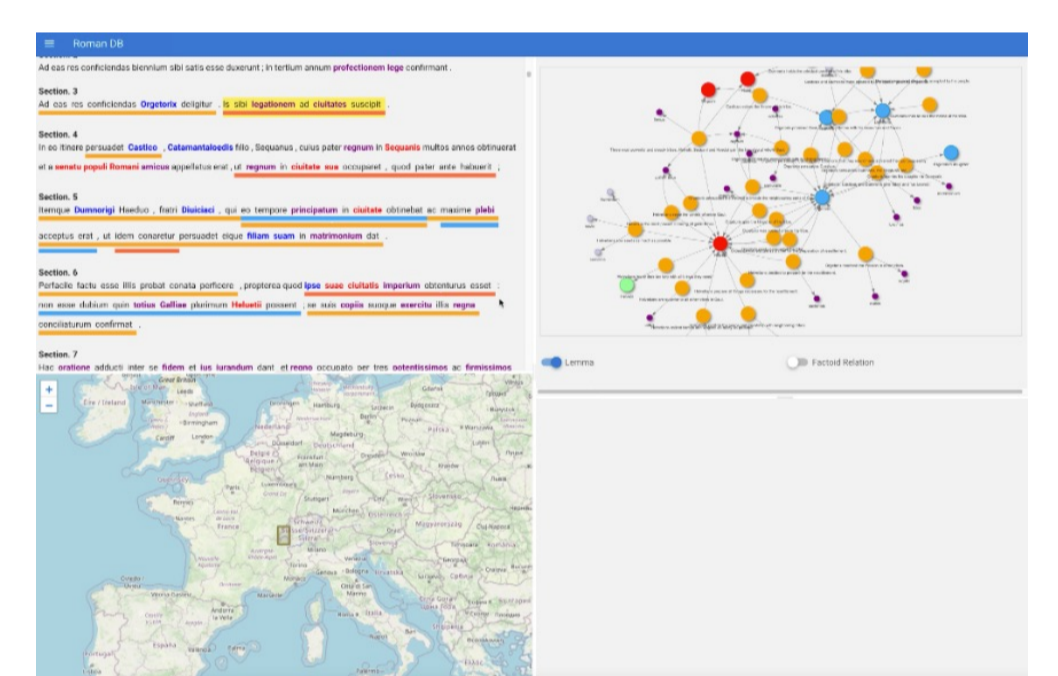
特徴は？

- 資料の具体的な「記述」に依拠
- 曖昧性や解釈の表現
- 異質な資料同士の接続
- 資料空間と実体空間の接続

テキスト資料のマイクロナレッジ

```
<div type="book" id="1" >
  <div type="chapter" id="1" >
    <div type="section" id="1" >
      <div type="text" id="1" >
        <span id="caes_bg_1_char_1_1_1" data-bbox="115 451 135 461" style="color: red; font-weight: bold; font-size: small;">Gallia est omnis diuisa in partes tres, quarum unam incolunt Belgae, aliam Aquani, tertiam qui ipsorum lingua Celtae, nostra Galli appellantur.
      </div>
    </div>
  </div>
</div>
```

```
section. 1
Gallia est omnis diuisa in partes tres, quarum unam incolunt Belgae, aliam Aquani, tertiam qui ipsorum lingua Celtae, nostra Galli appellantur.
<span id="caes_bg_1_char_1_1_1" data-bbox="115 451 135 461" style="color: red; font-weight: bold; font-size: small;">Gallia est omnis diuisa in partes tres, quarum unam incolunt Belgae, aliam Aquani, tertiam qui ipsorum lingua Celtae, nostra Galli appellantur.
<span id="caes_bg_1_char_1_1_2" data-bbox="135 451 155 461" style="color: red; font-weight: bold; font-size: small;">est omnis diuisa in partes tres, quarum unam incolunt Belgae, aliam Aquani, tertiam qui ipsorum lingua Celtae, nostra Galli appellantur.
```



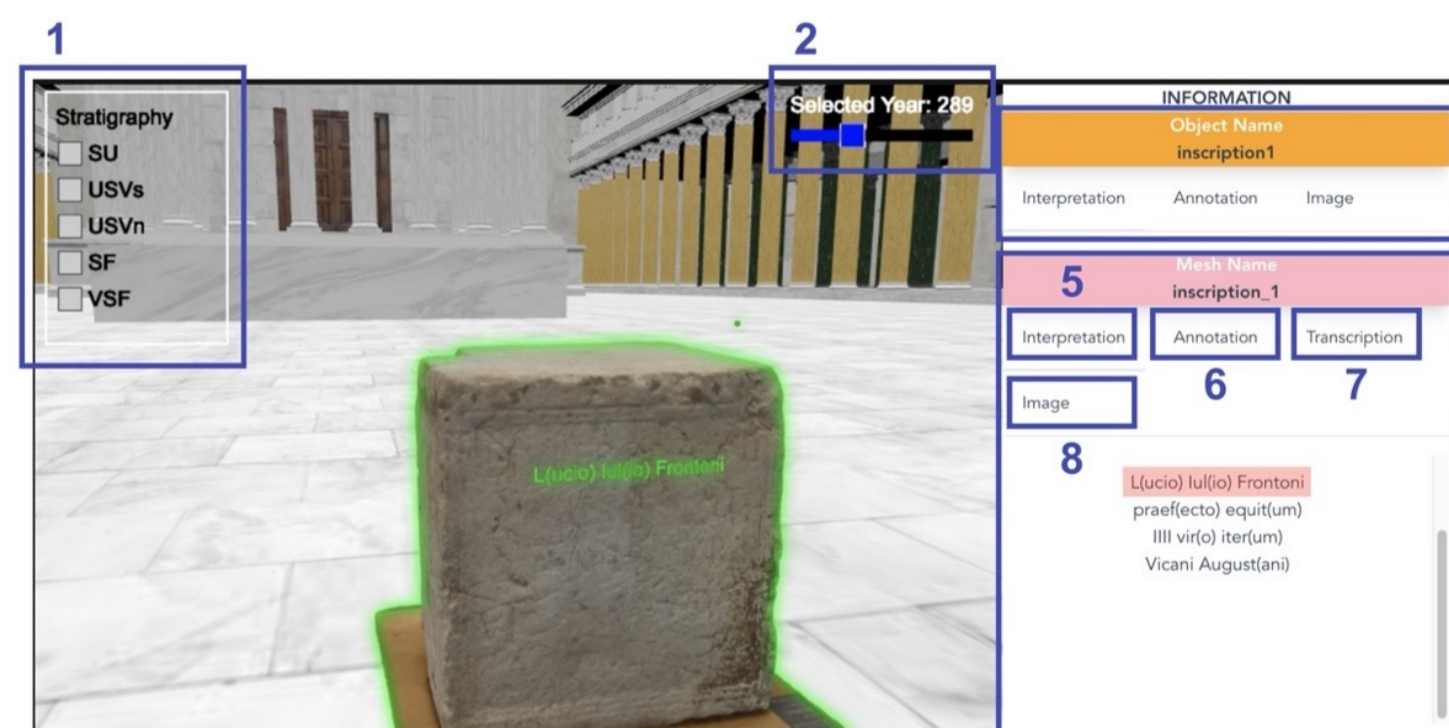
TEI/XMLによるテキスト・マークアップ
テキストそのものの構造や校訂情報の表現に優れたTEIでベーステキストを作成することで、RDFによる知識構造化の根拠となる一次資料記述への高い参照性を確保

RDFによる知識「言及」の構造化
テキストそのものの構造や校訂情報の表現に優れたTEIでベーステキストを作成することで、RDFによる知識構造化の根拠となる一次資料記述への高い参照性を確保

分析・可視化への応用
SPARQLを用いたデータ検索、TEI/XMLファイル参照により、テキスト・知識グラフ・GIS等を横断したインタラクティブな情報抽出と分析・可視化が可能

3Dマイクロナレッジ

3D空間の中の個々のオブジェクト（柱一本や石碑一つなど）をRDFリソースとして、関連情報を構造化する！



- ① 考古学的情報
- ② 時間情報
- ③ 「空間」に関する情報
- ④ 個々のオブジェクトに関する情報
- ⑤ オブジェクト作成の根拠となる「解釈」情報。
- ⑥ 注釈情報
- ⑦ オブジェクト上のテキストに関する情報
- ⑧ 関連する画像データ

オブジェクトに関する情報をRDFで記述（「解釈」リソースを中心としたモデル）
テキスト情報はTEI/XMLで構造化し3Dオブジェクトと接続
構造化にあたっては独自のオントロジーを定義し、個々のオブジェクトについての「解釈」を軸に、3Dモデルデータや関連資料、注釈等をLinked Dataとして構造化している。
テキストデータは、TEI/XMLで構造化。テキスト構造や校訂情報はもちろん、テキストの3D空間における座標情報も含まれ、TEIでマークアップしたテキストを自動的に3D空間にレンダリングすることが可能。

マイクロナレッジの利点

- 資料全体ではなく、その中の個々の「言及」を構造化の基本単位とすることで、資料タイプに基づく差異を最小化し、ある程度共通の構造でデータ化が可能。
- 各資料における個々の「言及」の内容について、より粒度の細かい構造化を実現。
- 資料の部分部分を個別のデータとして構造化するため、資料をまたいだデータリンクの構築に際して、各資料の特定の箇所同士を接続するなど、より精度の高いデータ接続が可能。

データ構築システム

HIMIKOエディタ

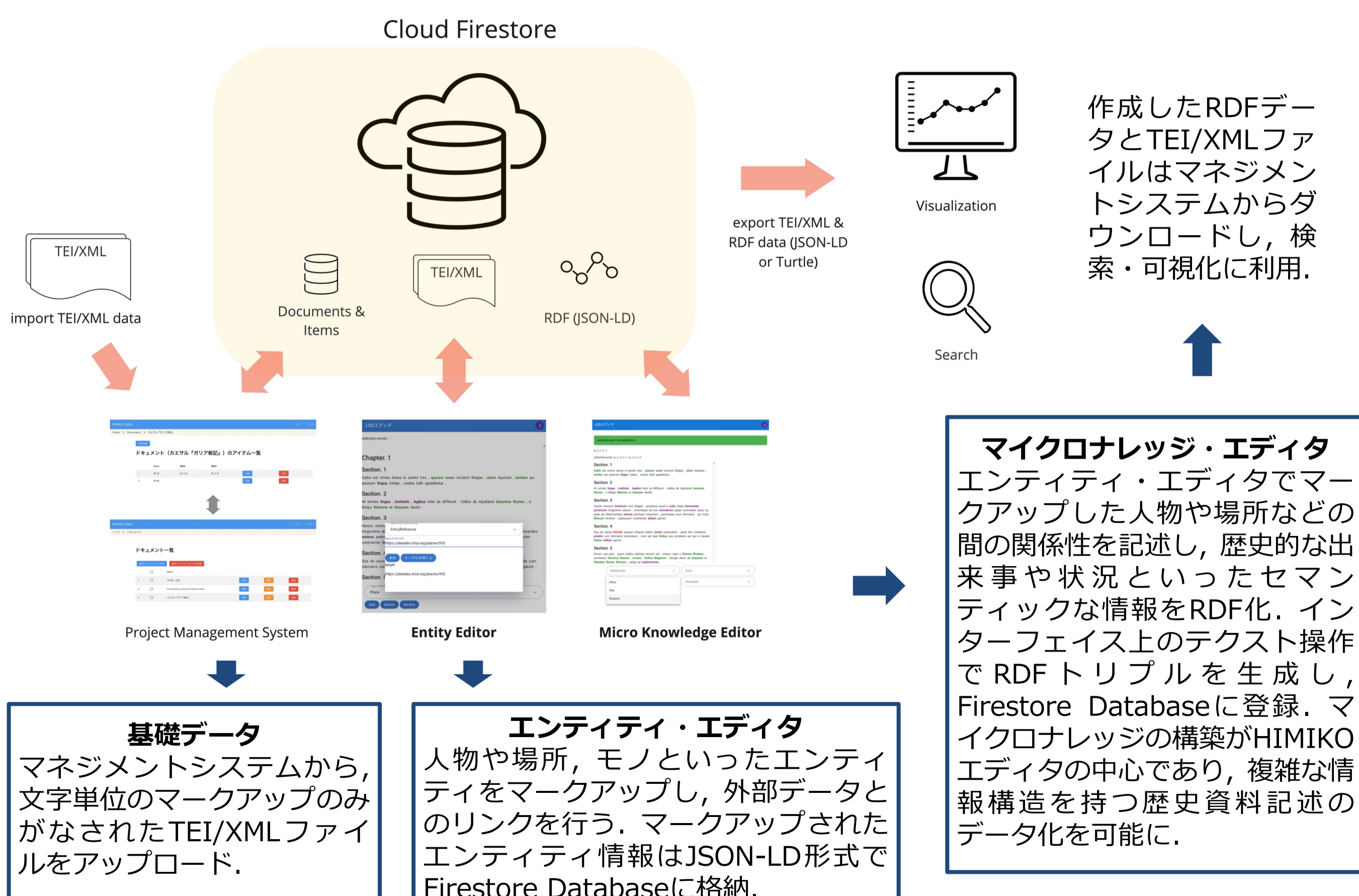
HIMIKOエディタとは？

テキスト資料のマイクロナレッジ・データ構築を支援するシステムで、テキスト・インターフェイス上でRDFデータを作成することが可能

何のため？

歴史研究に携わる専門家が、これまでの史料読解と近い感覚でテキストを読みつつ、データ構築に参画できるようにするため

将来的には、テキスト資料のマイクロナレッジのみならず、3Dマイクロナレッジや画像資料への拡張も！



研究実践（共同研究）

- **イスラム史プロソポグラフィ研究**
イスラム史研究者との共同研究。ムムルーク朝におけるイスラム宗教指導者間の師弟関係、免状授受関係を中心に、人物の経歴や社会関係を分析。人物録資料の記述内容をRDFでデータ構造化する試みを進めている。
- **古代ローマ碑文のデジタル分析**
古代ローマ史研究者との共同研究。大規模なラテン語碑文テキストデータを対象に、内容分析の手法を検討中。既存のRDFデータを用いた碑文の検索・収集、碑文内容のRDF構造化等を進めている。

その他、アイスランドのサガへの応用なども...