



KING'S
College
LONDON

SPQR: Supporting Productive Queries for Research

Mark Hedges, Tobias Blanke: King's College London

Mike Jackson, David Scott: University of Edinburgh

Stefan Gradmann: Humboldt University, Berlin

JISC



Background

- Discipline: classics, ancient history, ancient documents (inscriptions, papyri)
- Many hand-crafted datasets, e.g. XML documents (TEI/EpiDoc), databases,
- Datasets narrow in focus but conceptually related, small parts of a bigger picture
- Datasets much more useful if they could be explored as a single landscape

Aims and Objectives

- Broad aim: break data out of the silos so that it can be searched/browsed as a whole
- How: transform into RDF triples, extracting and representing entities and connections explicitly.
- Deliver corpus of triples as demonstrator, (combination of automated / manual)
- Deliver case study, evaluation, guidelines

Challenges of data

- Data is messy
- Datasets may not be optimally designed
- Datasets have embedded, implicit semantics
- Information is fuzzy, incomplete, ambiguous, contradictory
- Some standardisation (e.g. EpiDoc), but very broad in application (e.g. Use of vocabularies)
- Highly interpretative, contextual

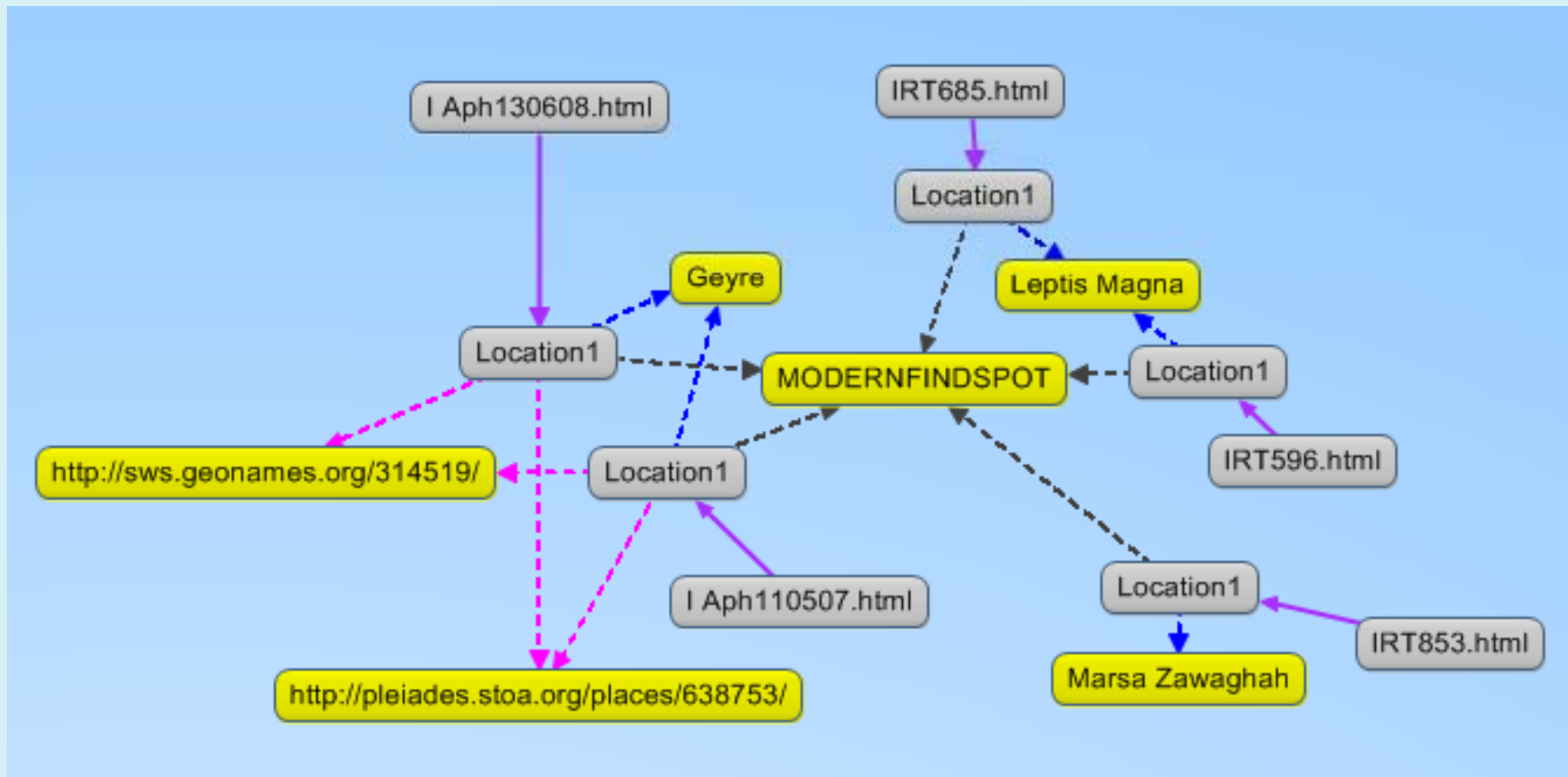
Challenges of researcher practices

- Hard to pin down to specific queries
- Researcher workflows / work processes are complex, difficult to predict, highly interactive
- Results of one search (together with other information) influence questions asked of others
- Authority based on provenance/citation of sources and peer assessment of choices/inferences
- Focus on – who? when? where? – for research questions and resources linking

Approach / Solutions

- TEI (and database) to RDF conversion:
 - Bespoke scripts exploiting knowledge of schema
 - Entities / relationships based on mark-up.
- Entity extraction and mapping:
 - NLP for named entity extraction, mapping to existing URIs (ancient places, personal names)
- Combination of automated and manual:
 - Researcher interpretation, e.g. Which Alexandria?
Which Gaios?

Example: references to place



Links with other projects

- Europeana – portal for European cultural and scientific heritage
 - Publishing data to Europeana
 - Integration with Europeana Data Model
- PELAGIOS – JISC-funded project (jiscGECO) creating generic format (ontology) for referring to ancient places in Web documents
 - We are a “data partner”, will use ontology

Questions / Issues

- How to best combine automatic processing and manual annotation?
- What are the semantic losses when transforming data (e.g. From TEI to RDF)?
- What can LD do better than TEI (and vice versa)?
- Quantity versus quality
- Trial with researchers using browser (Gruff) to obtain feedback