Capturing contextual information in Semantic Web linked data

Graham Klyne, e-Research Centre, University of Oxford; graham.klyne@oerc.ox.ac.uk June 2018

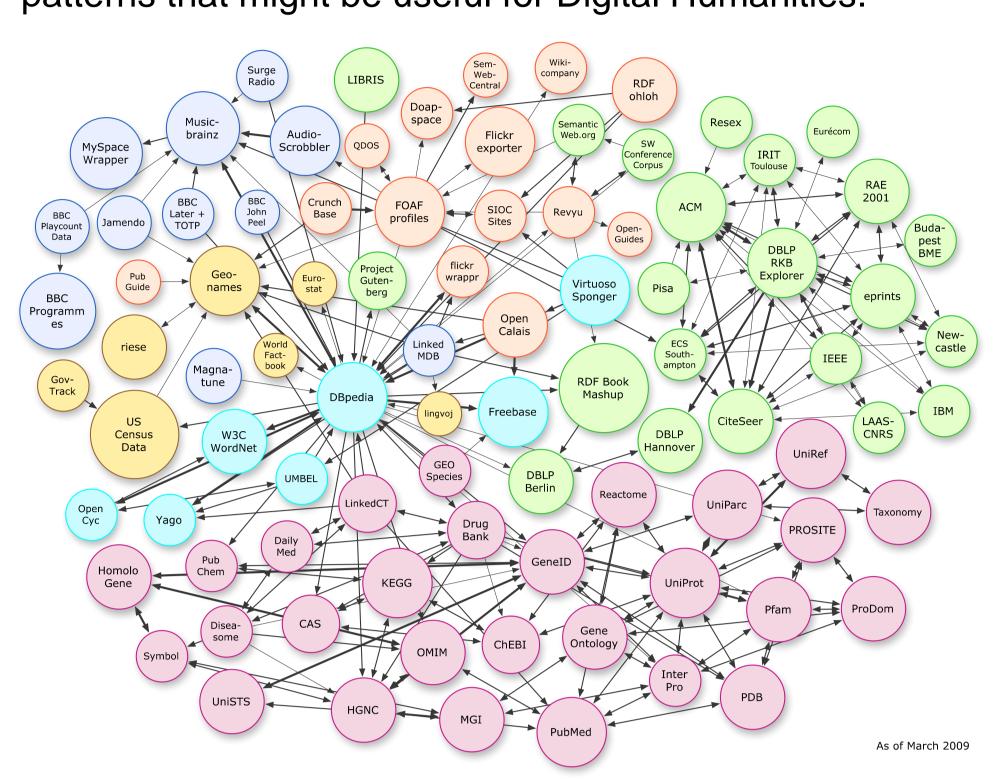
Keywords: Semantic Web, Linked data, Information models, Musical performance, Historical places



Abstract

The Resource Description Framework (RDF) [1] has been widely used for scientific and observed data [2]. Humanities require a greater emphasis on capturing the context in which information arises.

The simplicity of the RDF data model creates challenges and opportunities. Event-based ontologies, such as CIDOC CRM [13] and PROV [7], capture some aspects of context. We present some recent RDF modelling work for musical performances and historical places that capture other aspects of context, and suggests some possible modelling patterns that might be useful for Digital Humanities.

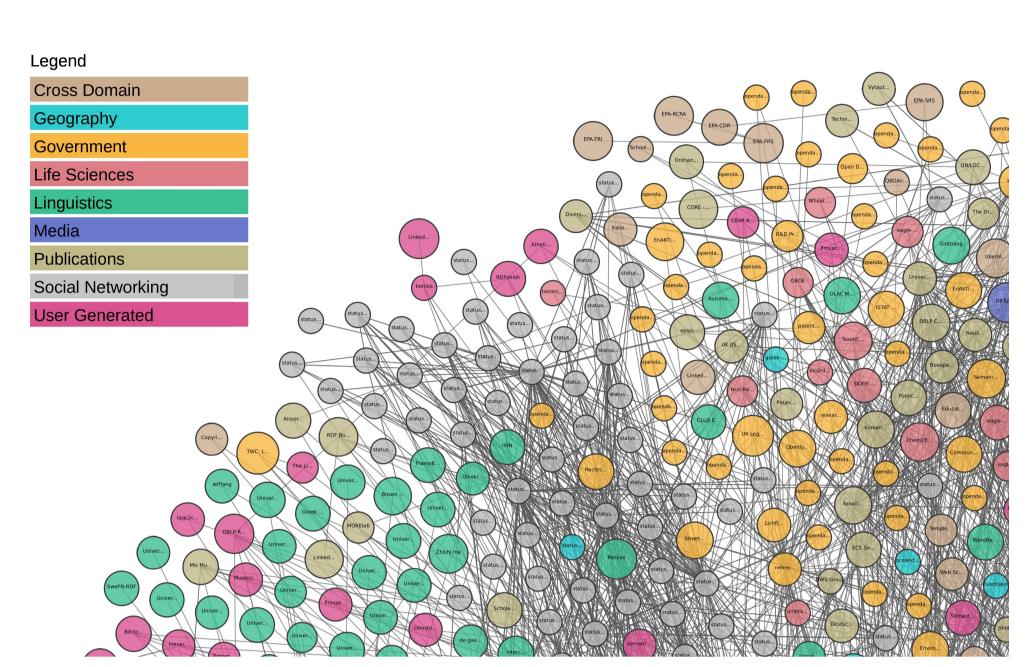


Linked open data cloud in 2009 [3]

Background

Early growth in the use of RDF linked data was for social networks, media, governmental data, academic publishing and life sciences [3].

Nearly a decade later, we see growing interest in linked data for humanities, though relatively little impact, as yet, is visible in the Linked Open Data cloud [2]. One of the challenges is that humanities data can be rather more subtle and context-dependent than direct observations and scientific data. We have recently designed RDF data models for musical performances and historical places that attempt to address these challenges.



Part of linked open data cloud in 2018 [2]

The challenge

The Resource Description Framework (RDF) data model [1] is very simple, and can be seen in outline in Tim Berners-Lee's early proposal for the World Wide Web [4]. The essential idea is that different resources (the shapes) are related by labeled connections. But this simple model does not directly support qualification of the relations. There have been various suggestions for capturing context of relations in RDF (e.g. see [5]), some of which involve revisions to the RDF model, which in turn can limit the choice of software tools, and also applicability of RDF's formal semantics.

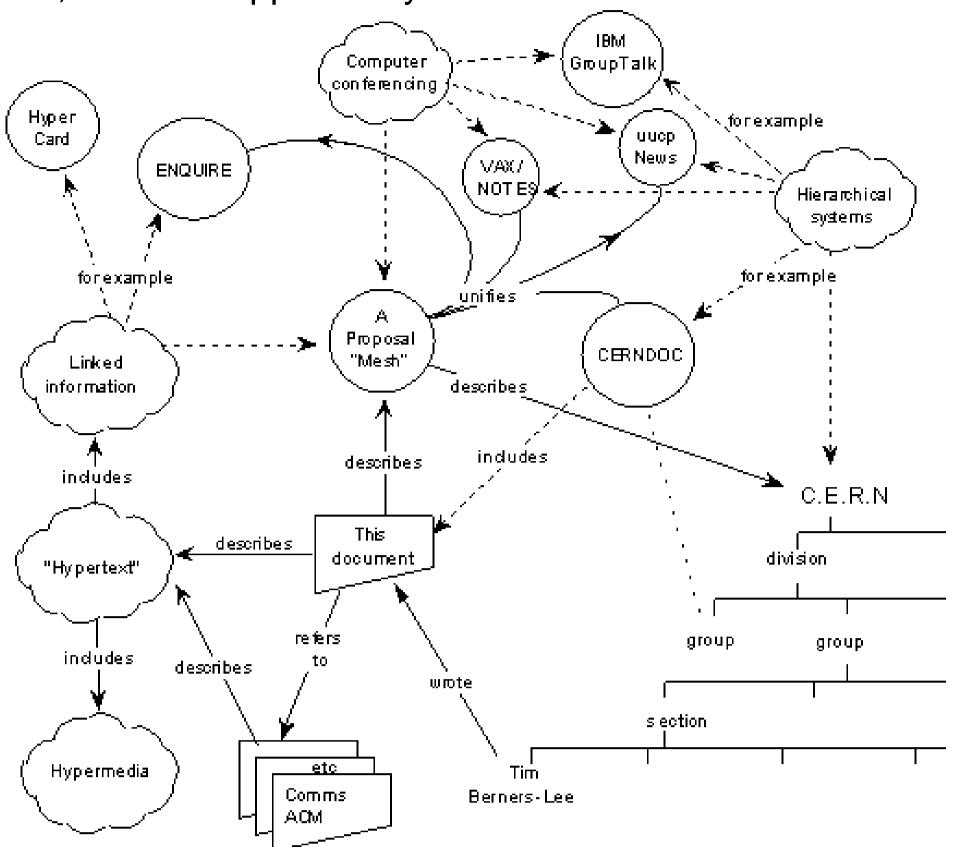


Diagram from original WWW proposal [4]

Performance

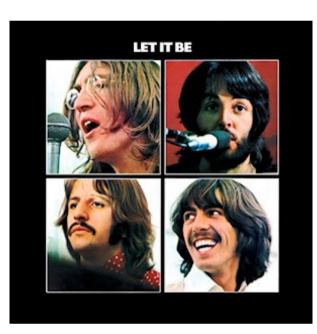
(Across the Universe)

hadRole

Pianist

qualifiedAssociation

Example: Musicians and performances

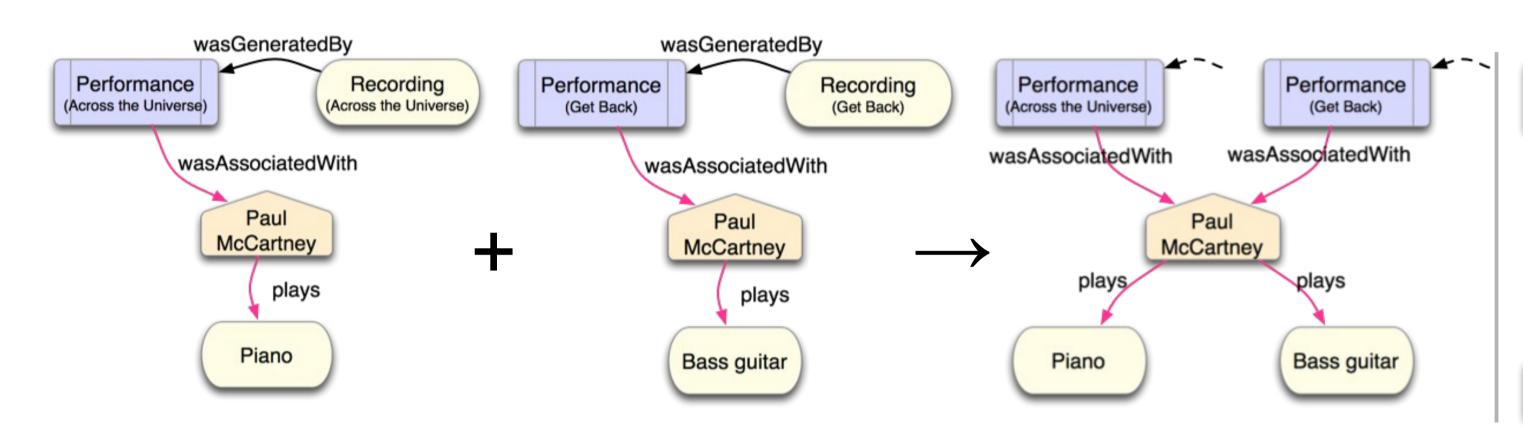


https://en.wikipedia.org/wi ki/Let_It_Be

On this 1970 Beatles album, Paul McCartney plays several instruments:

On the song "Across the Universe", he plays piano, but on "Get Back", he plays bass guitar.

How might we express this as linked data?



(We use here the PROV [7] model that relates the song to the artist via a "Performance" activity)

Connecting the performance role directly to the artist means that, when multiple performances are combined, contextual information about their role in a particular performance is lost.

So we use a PROV "Association" for the relationship between the artist and a particular performance, which allows us to capture their role in that particular context.

wasAssociated

McCartney

20180305-EMPlaces-data-model-using-annotations.graffi

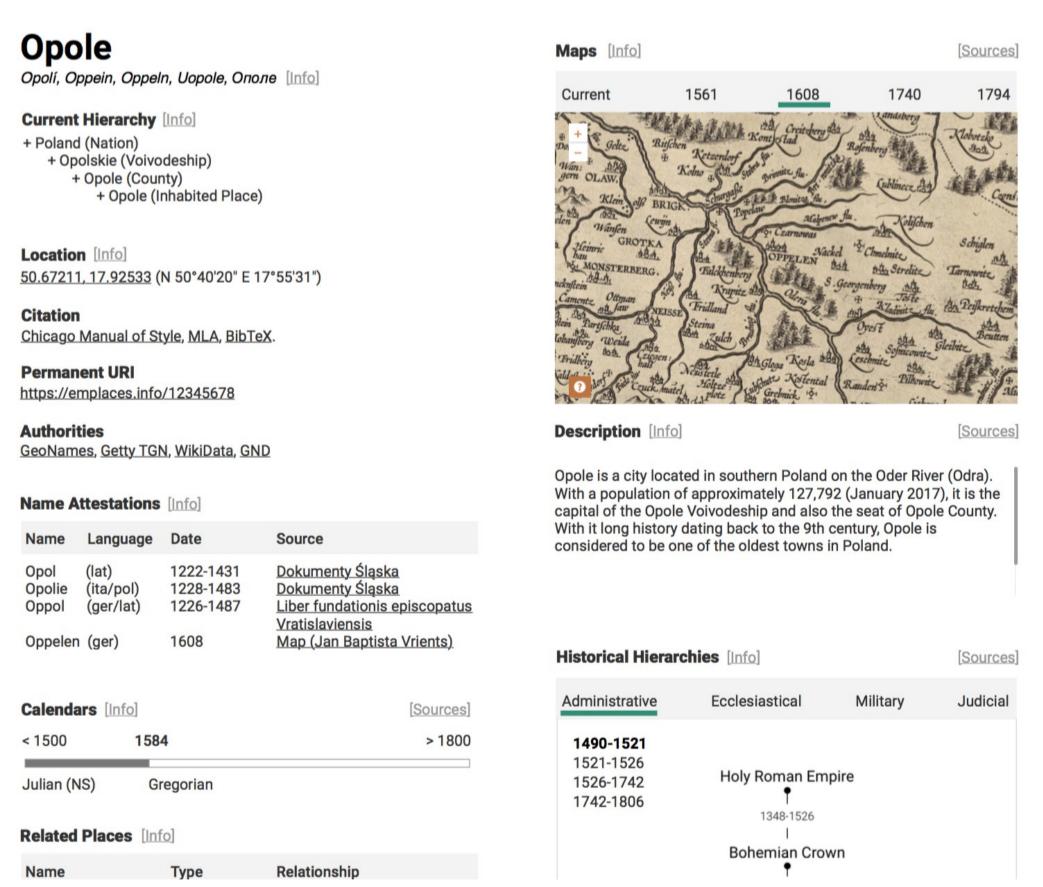
Performance

(Get Back)

Bass guitarist

qualifiedAssociation

Example: Historical places – Opole, 1222 - 1806



Part of web page mock-up by Arno Bosse

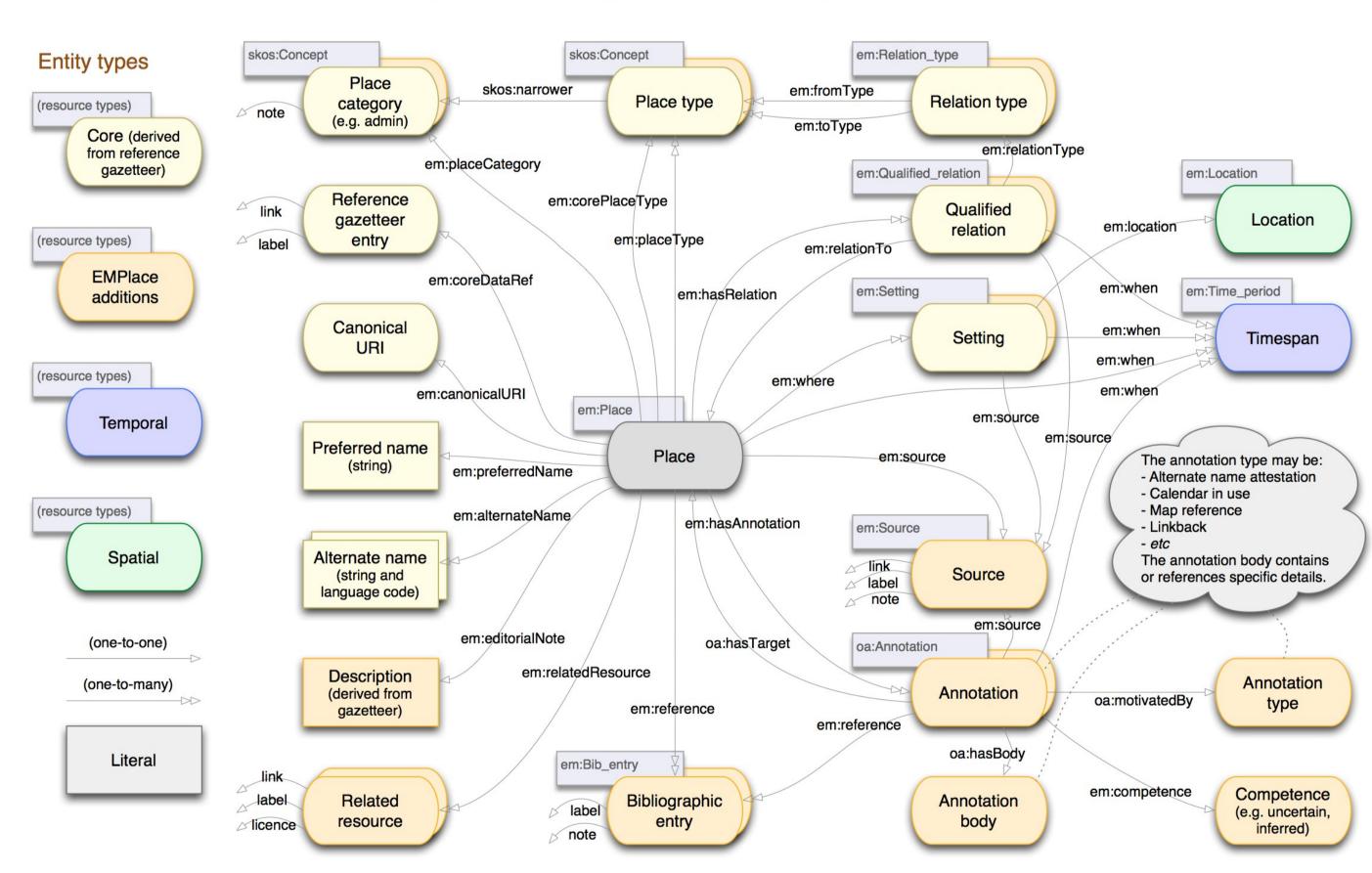
Places, as "constructed by human experience"*, may change over time.

Their names, relations to other places, the calendars they use, and even their geographical location may vary through different periods in their history.

This means that if we make simple uncontextualized claims about a place, they may not universally true.

To reflect our knowledge of historical places in linked data, we must capture and represent their changing context.

Data model for EMPlaces using annotations (draft 2018-05-29)

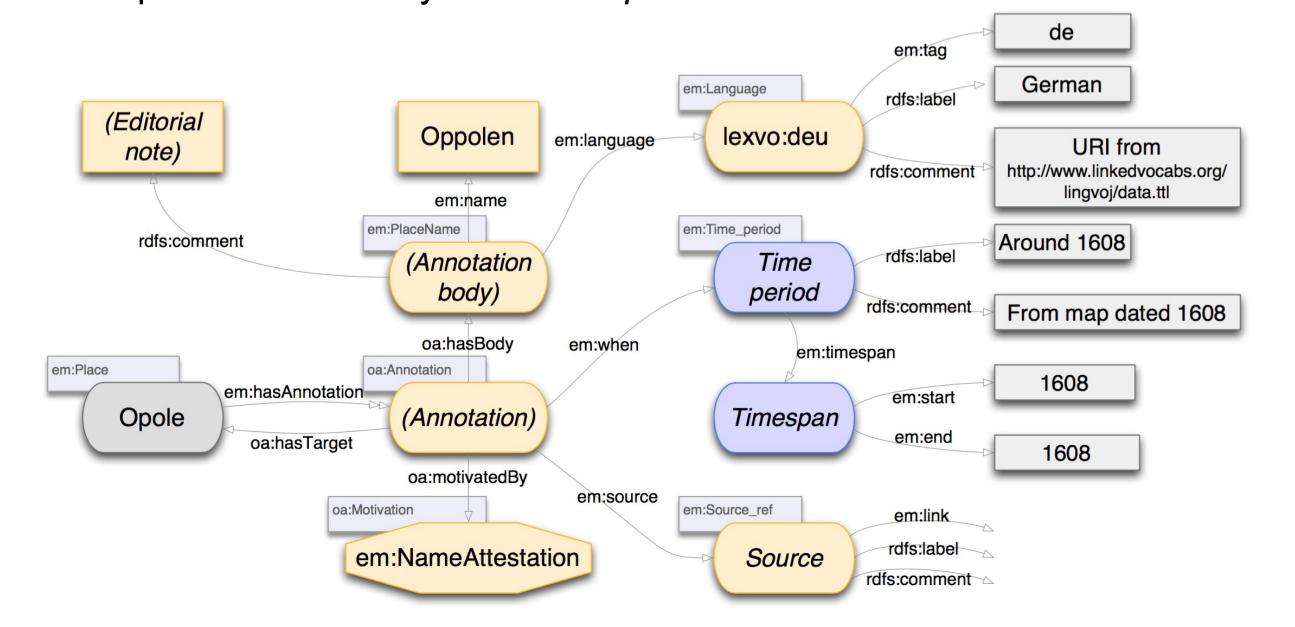


Our data model uses a "Qualified relation" to capture changing relationships between places, and Web Annotations for other contextualized information such as name attestations, calendars in use, etc.

Name attestation using Web Annotations

A pattern we are exploring uses the Web Annotation standard [12], originally designed for distributed annotation of web pages.

The structure can also record contextual information. Here, the annotation includes the date when an alternative German name, Oppolen, was used for the place we currently know as *Opole*.



Conclusions

Care is needed when building linked data for humanities to capture contextual information that is needed for their correct interpretation.

Several approaches have been surveyed [5], and a pattern that appears effective and reasonably easy to use is selective "reification" of relations -i.e., to give form and identity to individual relations so that additional statements can be made about them.

One form of this approach uses Web Annotations, an established linked data standard, increasingly used and supported by software tools. Web Annotations can support multiple and potentially conf lcting interpretations, with inclusion of additional contextual information for reconciling any conflcts.

References

[1] www.w3.org/TR/rdf11-concepts/

[2] lod-cloud.net/versions/2018-30-05/lod-cloud.png

[3] lod-cloud.net/versions/2009-03-05/lodcloud_colored.svg

[4] www.w3.org/History/1989/proposal.html

[5] github.com/gklyne/notes/blob/master/20180307-

qualified-property-links.md

[6] musicontology.com/specification/ [7] Www.w3.org/TR/prov-o/

[8] patterns.dataincubator.org/book/reifiedstatement.html

[9] github.com/culturesofknowledge/emplaces/ blob/master/models/PDFs/20180305-EMPlacesdata-model-using-annotations.pdf

[10] new.cidoc-crm.org/Resources/

modelling-properties-of-properties [11] www.snee.com/bobdc.blog/2018/04/

reification-is-a-red-herring.html

[12] www.w3.org/TR/annotation-model/

[13] www.cidoc-crm.org

^{*} pleiades.stoa.org/help/data-structure