summer school

Trends and Transients

18 September 2019

Linked Data in Digital Humanities

Graham Klyne

University of Oxford

Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

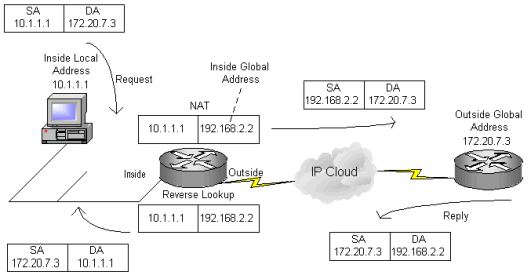
Linked Data in Digital Humanities

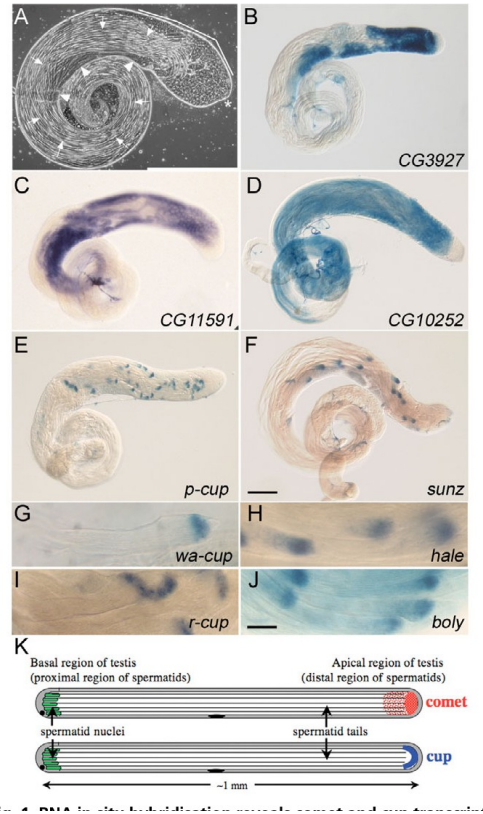
Graham Klyne

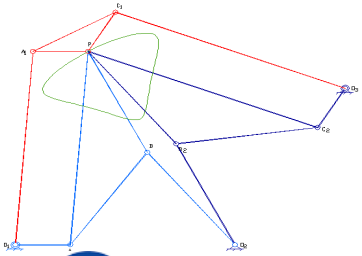
XML Summer School 2019

Trends and Transients

About me







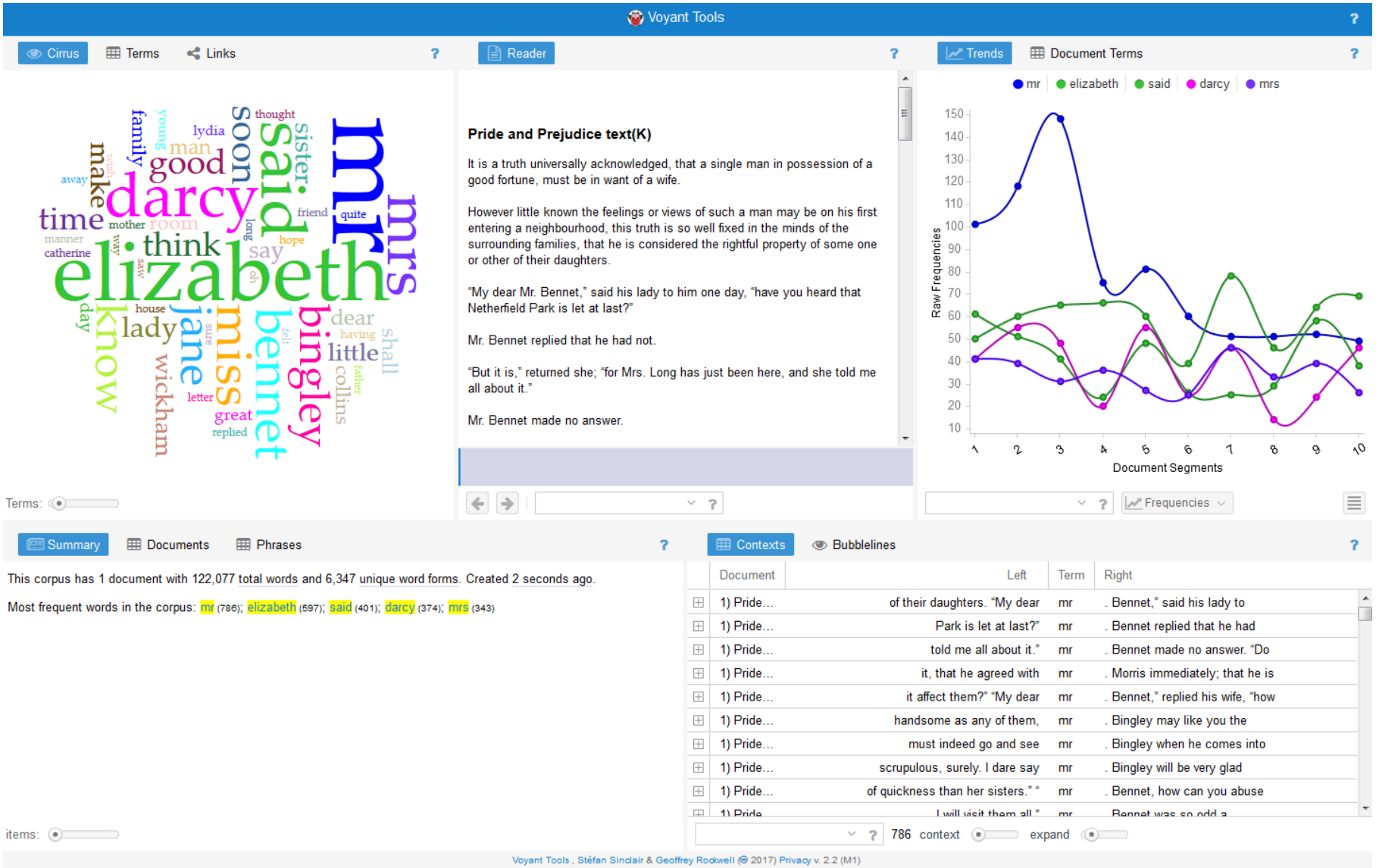
graham.klyne@oerc.ox.ac.uk

Humanities

“Humanities are academic disciplines that study aspects of human society and culture.”

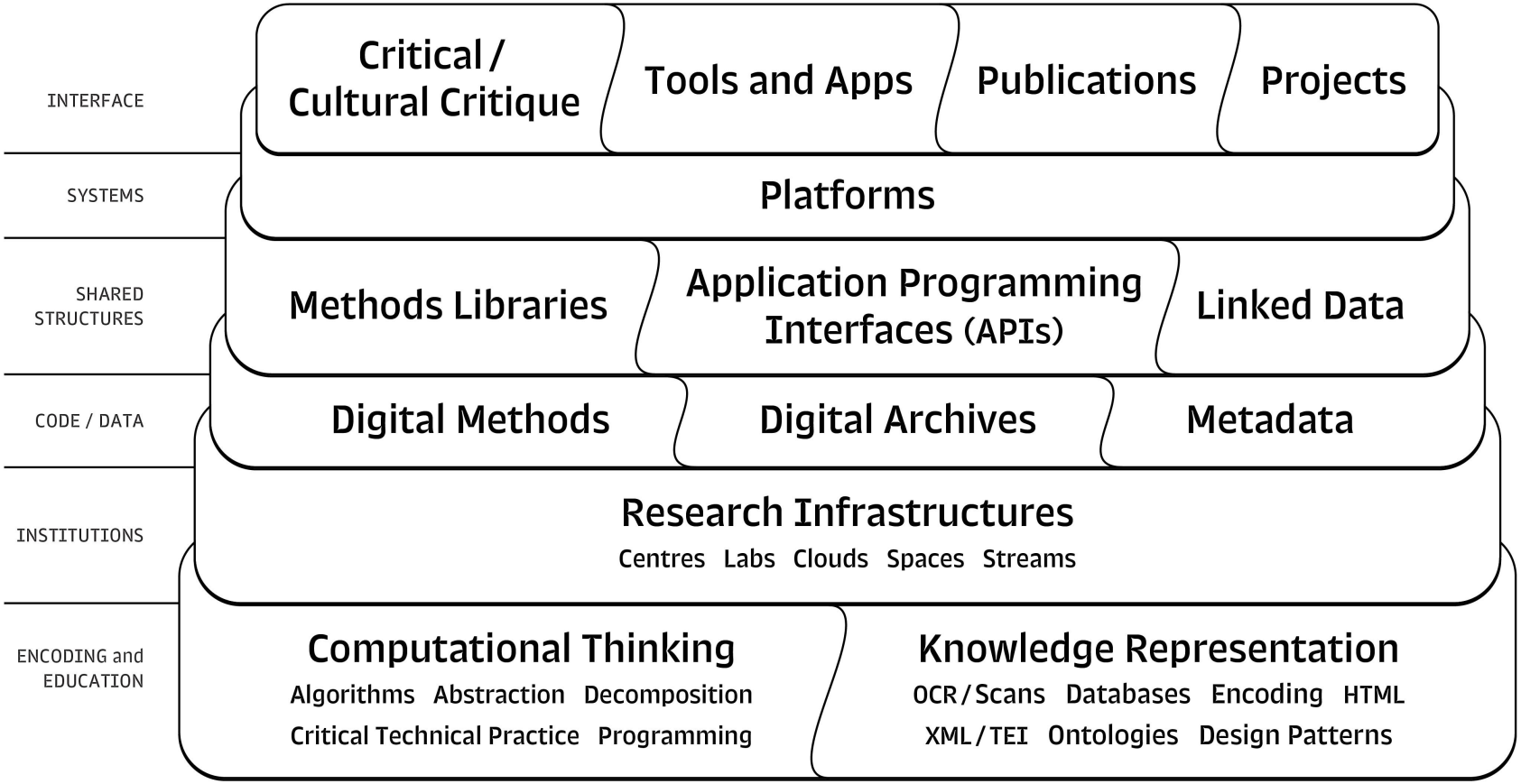
— *Wikipedia* (*https://en.wikipedia.org/wiki/Humanities*)

Digital Humanities



Text of Jane Austen's Pride and Prejudice in textual analysis program Voyant Tools
- *https://commons.wikimedia.org/wiki/File:Pride_and_Prejudice_in_Voyant_Tools.png*

Digital Humanities Stack?

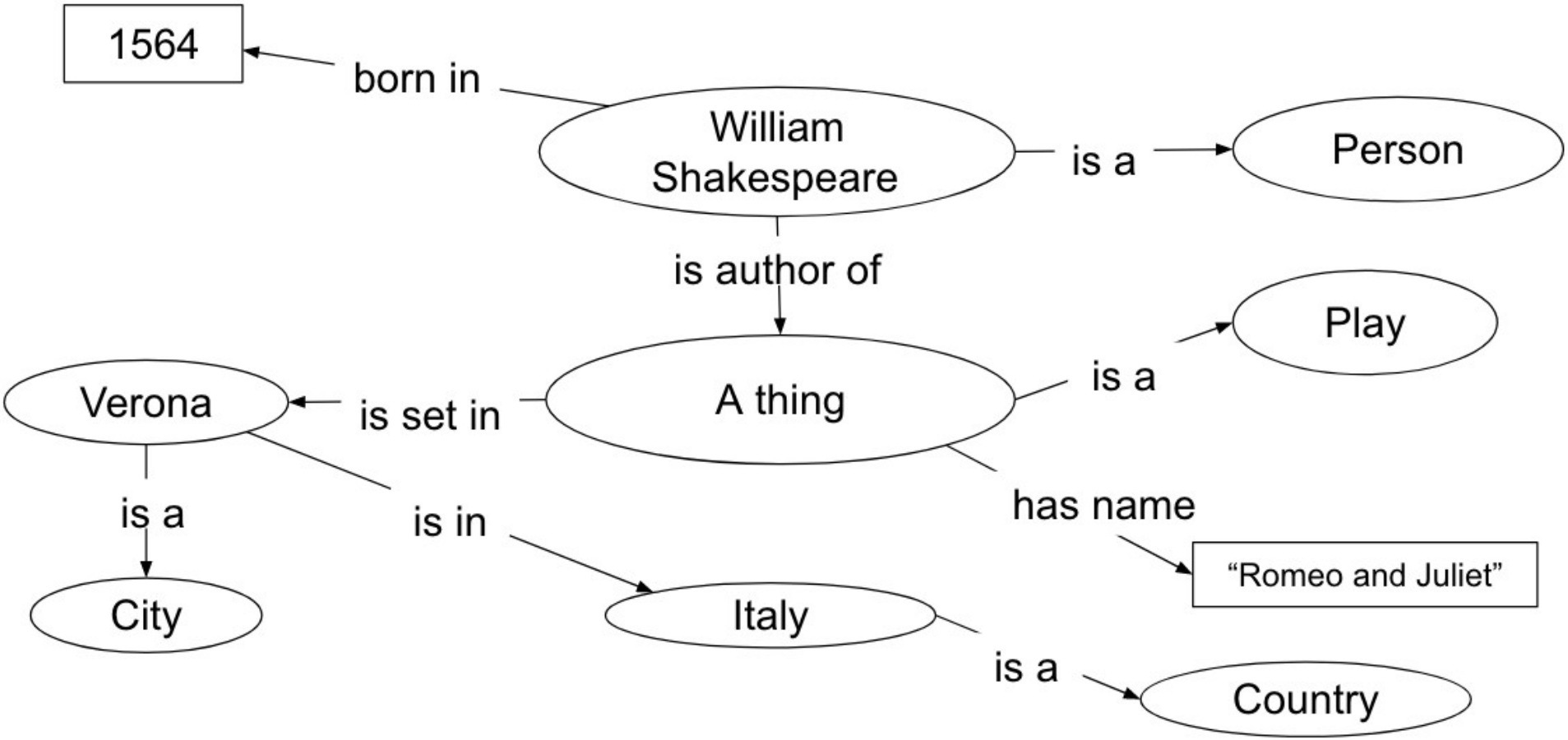


Digital Humanities Stack
- https://commons.wikimedia.org/wiki/File:Digital_Humanities_Stack_(from_Berry_and_Fagerjord_2017-_18).jpg

from Berry and Fagerjord, *Digital Humanities: Knowledge and Critique in a Digital Age*

Introducing Linked Data (RDF)

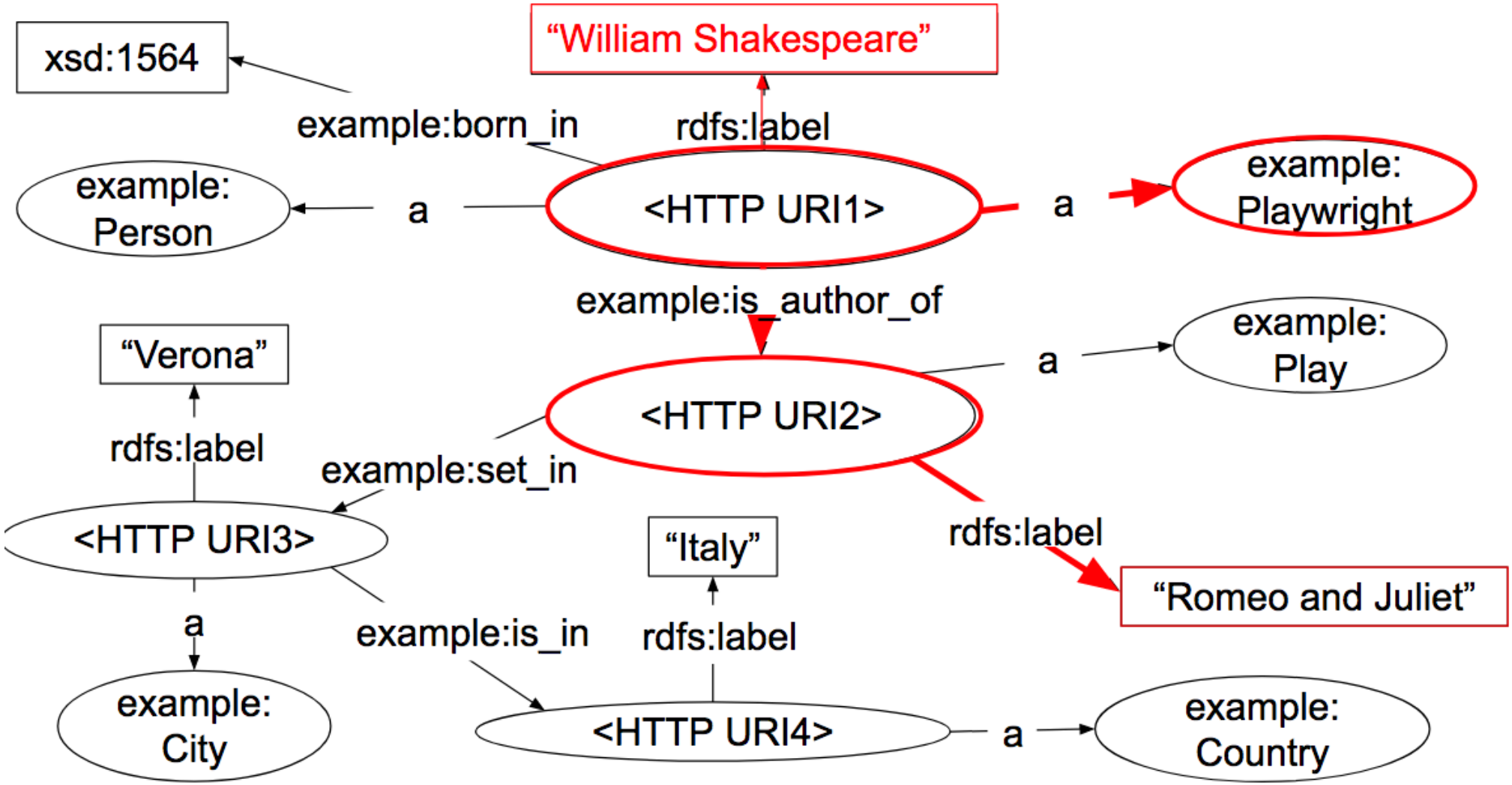
Simple example



Slide credit: Terhi Nurmikko-Fuller

Introducing Linked Data (RDF)

A bit closer to the truth



Slide credit: Terhi Nurmikko-Fuller

Introducing Linked Data (RDF)
The painful reality?

“People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated.”

“While you can avoid RDF, it is harder to avoid complicated data and complicated computer problems. RDF brings together data across application boundaries and imposes no discipline on mandatory or expected structures.”

(Dan Brickley and Libby Miller)

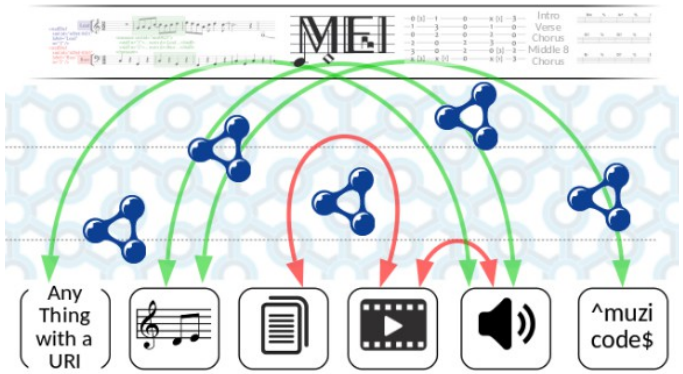
- https://book.validatingrdf.com/bookHtml005.html

About this talk

1. EM Places: Early Modern historical places

The Letter Catchers by Annaliese Stoney for Cultures of Knowledge (with inspiration from Pier Francesco Mola)

2. MELD: Music Encoding and linked data



summer school

EM Places

Early Modern historical place data

Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

EM Places

Early Modern historical place data

EM Places

CULTURES of KNOWLEDGE

Networking the Republic of Letters, 1550-1750

Established in 2009, we are a collaborative, interdisciplinary research project based at the University of Oxfo
Mellon Foundation and the Arts and Humanities Research Council. We are using digital methods to reassen
networks of the early modern period.

The Letter Catchers by Annaliese Stoney for Cultures of Knowledge (with inspiration from Pier Francesco Mola)

EMLO

Early Modern Letters Online is our growing union catalogue of correspondence from the sixteenth, seventeenth, and eighteenth centuries. Developing EMLO is now at the heart of what we do.

EM Places requirements

Opole [Info] [Provenance]

Opolė, Òpòle, Opolí, Oppein, Oppeln, Uopole, Горад Аполе, Ополе, אופולה, أبولوسكي, اوپول, اوپوله, ออปอเล, 오폴레, オポーレ, 奥波莱

Administrative Hierarchy [Info]

+ Poland (Country)
  + Opole Voivodeship (ADM 1)
    + Opole (ADM 2)
      + Opole (ADM 3)
        + Opole (Populated Place)

Location [Info]

Greenwich Meridian: 50.67211, 17.92533 (N 50°40'20" E 17°55'31")

Citation [Info]

Chicago Manual of Style, MLA, BibTeX, RIS

Persistent URI [Info]

https://emplaces.info/ark:/12345/abc67890

Authorities [Info]

GeoNames, Getty TGN, WikiData, GND

Name Attestations [Info] [Provenance]

Name	Language	Date	Source
Opol	(lat)	1222-1431	Dokumenty Śląska
Opolie	(ita, pol)	1228-1483	Dokumenty Śląska
Oppol	(ger, lat)	1226-1487	Liber fundationis episcopatus Vratislaviensis
Oppelen	(ger)	1608	Jan Baptista Vrients

Calendars [Info] [Provenance]

1500 — 1584.01.29 (Inferred) — 1800

Julian (Jan. 1) — Gregorian

Connections [Info] [Provenance]

Name	Type	Relation
St. Adalbert	Church	Located within

Related Resources [Info]

- Herder Institute: *Historical-Topographical Atlas of Silesian Towns*
- ARTFL: Diderot and d'Alembert, *Encyclopédie* (Opole)

Bibliography [Info]

There are 16 publications associated with Opole.

Maps [Info] [Provenance]

Current | 1561 | 1608 | 1740 | 1794

Description [Info]

Opole is a city located in southern Poland on the Oder River (Odra). With a population of approximately 127,792 (January 2017), it is the capital of the Opole Voivodeship and also the seat of Opole County. With it long history dating back to the ninth century, Opole is considered to be one of the oldest towns in Poland.

Historical Hierarchies [Info] [Provenance]

Administrative | Ecclesiastical | Military | Judicial

1490-1521
1521-1526
1526-1742
1742-1806

Holy Roman Empire
1348-1804
Bohemian Crown
1490-1742
(Governorate of) Duchy of Silesia
1422-1521
Duchy of Opole
1281-1521
Opole

EM Places require-ments

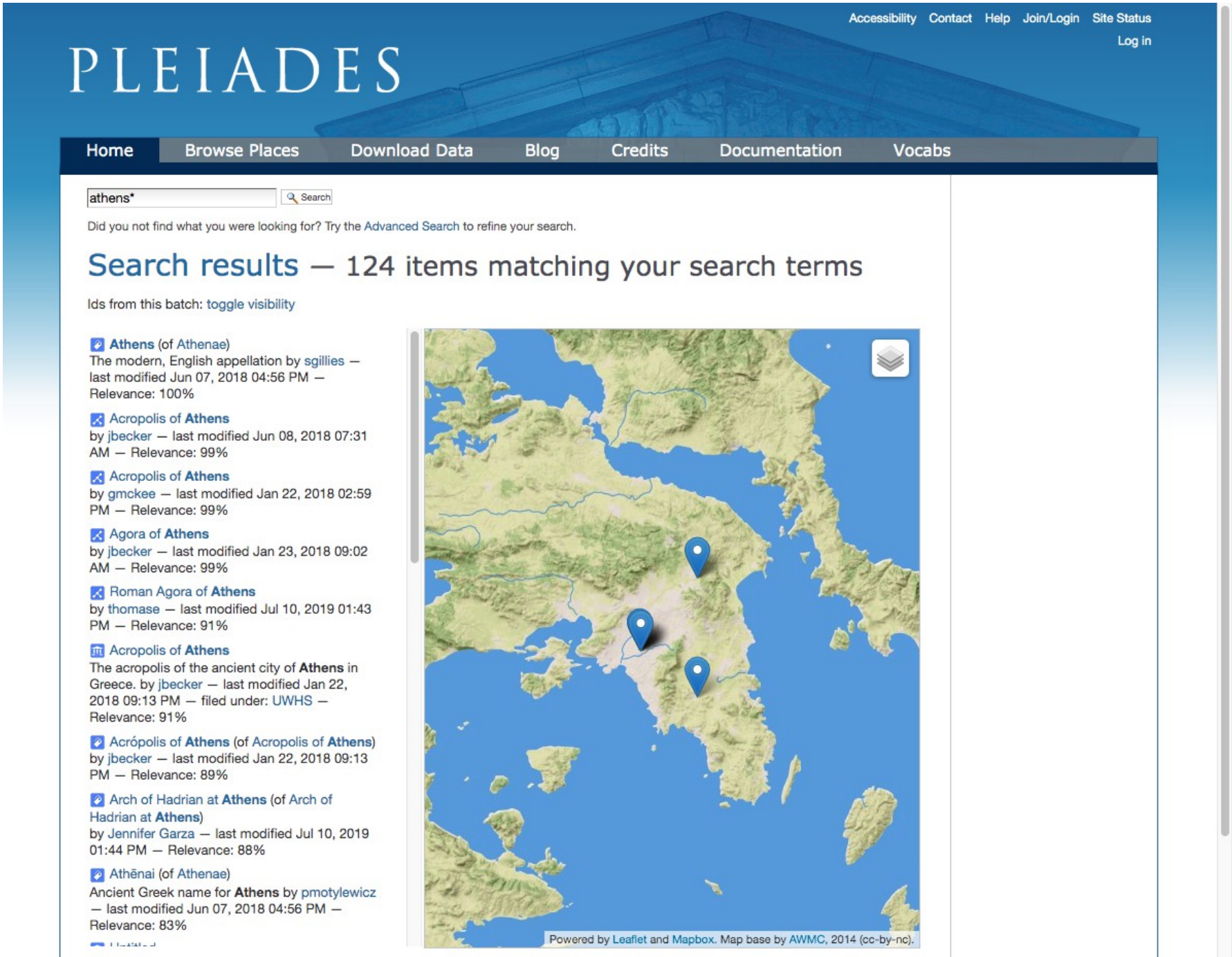
Historical Hierarchies [Info] [Provenance]

Administrative | Ecclesiastical | Military | Judicial

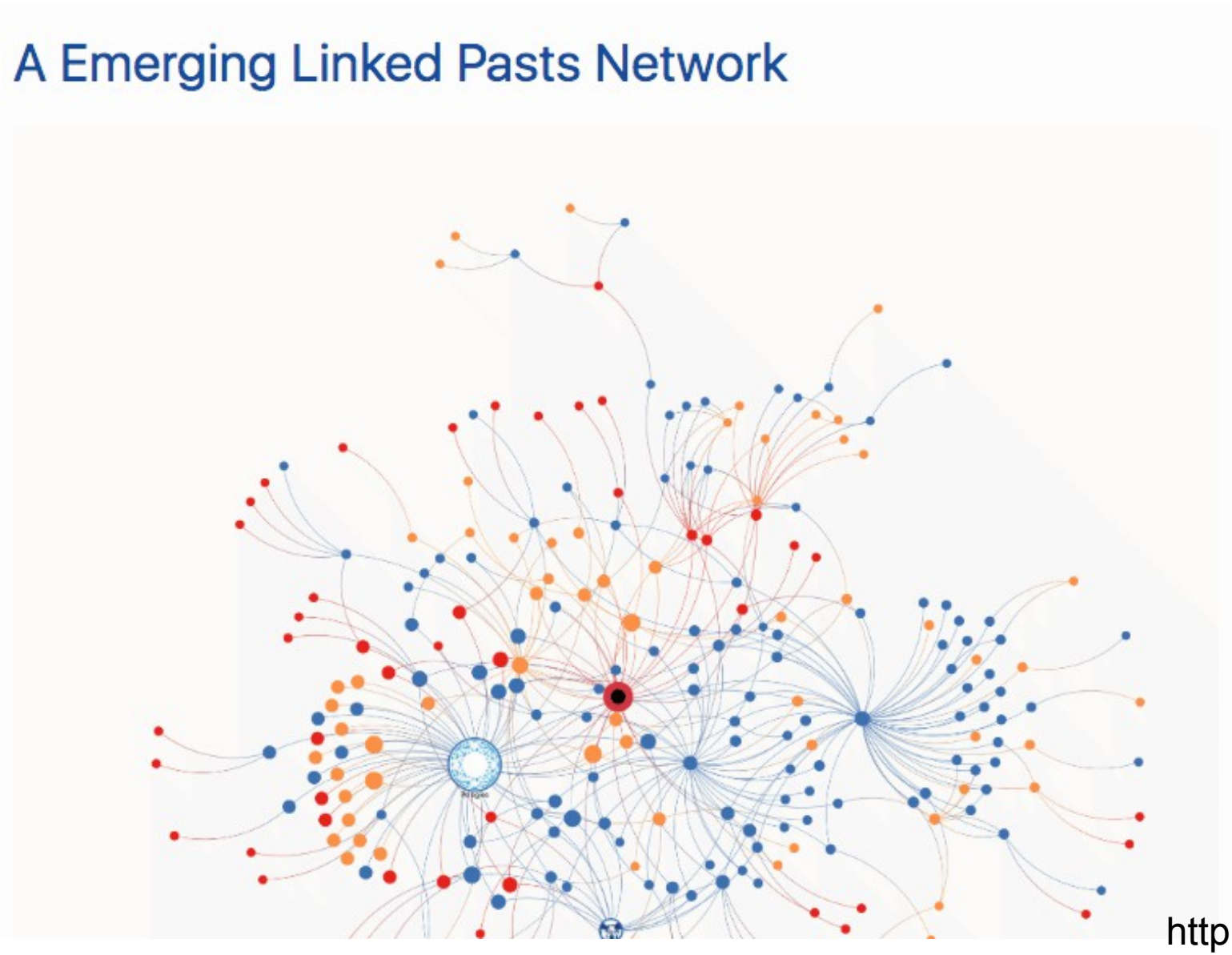
1490-1521
1521-1526
1526-1742
1742-1806

Holy Roman Empire
1348-1804
Bohemian Crown
1490-1742
(Governorate of) Duchy of Silesia
1422-1521
Duchy of Opole
1281-1521
Opole

Related work: Pleiades



Related work:
Linked Pasts/Linked Places



http://linkedpasts.org/

What is a place?

“constructed by human experience”

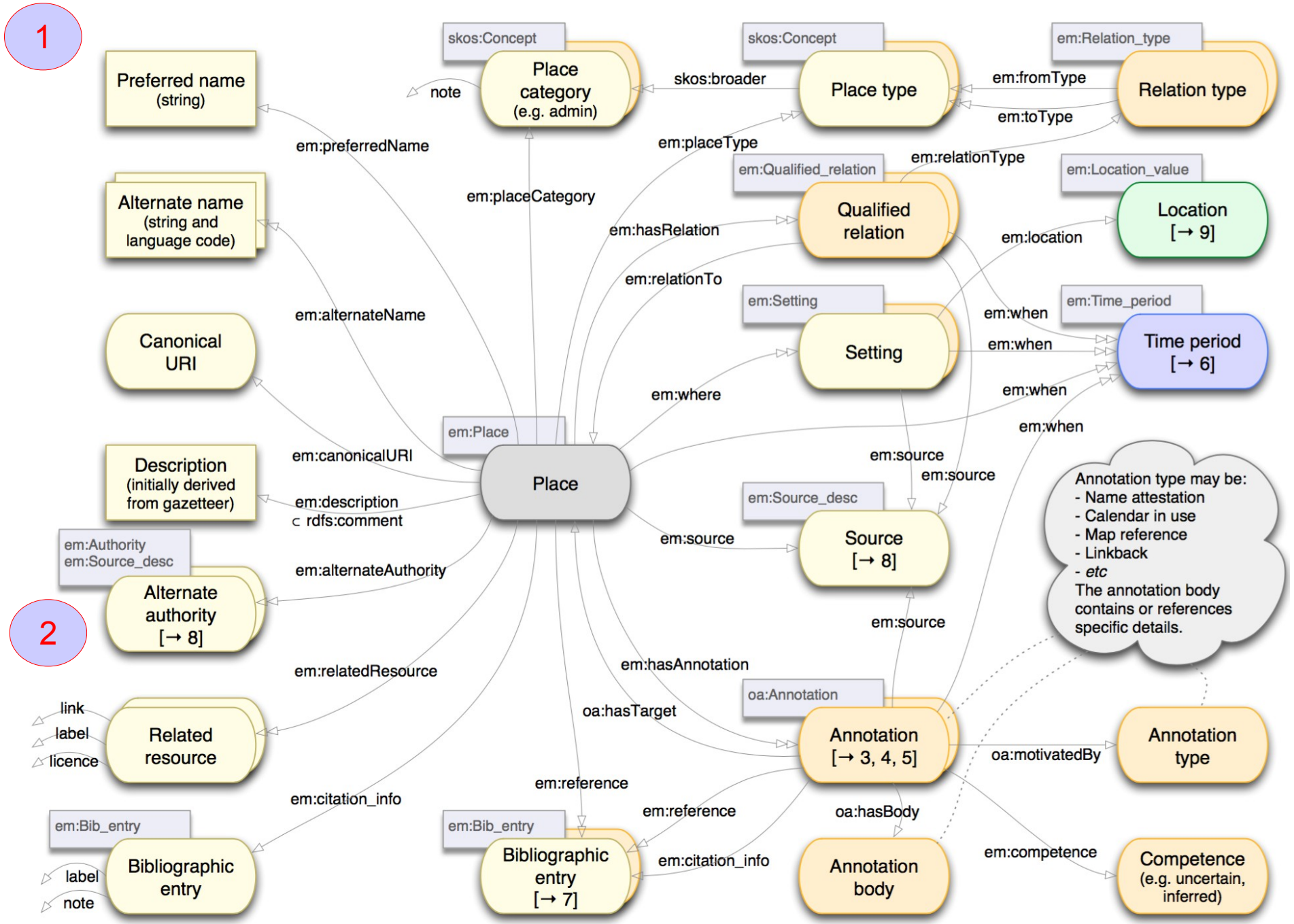
Pleiades content is organized into three types of "information resource": *places*, *locations*, and *names*.

Places

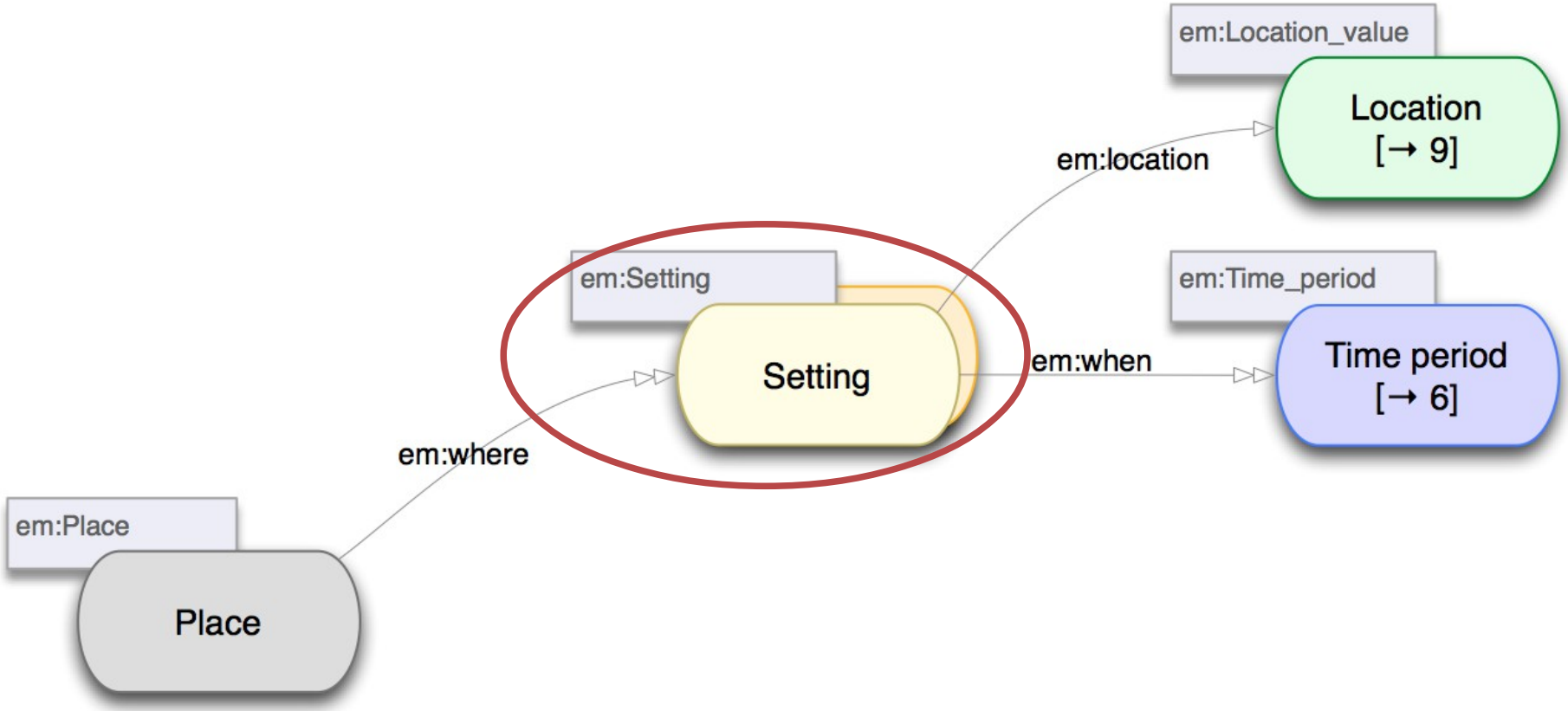
In *Pleiades*, *places* are conceptual entities. They are objects of thought, speech, or writing, not tangible, mappable points on the earth’s surface. Even though they are usually tied closely to some aspect or feature of the physical world — a sea, a bay, a river, a mountain range, a pass, a road, a settlement, or an ethnic region — They have no spatial or temporal attributes of their own. Their primary quality is that, in the words of Yi-fu Tuan [1], they are "constructed by human experience."

-- https://pleiades.stoa.org/help/data-structure

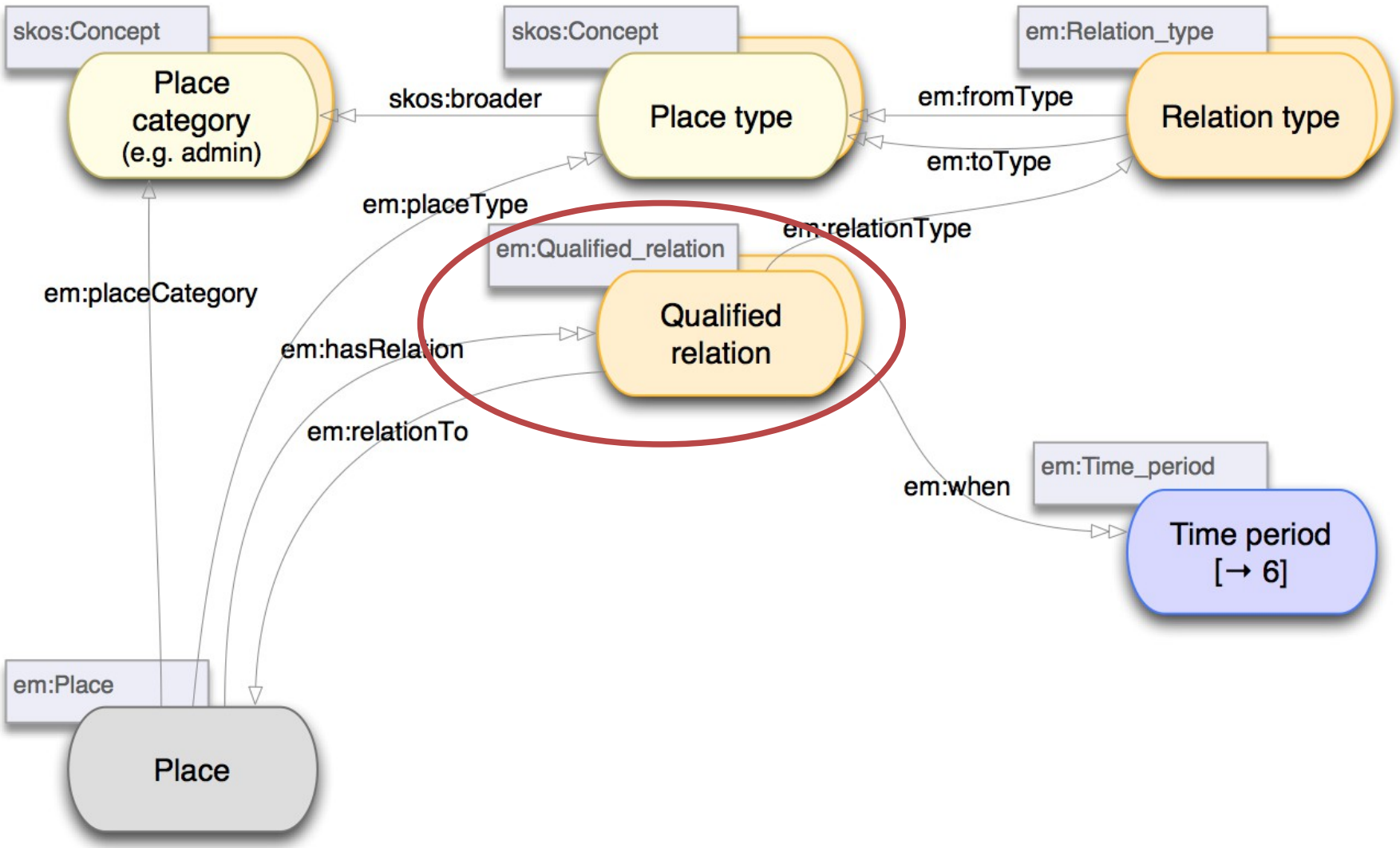
Overall place model



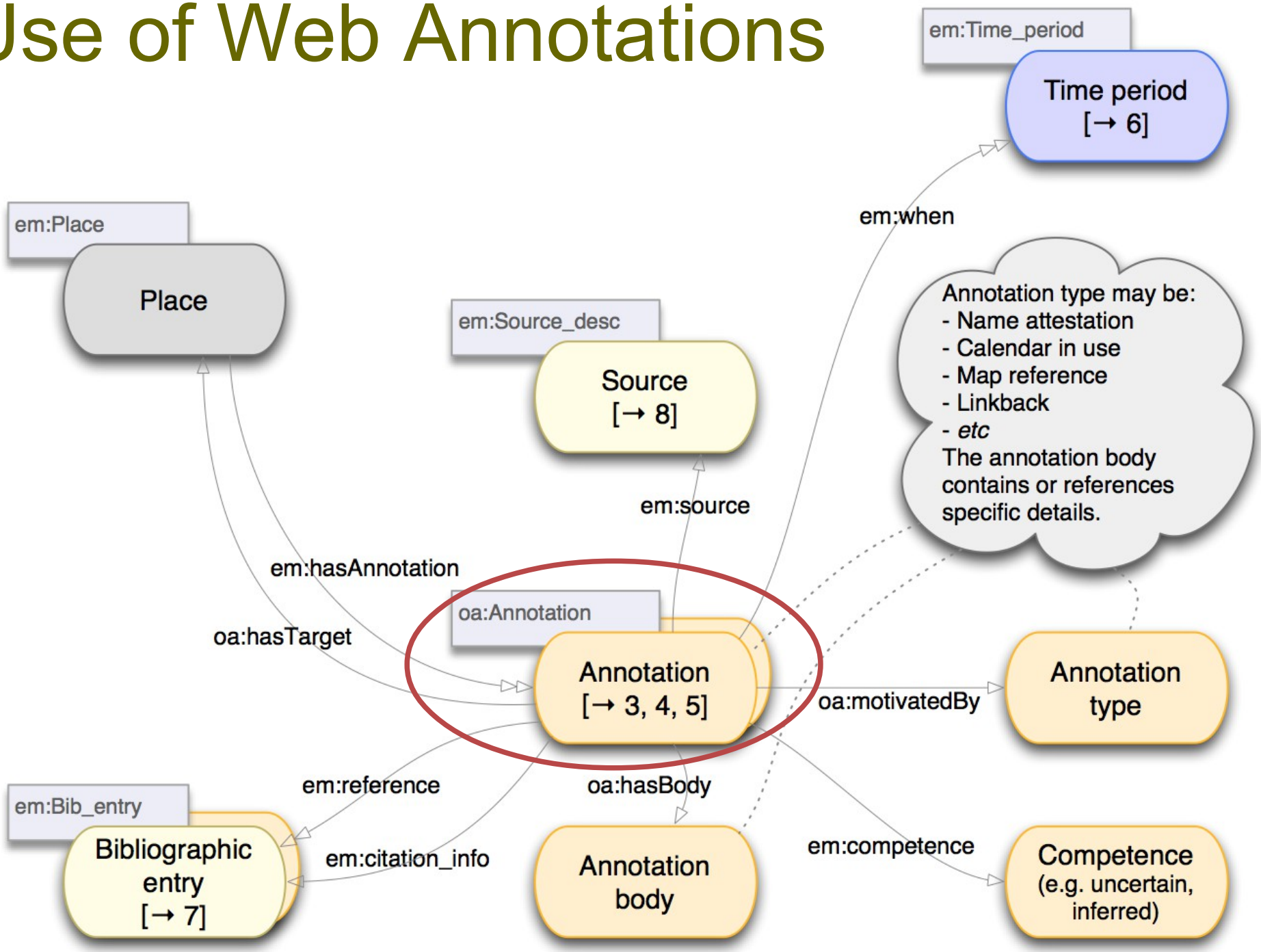
Setting: space *and* time



Qualified relations between places



Use of Web Annotations



Reflection on recording context

“People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated.”

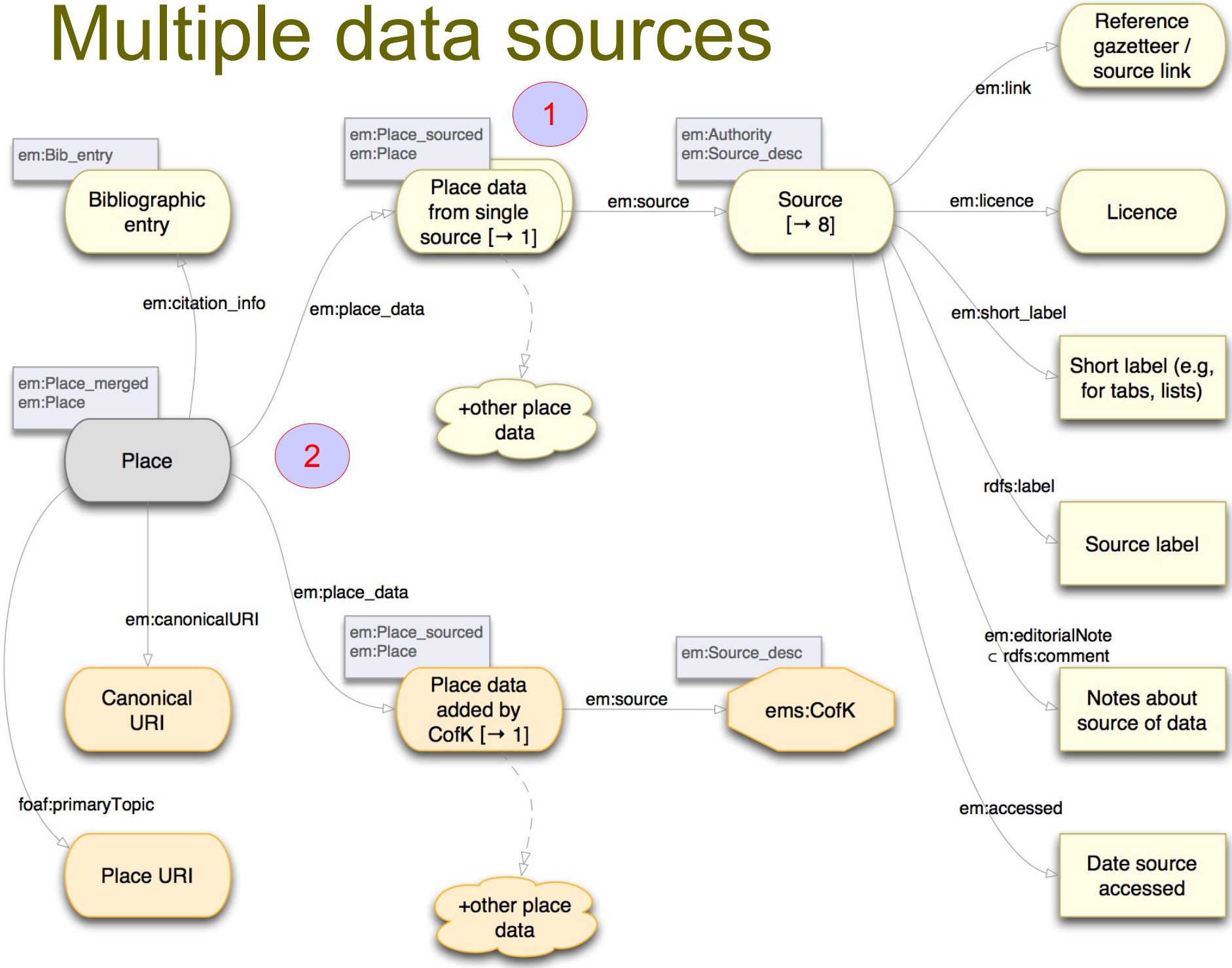
(Dan Brickley and Libby Miller)

- https://book.validatingrdf.com/bookHtml005.html

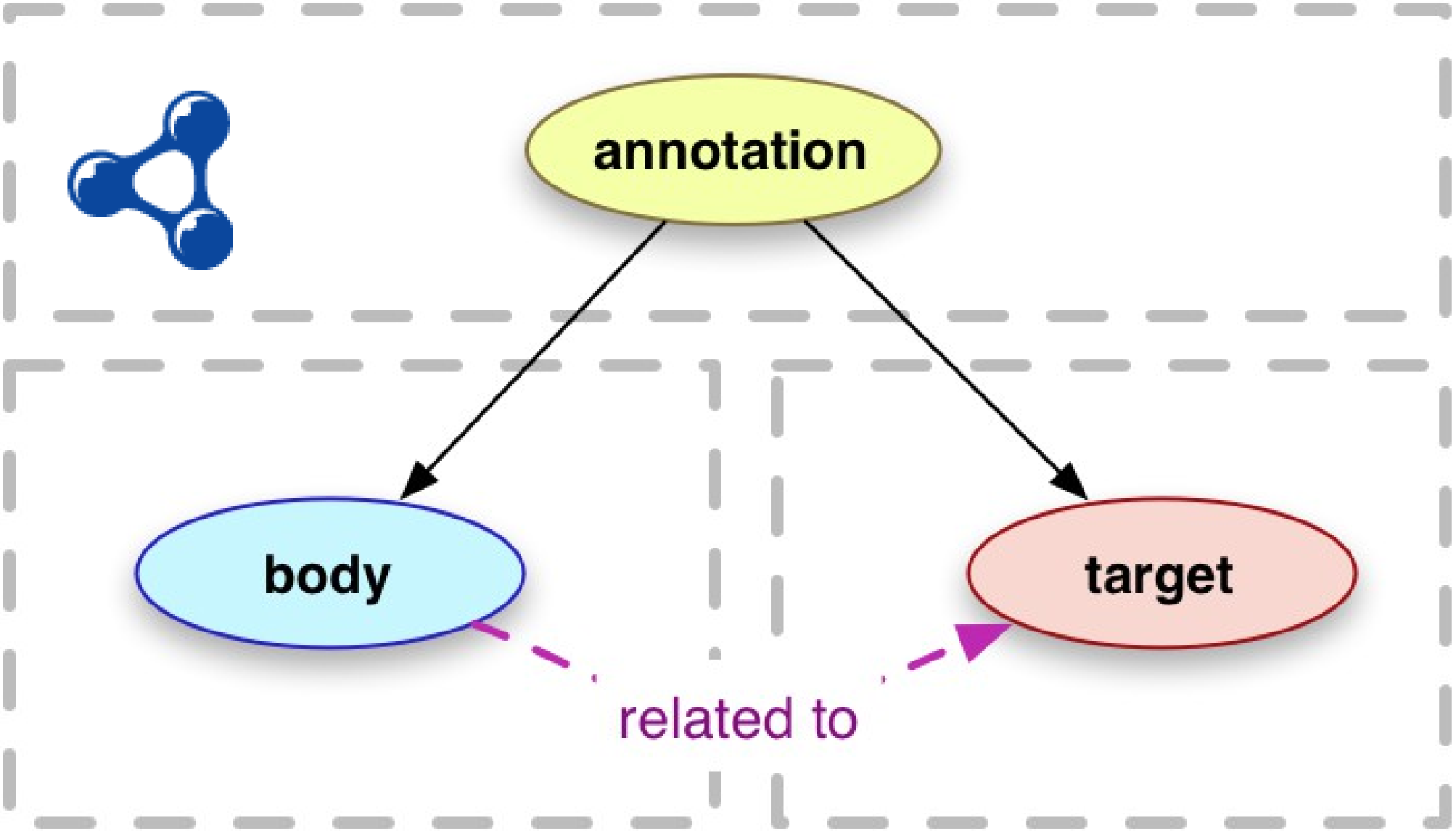
See also:

https://github.com/gklyne/notes/blob/master/20180307-qualified-property-links.md

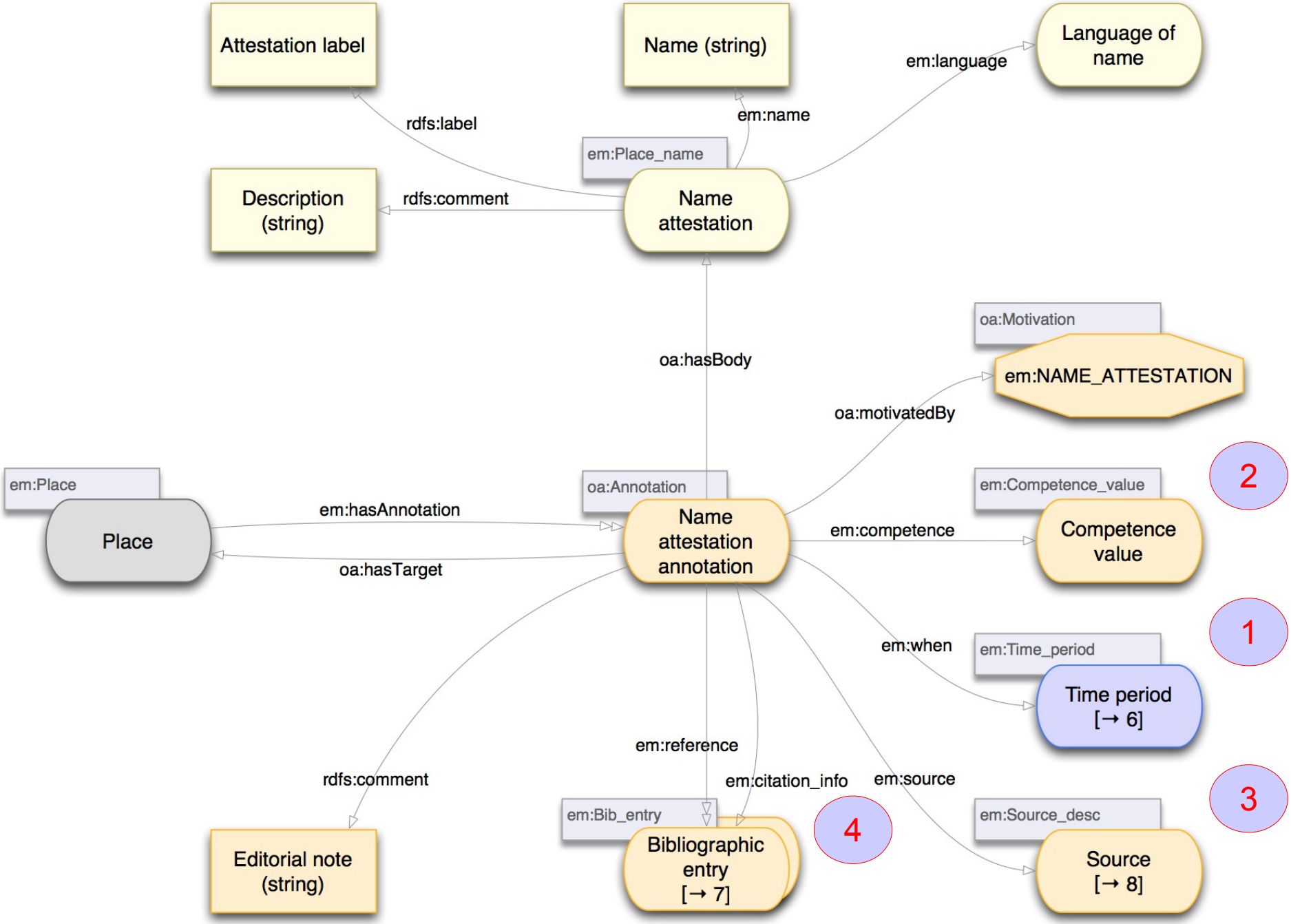
Multiple data sources



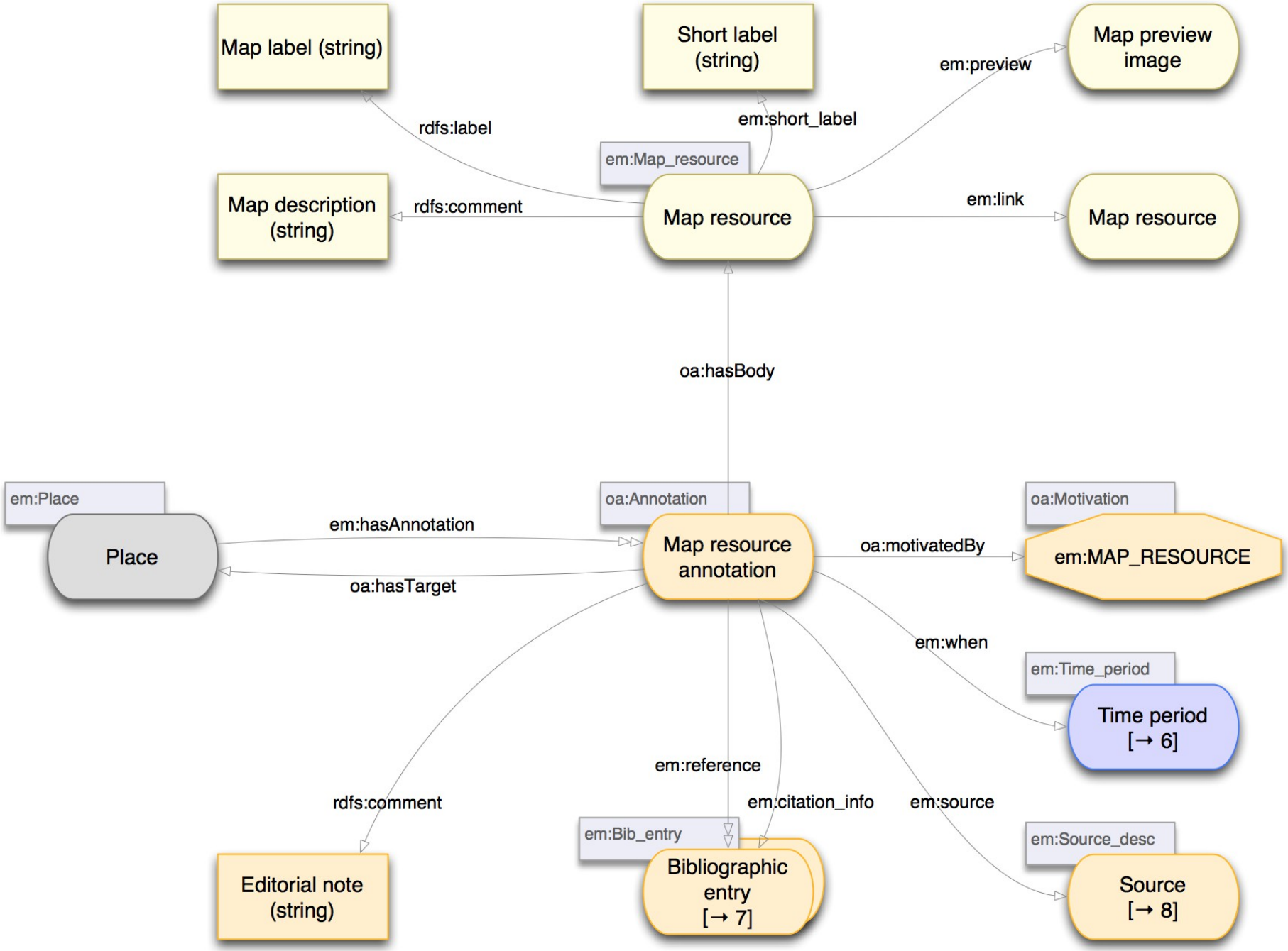
Web Annotation summary



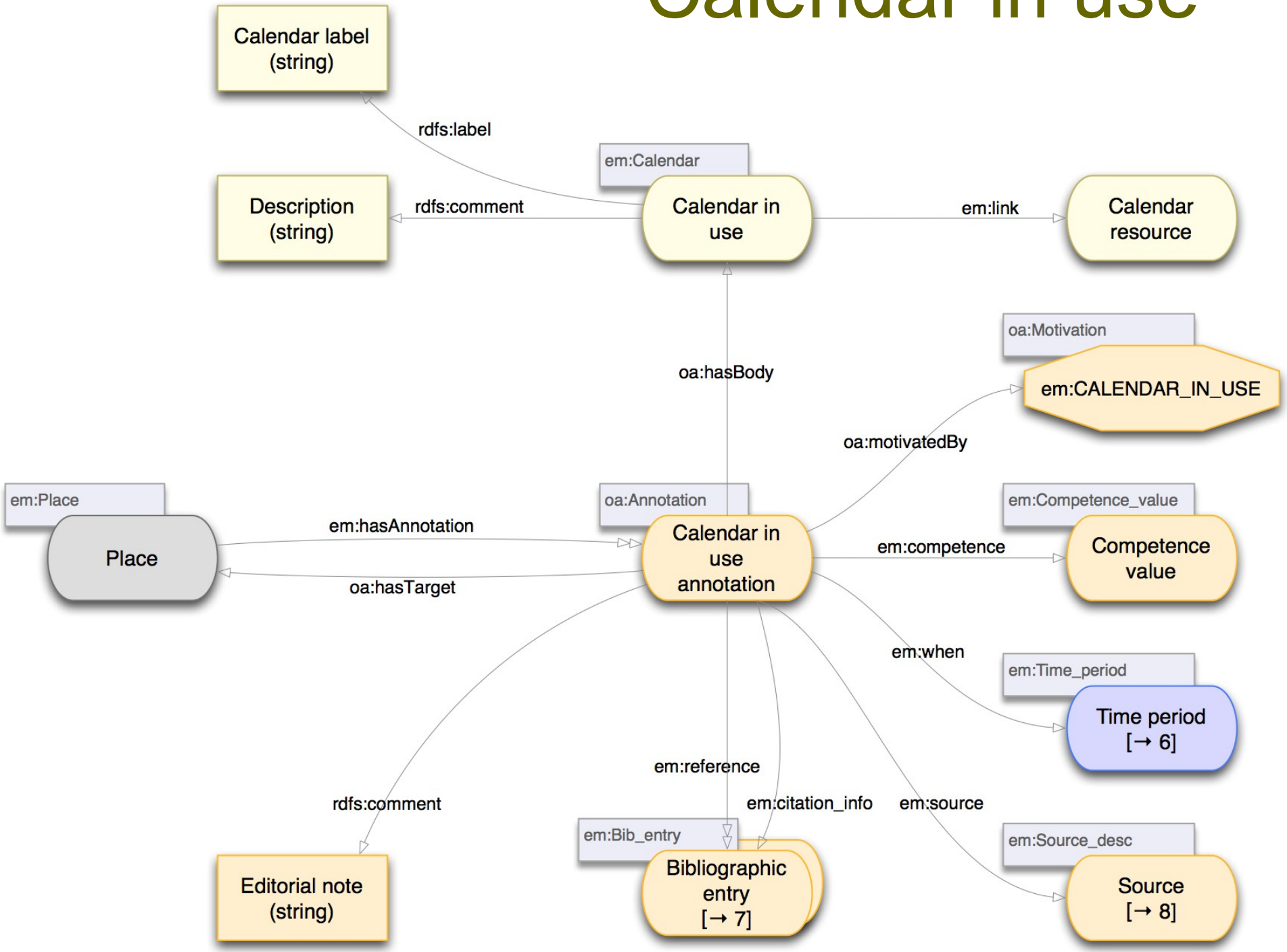
Additional details: name attestation



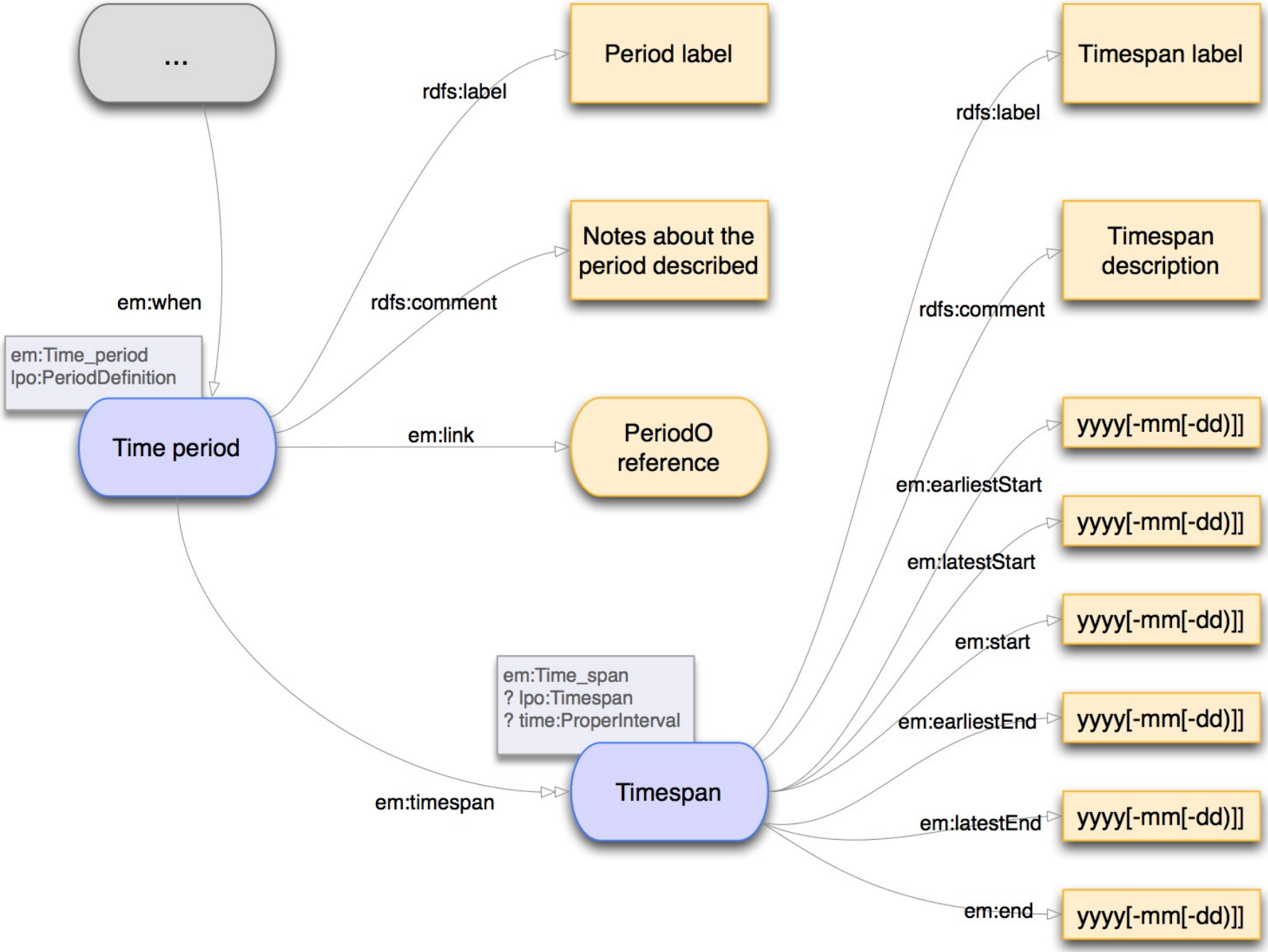
Map resource



Calendar in use



Time periods and time spans



EM Places model summary points

Context and provenance

- Expanded data model

Ontology sharing/reuse: parallel efforts

- Tyranny of timescales

Ontology sharing/reuse: local needs

- Specific application needs vs concept sharing

Data model references

Place model description files:

- https://github.com /culturesofknowledge/emplaces/tree/master/models

Place model diagrams:

- https://github.com /culturesofknowledge/emplaces/blob/master/models /PDFs/20180802-EMPlaces-data-model-multisource.pdf

Place model sample RDF data:

- https://github.com /culturesofknowledge/emplaces/blob/master/models /20180802-opole-example-multisourced.ttl

summer school

MELD

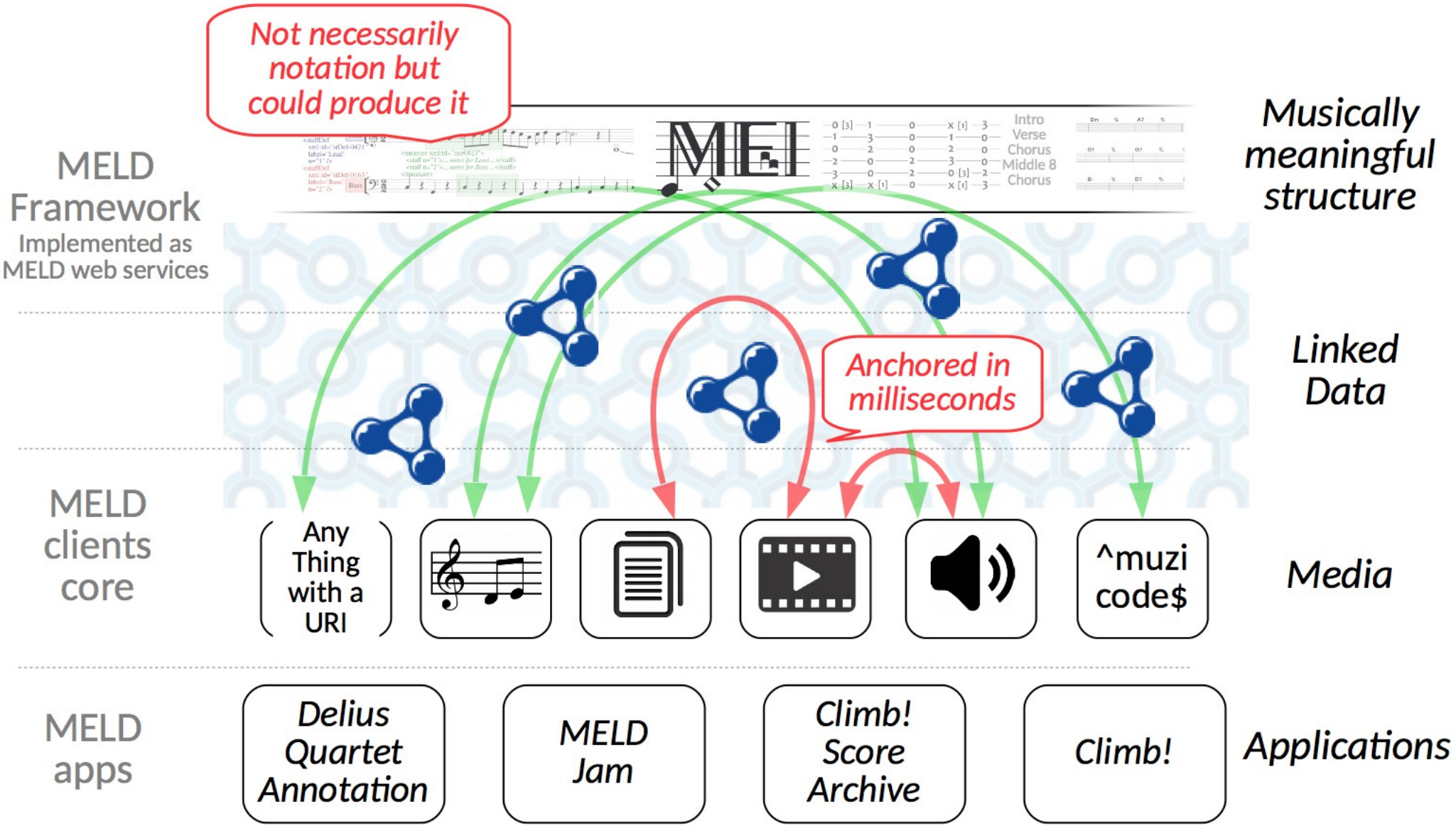
Music Encoding and Linked Data

Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

MELD

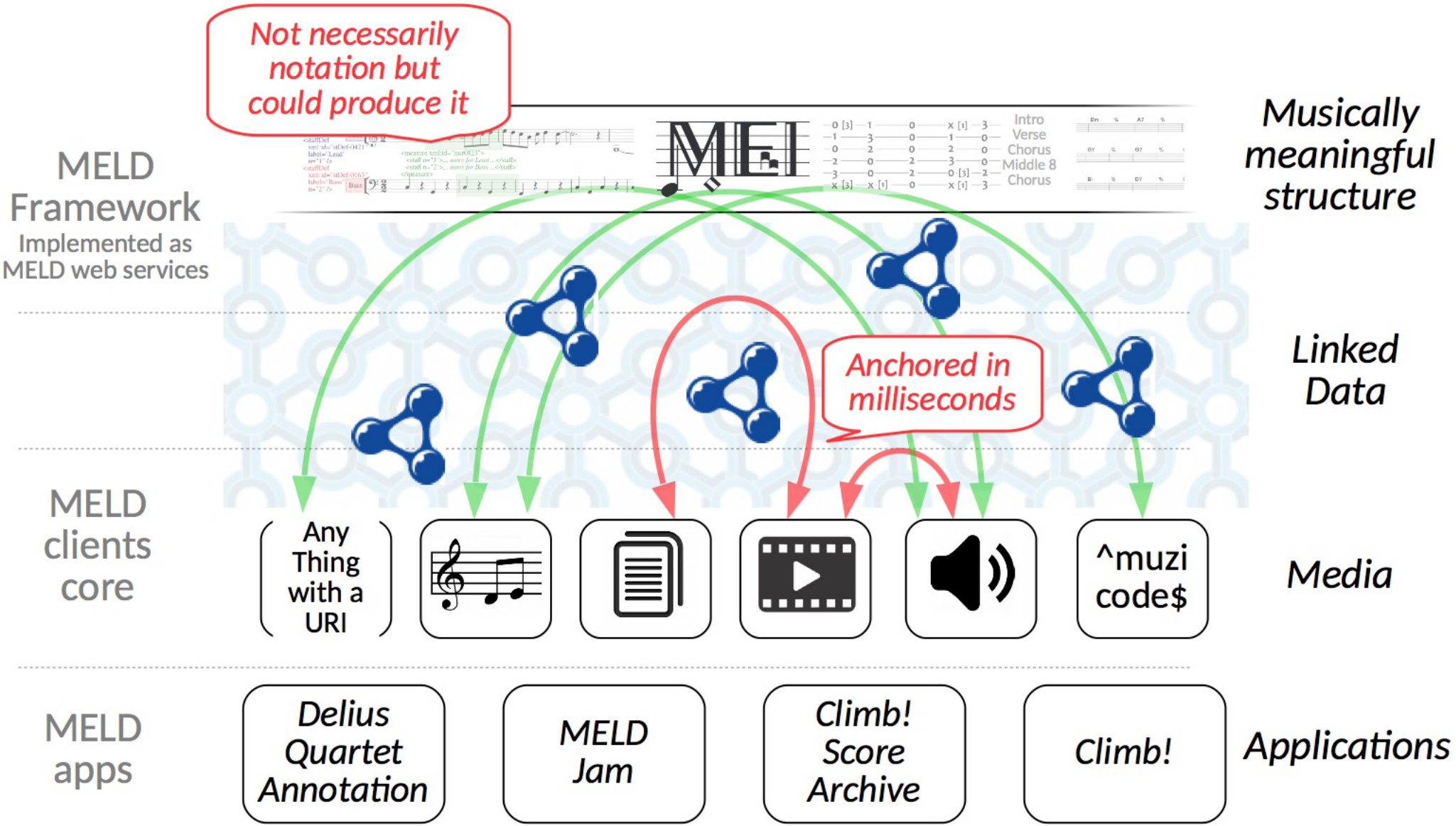
Music Encoding and Linked Data

Music Encoding and Linked Data (MELD) - overview



Kevin Page - kevin.page@oerc.ox.ac.uk – **05/03/2018**
MELD team: David Weigl (lead developer), David Lewis
Please reference: https://ora.ox.ac.uk/objects/uuid:945287f6-5dd3-4424-940c-b919b8ad2768

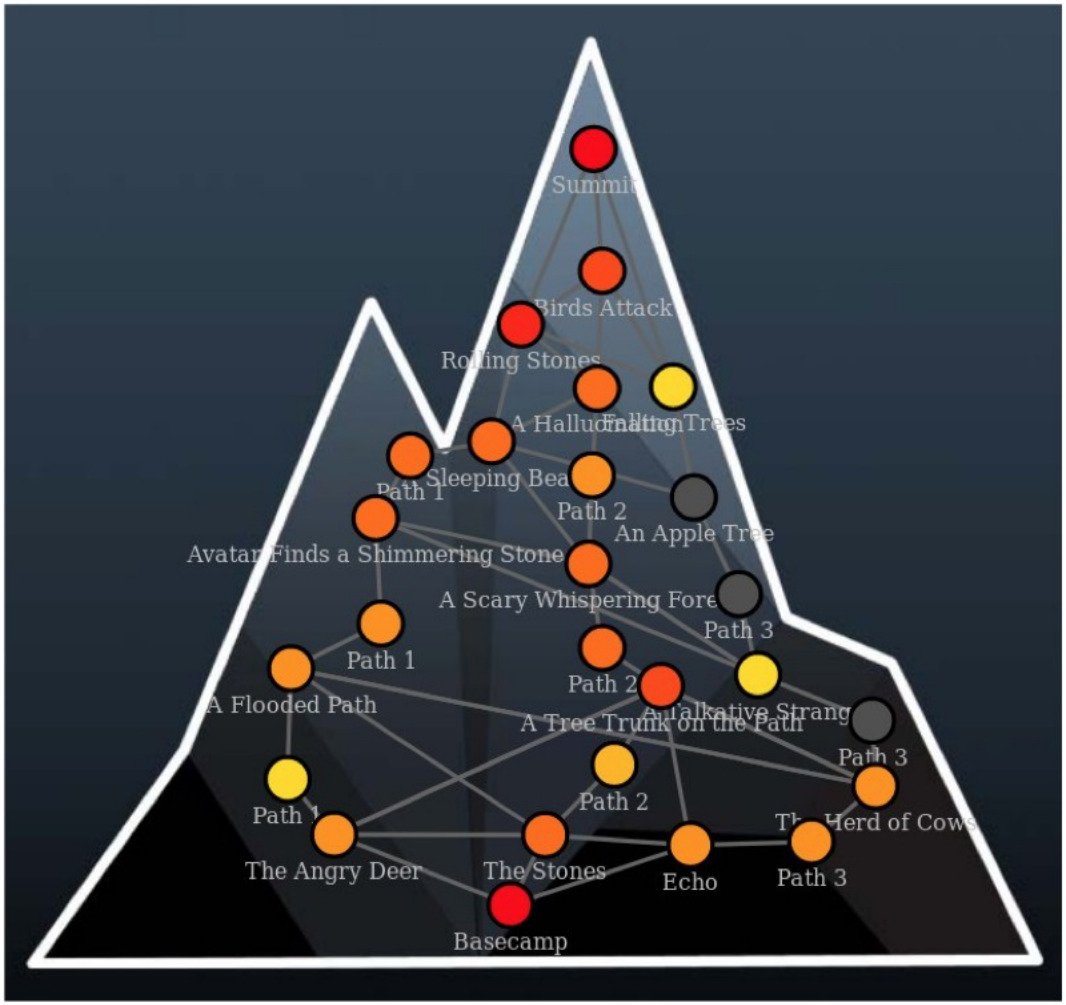
Music Encoding and Linked Data (MELD) - overview

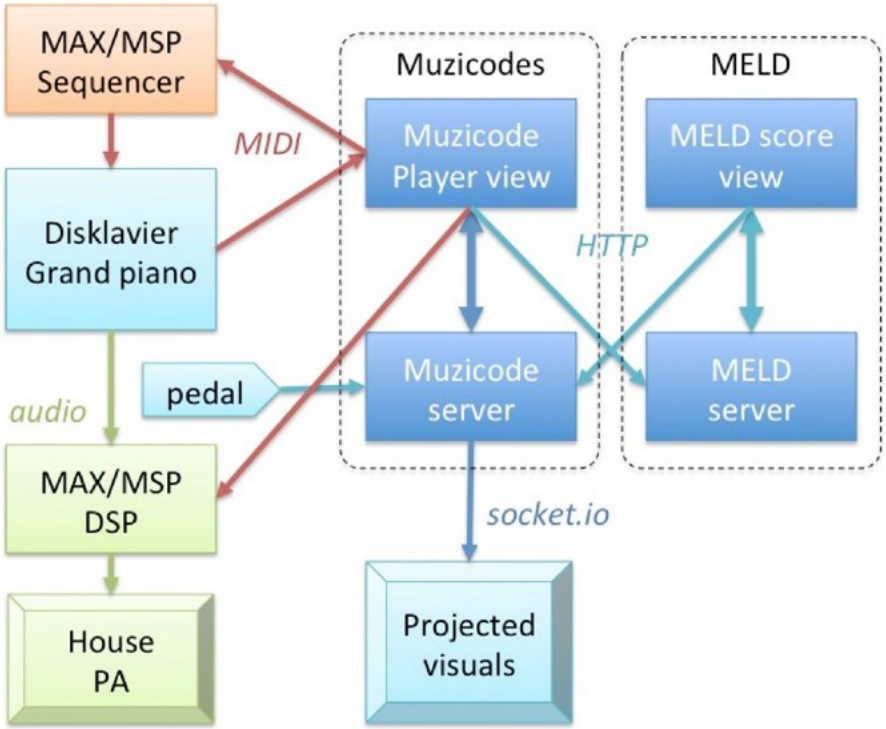


Kevin Page - kevin.page@oerc.ox.ac.uk – **05/03/2018**
MELD team: David Weigl (lead developer), David Lewis
Please reference: https://ora.ox.ac.uk/objects/uuid:945287f6-5dd3-4424-940c-b919b8ad2768

Climb!

"Climb!" non-linear musical work for Disklavier and electronics composed by Maria Kallionpää





Each section is encoded as an MEI part, with the paths, choices, narrative and actions encoded via MELD annotation

OXFORD e-Research CENTRE

(Kevin Page)

Climb! archive

Music Archive

Climb!

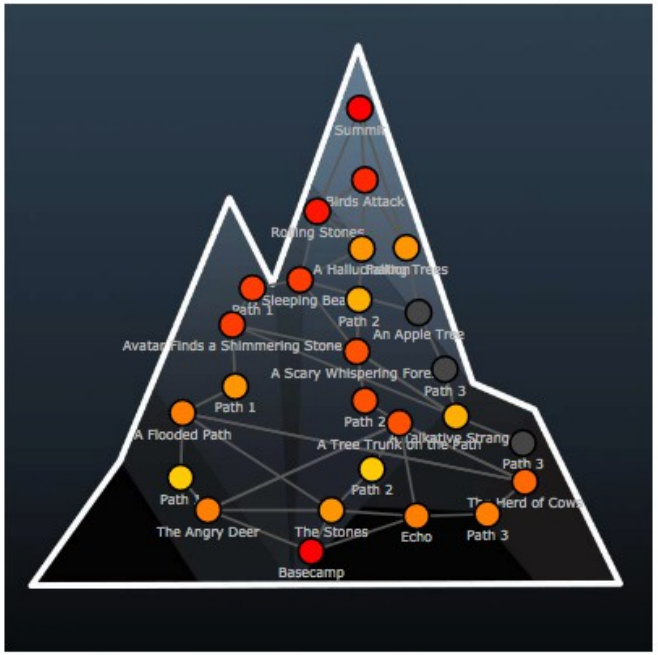
More about Climb!...

Select ☐ a Performance then click on its parts, or a click on and select ☐ a Part then play ⊙ the different versions available.

Performances and Playlists

☑ All Performances	☐ Dress rehearsal of Climb! *Maria Kallionpää*	☐ Premiere performance of Climb! *Maria Kallionpää*
☐ Second performance of Climb! *Maria Kallionpää*	☐ Climb! at All Your Bass *Anne Veinberg*	☐ Climb! at All Your Bass (2) *Zubin Kanga*
☐ Climb! at RCC 2018 (1) *Maria Kallionpää*	☐ Climb! at RCC 2018 (2) *Maria Kallionpää*	☐ Climb! in Finland (1) *Maria Kallionpää*
☐ Climb! in Finland (2) *Maria Kallionpää*	☐ Climb! at FAST Industry Day *Maria Kallionpää*	☐ Climb! test at FAST Industry day *Maria Kallionpää*
New Playlist		

Parts (sections) List Map



http://music-mrl.nott.ac.uk/1/archive/explore/Climb

SOFA Remixer

Numbers into Notes

1. Generate

The first stage in this interactive demonstration is to **generate** a sequence of numbers using an algorithm that might have been run on the Analytical Engine. In stage 2 you **reduce** these numbers to a smaller range of numbers, using 'clock arithmetic'. Then in stage 3, you can visualize and **play** the number sequence interactively, and explore the mapping of the numbers to musical notes. If this is your first visit, try the green buttons. NB This program has been developed and tested in Chrome and Firefox browsers.

Fibonacci

Each number is the sum of the previous two. If we start with 0, 1 we get the sequence commonly known as the Fibonacci numbers 0, 1, 1, 2, 3, 5, 8, 13, ... (could also start 1, 1). If we start with 2, 1 we get the Lucas numbers 2, 1, 3, 4, 7, 11, 18, ... (could also start 1, 3). The sequences are described mathematically by the relation $F_{k,n+1} = kF_{k,n} + F_{k,n-1}$ where k = 1.

Fibonacci *n0:* 0 *n1:* 1 *k:* 1

Powers

The sequence of powers of 2 is 2^0, 2^1, 2^2, 2^3, ... i.e. 1, 2, 4, 8, 16, 32 etc. Here you can generate the sequence of powers of 2 or any other number, known as the *base*.

Powers *base:* 2

Other Algorithms

How might Ada Lovelace have calculated these?

Primes | Pi | Golden Ratio

Random 84

Arithmetic progression

The sequence starting at *a* and adding *d* each time, so 0 and 1 gives 0,1,2,3,4, ... Useful to illustrate mappings of numbers to notes.

Arithmetic Progression

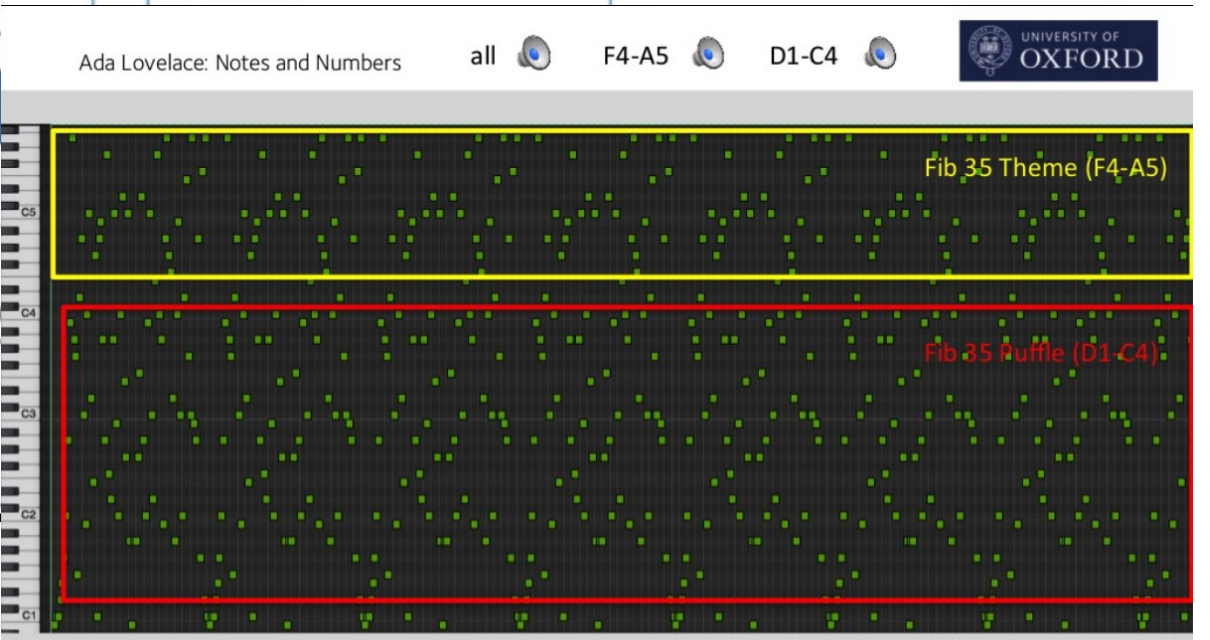
a: 0 *d:* 1

Triangular and Factorials

Triangular numbers are the sequence of sums 1, 1 + 2, 1 + 3, 1 + 2 + 3 + 4, ...

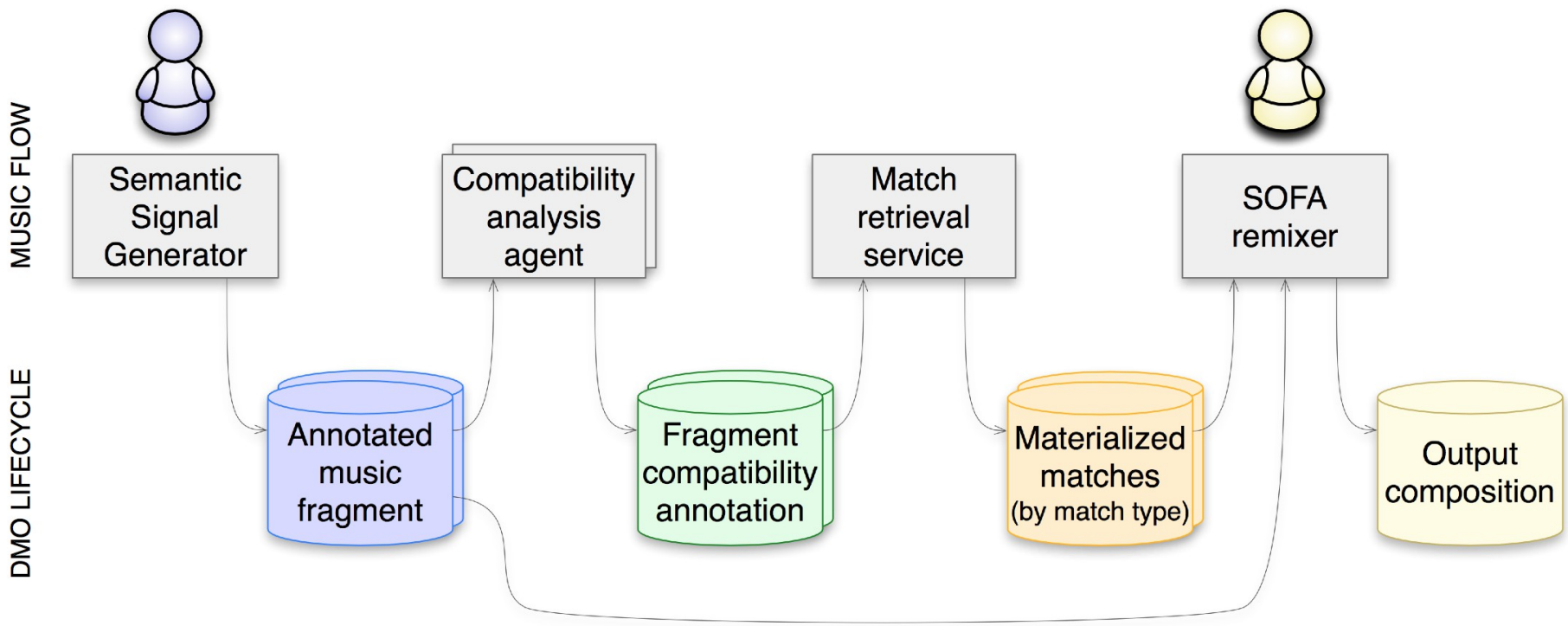
Factorial numbers are the sequ of products 1, 1 x 2, 1 x 2 x 3, x 3 x 4, ...

Triangular | Factorial



https://www.slideshare.net/dder/ada-lovelace-numbers-and-notes

SOFA Remixer
(SOFA Ontological Fragment Assembler)

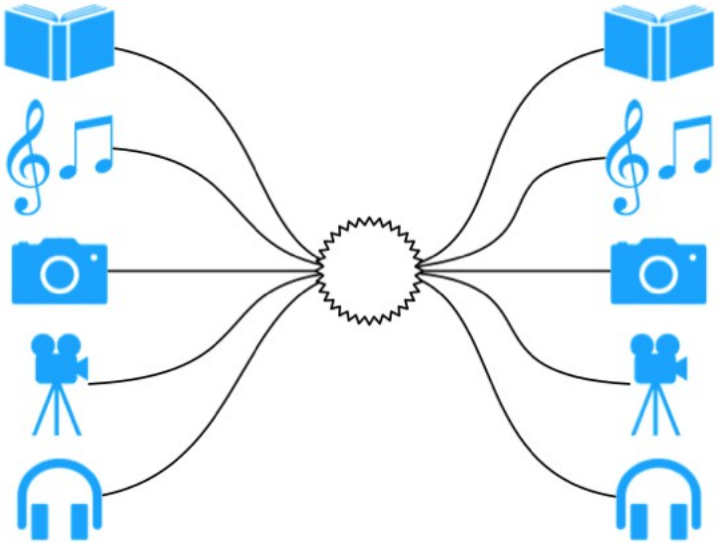


SOFA Remixer: sample interface

Linking to music

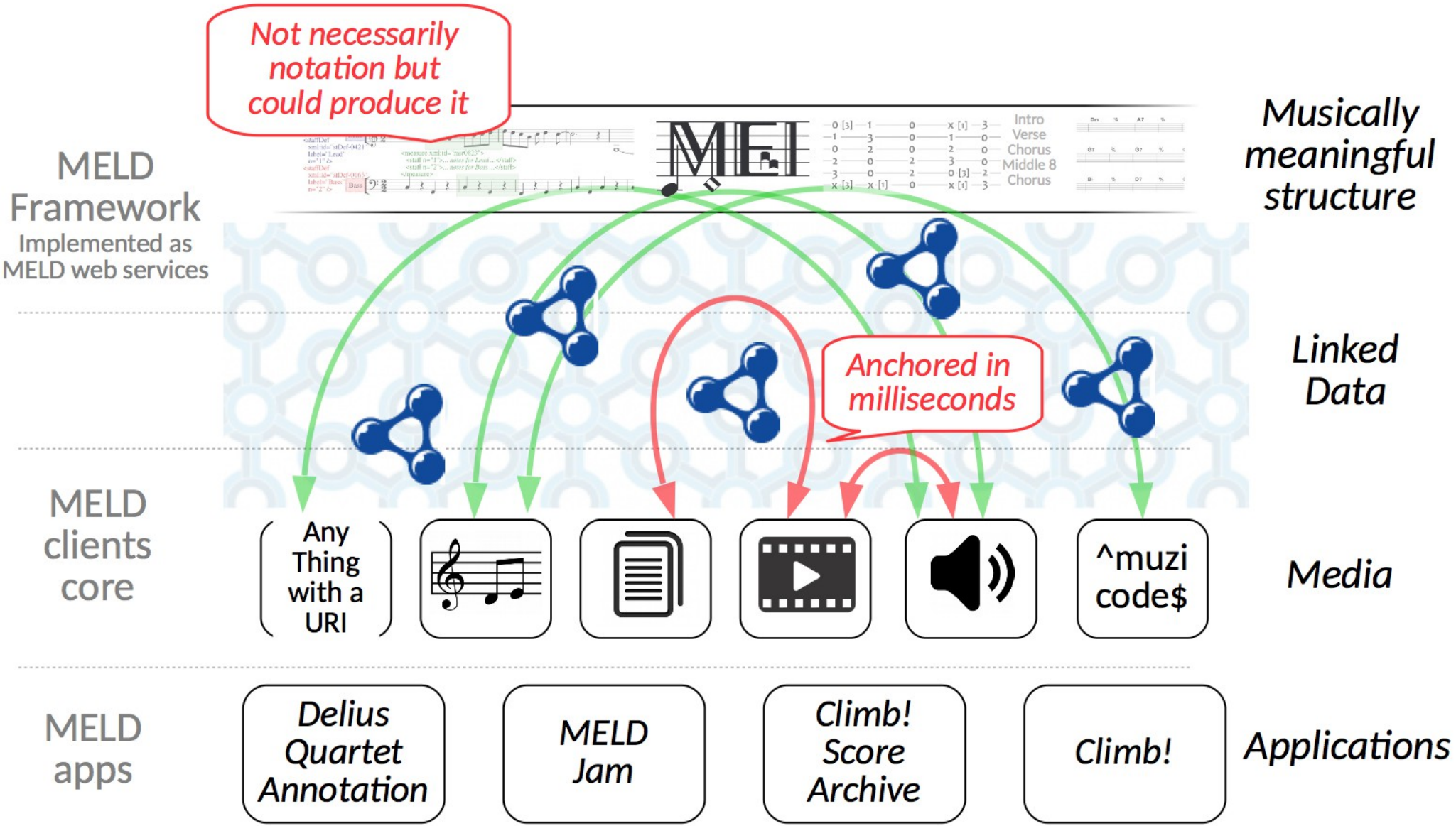
Music can be encapsulated in multiple different digital formats. To illustrate musicological insights using these digital encodings we need to: (i) relate the insight or observation to the media; (ii) interact with the media through actions and reactions which reflect the insight

Medium	Trigger Action	Realised action
Text (TEI)	Select anchor in text	Make text appear, Jump to region, Highlight region
Music notation (MEI)	Select highlighted notation elements	Make score appear, Scroll to region, Highlight region
Image (JPEG/IIIF)	Activate highlighted region	Make image appear, Switch page, Zoom to region, Highlight region
Video (MPEG)	Play or skip to timepoint	Make video appear, Jump to timepoint, Trigger playback
Audio (MPEG)	Play or skip to timepoint	Show audio controls, Jump to timepoint, Trigger playback



(Kevin Page)

Music Encoding and Linked Data (MELD) - overview

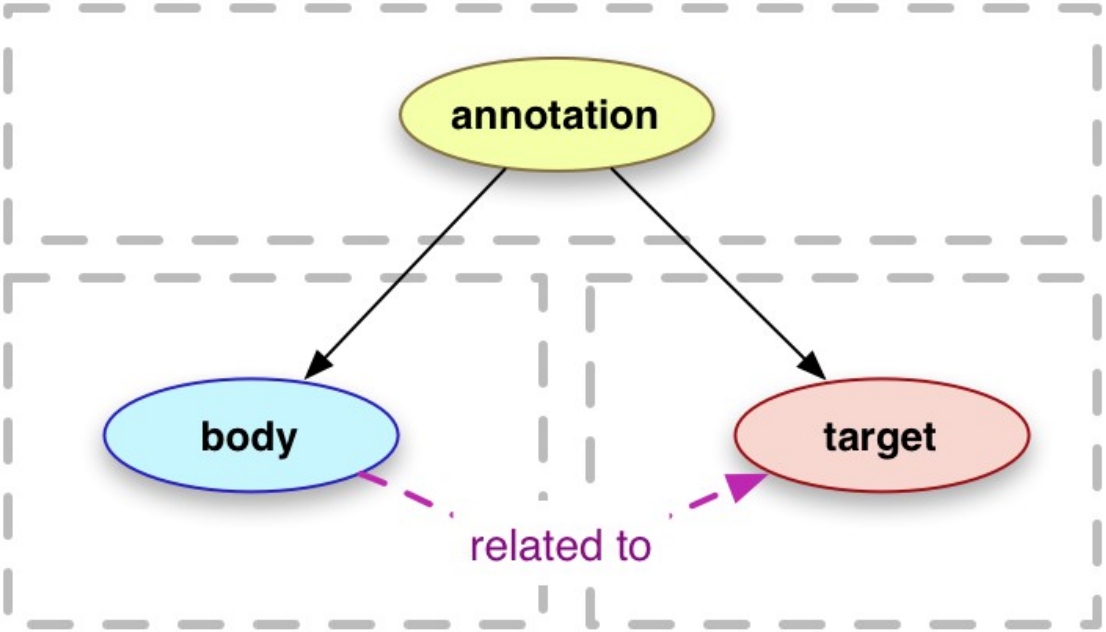


Kevin Page - kevin.page@oerc.ox.ac.uk – **05/03/2018**
MELD team: David Weigl (lead developer), David Lewis
Please reference: https://ora.ox.ac.uk/objects/uuid:945287f6-5dd3-4424-940c-b919b8ad2768

MELD:
Building on existing standards

RDF linked data

Web Annotations



Linked Data Platform (LDP)

- SoLiD

Music Encoding Initiative (XML)

Questions / discussion

(Reminders)

- Linked data connecting other formats (esp. XML)
- MELD
  - LDP + Web Annotations + musical structure...
- SoLiD (re-decentralizing the web)
  - Separation of data custody from web applications
- EM Places
  - Time-constrained attributes and relations
  - Extending basic data model to capture context and provenance

summer school

Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

The field of Digital Humanities explores the ways in which digital technologies can be used to support and enhance humanities research. This session will be about my experiences of using linked data on the Web, in the form of RDF and associated Semantic Web technologies, in some Digital Humanities applications.

I shall initially discuss some ways that application of Semantic Web technologies in humanities may differ from use of the same in scientific or other common applications. I shall then describe a couple of applications I've been working on recently, hopefully showing how affordances of linked data can support some of the particular requirements of humanities data.

///

Linked Data in Digital Humanities

Graham Klyne

XML Summer School 2019
Trends and Transients

The field of Digital Humanities explores the ways in which digital technologies can be used to support and enhance humanities research. This session will be about experiences of using linked data (RDF) and associated Semantic Web technologies in Digital Humanities applications.

I shall discuss some ways that application of Semantic Web technologies in humanities may differ from use of the same in scientific or other common applications. I shall then describe a couple of applications I've been working on recently, hopefully showing how affordances of linked data can support some of the particular requirements of humanities data.
///

I'll start with a brief summary of my background, to give an idea of my biases and prejudices…

I work as a research software engineer at Oxford University, in the multidisciplinary e-Research centre. Before I came to Oxford I was a commercial software developer, working on applications ranging from industrial process control, 3D motion capture for medical use, computer-aided mechanism design, computer networking, financial trading systems and home automation. While working for a now defunct start-up, I was involved in defining IETF standards for email enhancements and content negotiation, which led (by a roundabout route) to participation in the W3C working group that was defining RDF, where RDF is one of the technical standards that underpins linked data on the web.

When I came to Oxford, I joined David Shotton's Image Bioinformatics Research Group, to work on using Semantic Web technologies in a metadata-rich biological image database. At the time, I believe David's was the only group in Oxford doing any kind of Semantic Web related research, and I was keen to gain experience actually using the technologies I'd helped to specify. When I joined David's group, I expected it would be for just a short period of working in the university. But, somehow, 15 years later I'm still here. I promise it's not for the money!

During my time at Oxford, I've worked on a range of applications from bioinformatics to classical art to scientific workflow description to digital music to historical data. A common underlying theme has been management of research data, which I believe raises some challenges that are different from other applications, by not always being susceptible to full definition until data collection is well under way. In my opinion, it is this open-ended nature of research data that makes linked data particularly applicable.

I also work, in a personal capacity, on a linked data notebook tool [Annalist] (http://annalist.net/), which I've used extensively for prototyping linked data model designs.

///

Humanities

"Humanities are academic disciplines that study aspects of human society and culture."

— *Wikipedia (https://en.wikipedia.org/wiki/Humanities)*

So what are Humanities, and what makes humanities data different?

Broadly conceived, the humanities include ancient and modern languages, literature, philosophy, history, human geography, law, politics, religion, and art. My own interactions have been primarily with aspects of history and art (which are topics in which I claim no expertise!), so my comments here may be less applicable to other humanities fields.

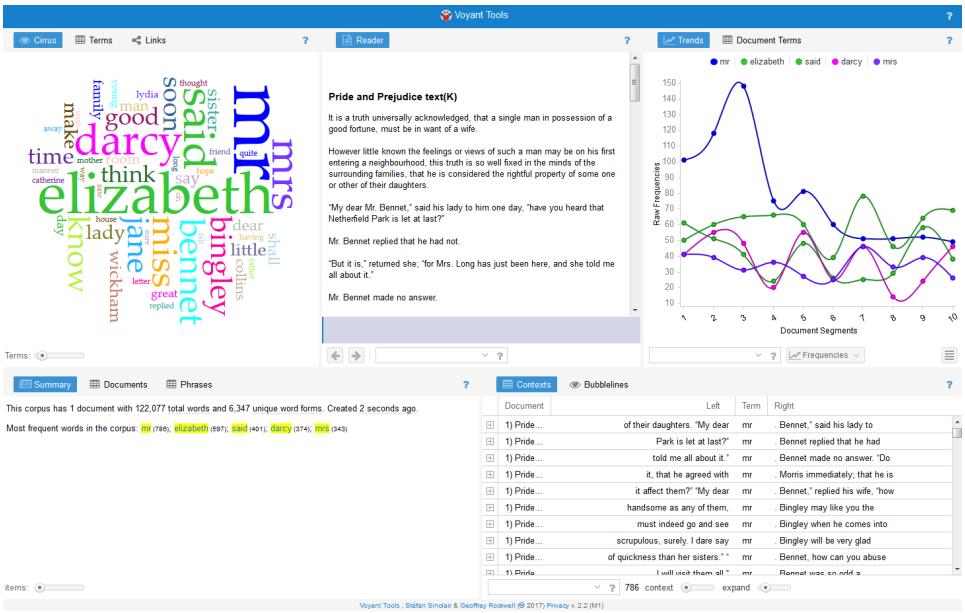
Compared with scientific research, which focuses substantially on observable and reproducible measures of the world around us, I perceive humanities research to be much more concerned with outputs that involve significant levels of human interpretation or reception.

This in turn depends crucially (among other things) on establishing a context within which interpretation or reception can take place. Different people may view the same information from different perspectives, different background contexts, and come to quite different conclusions.

Often (particularly when trying to understand historical events) the available information is sparse, and the best that can be achieved is to create a compelling narrative that accords with the available facts. There is plenty of scope for Disagreement! I think the point is not necessarily to arrive at a single unarguable conclusion, but to engender a deeper understanding of the material under study.

(I was recently involved in a discussion about reproducible research in the humanities, where I asked what reproducibility in humanities actually means. One respondent suggested that it might be the ability to consistently engender an attitude or conclusion in the mind of a person receiving the research.)
///

Digital Humanities



Text of Jane Austen's Pride and Prejudice in textual analysis program Voyant Tools - *https://commons.wikimedia.org/wiki/File:Pride_and_Prejudice_in_Voyant_Tools.png*

So what is Digital Humanities?

The field of Digital Humanities, which grew out of “Humanities Computing”, explores the ways in which digital technologies can be used to support and enhance humanities research.

There is some debate about whether “Digital Humanities” deserves to be recognized as a field in its own right - some say it’s all just “Humanities”. I don’t take a position on this, but I readily use the term as a shorthand to help contextualize my work.

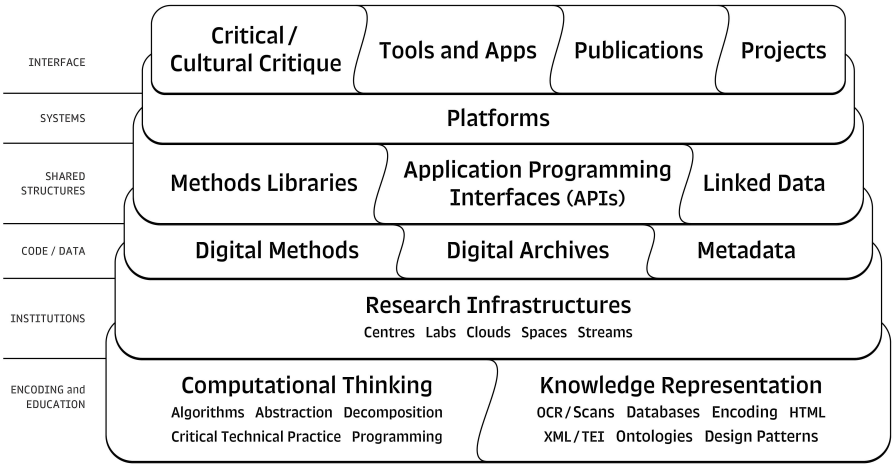
Digital Humanities supports researchers with a range of digital tools: digitized archives, text analysis, digital mapping, network analysis, data mining, data visualization, image analysis, audio analysis, 3D modelling, cataloging, and general research data management are all used in humanities projects..

The image here is an example lifted from Wikipedia, that shows panels doing things computers do easily: presenting text, visualization, graphing, counting, indexing, etc., and presenting results for human interpretation.

One of the early fruits of Digital Humanities research was the Text Encoding Initiative (TEI), which is an XML-based format for representing and tagging digital texts. More recently, a similar initiative, the Music Encoding Initiative (MEI) has emerged: this is an XML-based format for digitally representing musical structure. I shall refer back to this later in this talk.

///

Digital Humanities Stack?



Digital Humanities Stack
- https://commons.wikimedia.org/wiki/File:Digital_Humanities_Stack_(from_Berry_and_Fagerjord_2017-_18).jpg

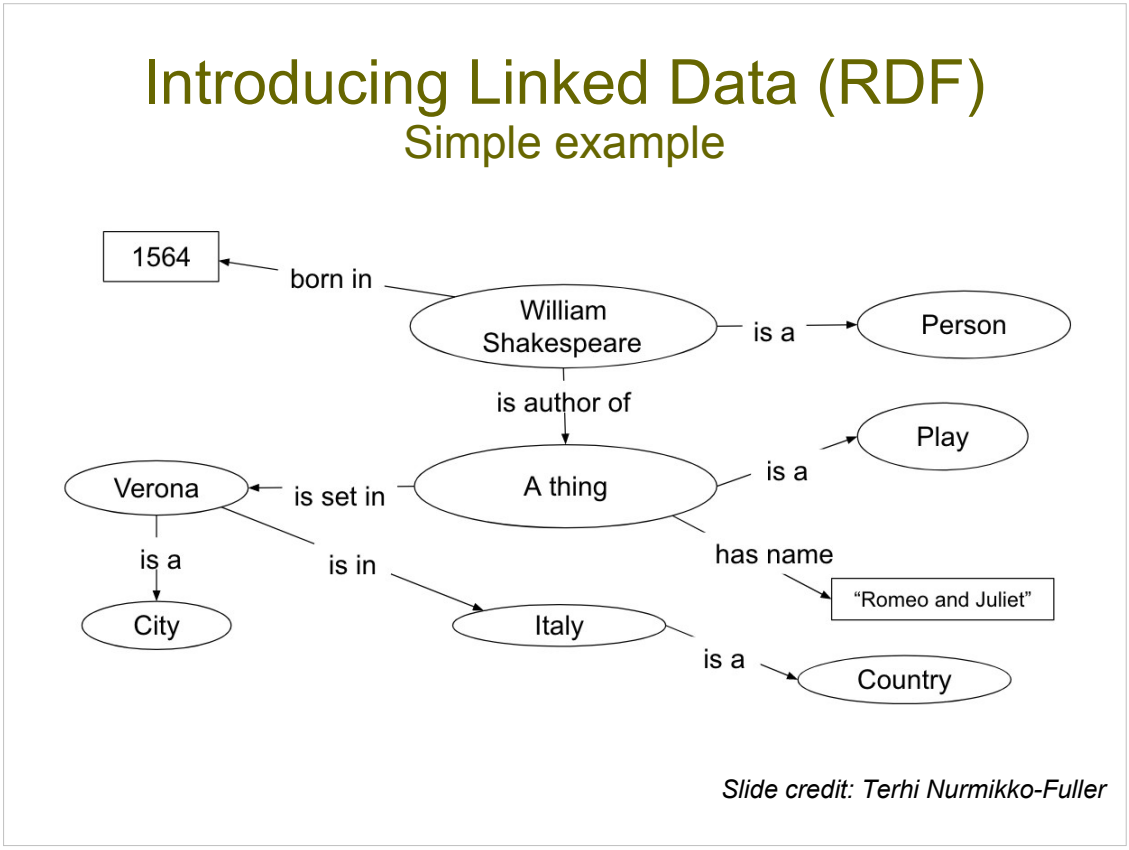
from Berry and Fagerjord, *Digital Humanities: Knowledge and Critique in a Digital Age*

By way of a slight diversion, this “Digital Humanities Stack” is something I came across while preparing these slides.
It is referenced by the Wikipedia page on Digital Humanities.

I'm not entirely sure what to make of it, but I notice that it substantially consists of conventional digital techniques with interpretive and communication activities layered on the top.

Also, the construction of the stack itself suggests a humanistic rather than a purely technical perspective (e.g. the way it builds up from “computational thinking” and “knowledge representation” at its base). I think this different perspective might conceivably be an interesting aspect of Digital Humanities compared with other uses of digital technologies..
///
Referenced by:

- [Wikipedia](https://en.wikipedia.org/wiki/Digital_humanities)

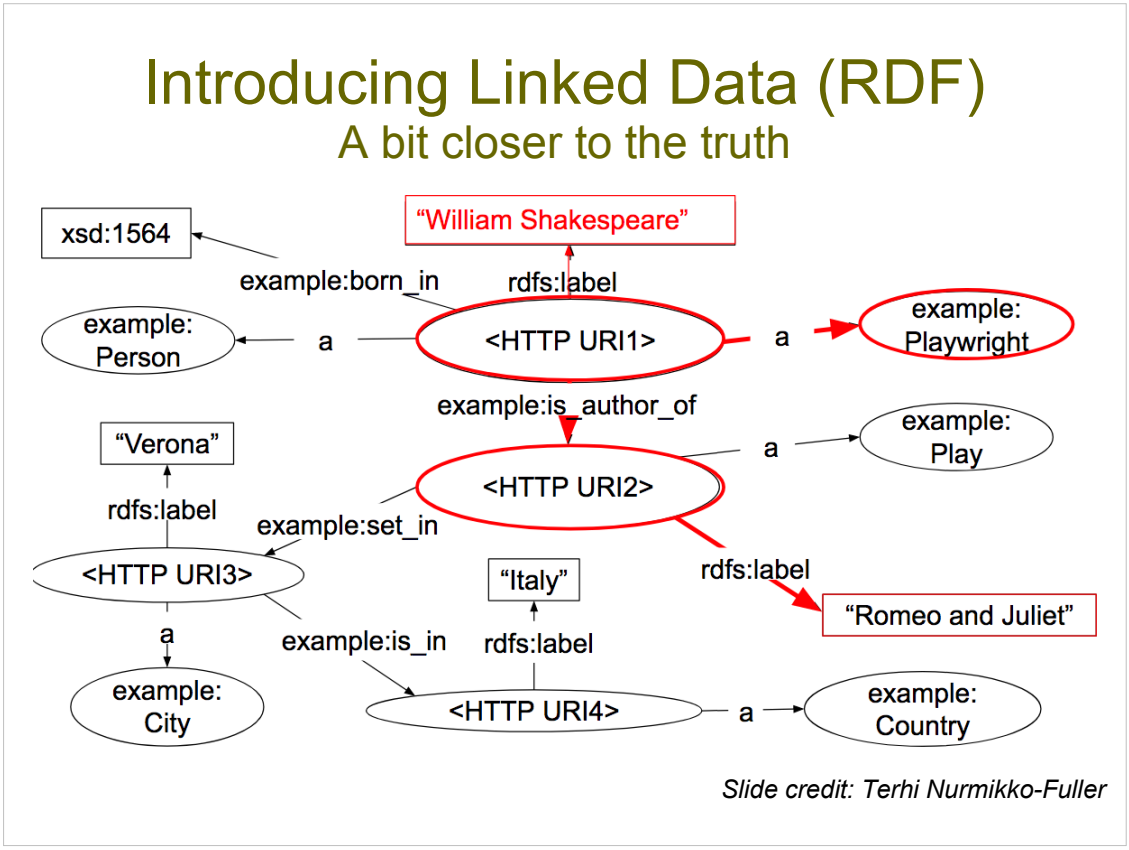


Turning now to RDF linked data... (ask who is not familiar)

I'm not going to attempt a detailed technical description of RDF linked data: this is intended to be just-enough background for you to be able to follow what I'll talk about later.

Linked data focuses on entities, and relations between them. So in this (simplified) example, the elements labelled "William Shakespeare" and "A thing" correspond to the playwright William Shakespeare, and his play Romeo and Juliet", respectively. The relation between them described here is that William Shakespeare is the author of Romeo and Juliet.

Each connecting line in the graph corresponds to a relation between two concepts, or between a concept and a literal value. When describing linked data, people often talk about "triples": taken together, the two related concepts and the type of relation between them constitute such a "triple".
///



In actual RDF linked data, the concepts and relation types are represented as URIs, which (preferably) are HTTP URIs that can be retrieved on the web.

This means that:
---
1. we have unique identifiers for every concept and relation type that is referenced, and
---
2. further information about any concept or relation type can (in principle) be retrieved from the Web using it's URI. Only the literals have predefined meaning, and that meaning is generally very limited; e.g., strings, numbers, dates, etc.

This use of URIs to identify things is what makes RDF be truly "of the Web", and not just another graph database that can be accessed over the Web.

And that's pretty much it. There are plenty of technical details, which some of you will have learned about in the linked data strand of this summer school. But this connecting of concepts with relations, where concepts and relation types are identified by URIs, is the essential nature of RDF linked data.
///

Introducing Linked Data (RDF)

The painful reality?

“People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated.”

“While you can avoid RDF, it is harder to avoid complicated data and complicated computer problems. RDF brings together data across application boundaries and imposes no discipline on mandatory or expected structures.”

(Dan Brickley and Libby Miller)

- https://book.validatingrdf.com/bookHtml005.html

I particularly like this quote from Dan Brickley and Libby Miller.
---
^^^

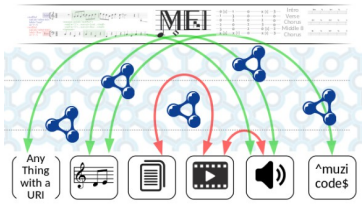
The rest of this talk could be viewed in this light: describing how we can use a “painfully simplistic” structure to represent complicated real-world problems, particularly as they arise in humanities research.
///

About this talk

1. EM Places: Early Modern historical places

The Letter Catchers by Annaliese Stoney for Cultures of Knowledge (with inspiration from Pier Francesco Mola)

2. MELD: Music Encoding and linked data



Having described some introductory concepts, the rest of this talk will look more closely at two of the digital humanities linked data applications I've worked on recently, hopefully showing how the affordances of linked data are suited (or not) to dealing with the data representation and processing issues that arise.

---

1. EM Places: recording Early Modern historical place information ([Cultures of Knowledge (CofK) project](https://www.history.ox.ac.uk/cultures-knowledge)). Cultures of Knowledge has created the Early Modern Letters Online (EMLO) database. The EM Places project is building an online reference resource of place information from EMLO. I shall discuss the development of a data model that captures historical contextual information, and also provides for incorporation of up-to-date information from other sources.

---

2. MELD: Music Encoding and Linked Data ([Fusing Audio and Semantic Technology (FAST) project](http://www.semanticaudio.ac.uk/)). The FAST project is exploring the intersection of semantic and music-related technologies. I shall talk about the MELD framework, conceived by my colleagues David Weigl and Kevin Page, which uses linked data in applications that establish and act upon connections between musical structure, music-related media and other data.

///

summer school

EM Places

Early Modern historical place data

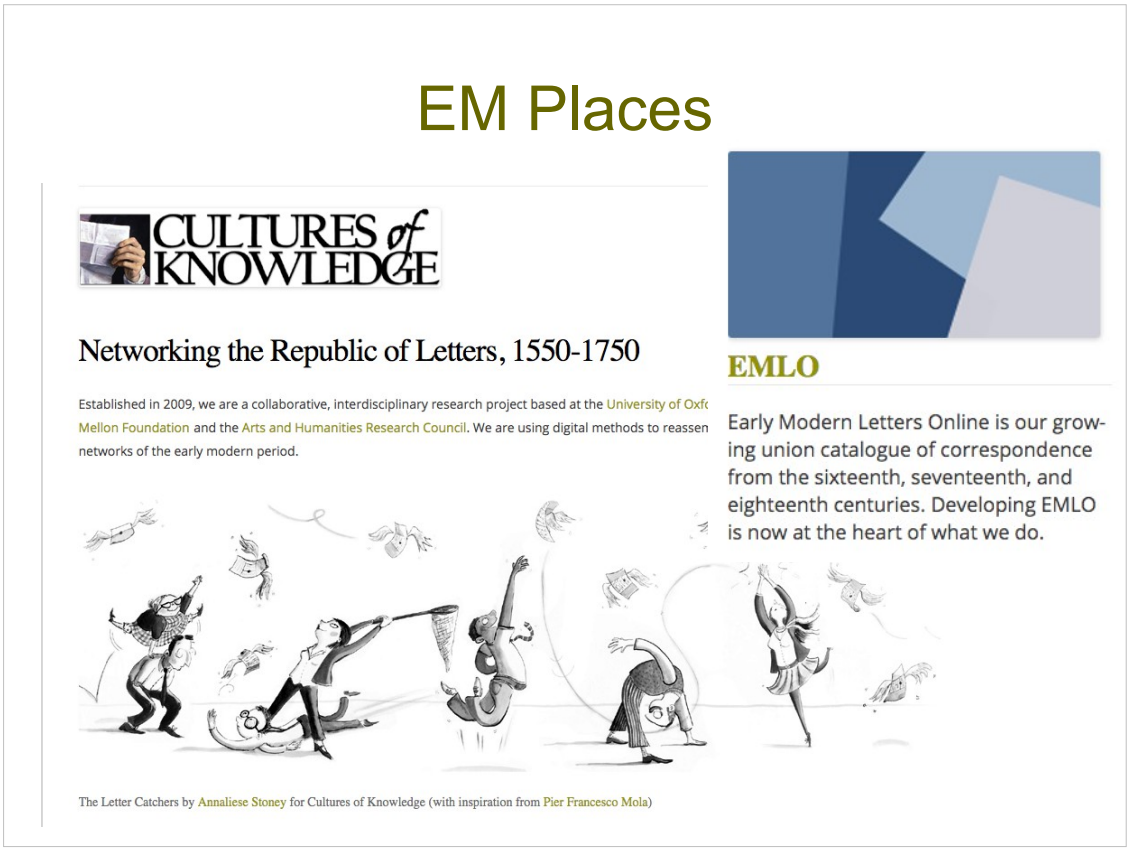
Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

xmlsummerschool.com

EM Places

Early Modern historical place data

EM Places



The Cultures of Knowledge project has created the Early Modern Letters Online (EMLO) catalogue, which contains information about selected correspondence by notable people in the Early Modern period. This period saw a flourishing of science and understanding that underpins modern society. The data collected includes information about the correspondents (people), correspondence dates, and places of origin and receipt of the letters.

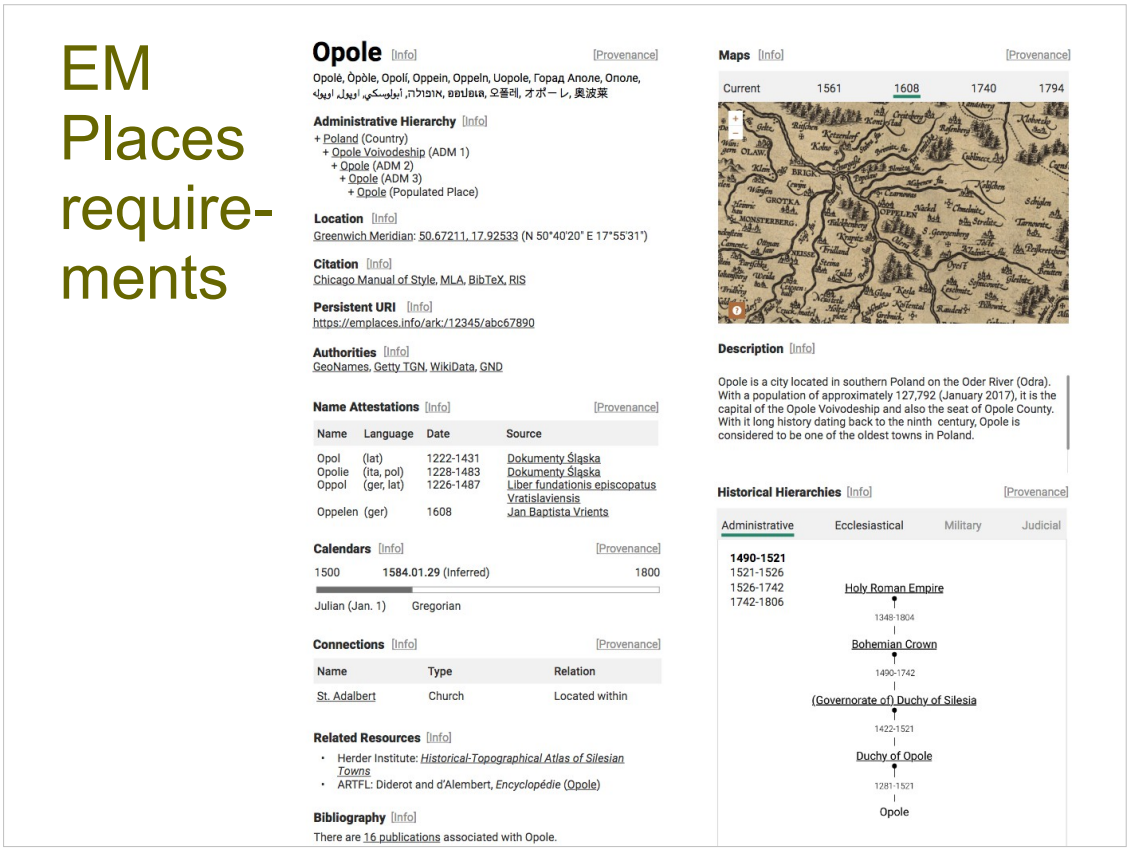
Currently, the place information is somewhat isolated from other similar resources. A fair number, but not all, of the places mentioned have GeoNames identifiers included in their descriptions, but that is as far as cross-referencing goes in EMLO.

Not all places mentioned in the letters are recorded in GeoNames, for a variety of reasons: GeoNames records current places, but places associated with the letters may no longer exist. Also, GeoNames tends to record larger entities such as cities and countries, but some letters refer to more local or obscure places.

Yet this place information may be important to allow for interpretation of of other related content; e.g. to determine the actual day in history that a letter was written may require knowledge of the calendar in use in the place where it was written, which may in turn depend on administrative relationships that were in place at the time.

///

EM Places require-ments



Opole [Info] [Provenance]

Opolé, Òpòle, Opolí, Oppein, Oppeln, Uopole, Горад Аполе, Ополе, אופולה, אופולה, أوپولە, أوپولە, 오폴레, オポーレ, 奥波莱

Administrative Hierarchy [Info]

+ Poland (Country)
  + Opole Voivodeship (ADM 1)
    + Opole (ADM 2)
      + Opole (ADM 3)
        + Opole (Populated Place)

Location [Info]

Greenwich Meridian: 50.67211, 17.92533 (N 50°40'20" E 17°55'31")

Citation [Info]

Chicago Manual of Style, MLA, BibTeX, RIS

Persistent URI [Info]

https://emplaces.info/ark:/12345/abc67890

Authorities [Info]

GeoNames, Getty TGN, WikiData, GND

Name Attestations [Info] [Provenance]

Name	Language	Date	Source
Opol	(lat)	1222-1431	Dokumenty Śląska
Opolie	(ita, pol)	1228-1483	Dokumenty Śląska
Oppol	(ger, lat)	1226-1487	Liber fundationis episcopatus Vratislaviensis
Oppelen	(ger)	1608	Jan Baptista Vrients

Calendars [Info] [Provenance]

1500 **1584.01.29** (Inferred) 1800

Julian (Jan. 1) Gregorian

Connections [Info] [Provenance]

Name	Type	Relation
St. Adalbert	Church	Located within

Related Resources [Info]

- Herder Institute: *Historical-Topographical Atlas of Silesian Towns*
- ARTFL: Diderot and d'Alembert, *Encyclopédie* (Opole)

Bibliography [Info]

There are 16 publications associated with Opole.

Maps [Info] [Provenance]

Current 1561 1608 1740 1794

Description [Info]

Opole is a city located in southern Poland on the Oder River (Odra). With a population of approximately 127,792 (January 2017), it is the capital of the Opole Voivodeship and also the seat of Opole County. With it long history dating back to the ninth century, Opole is considered to be one of the oldest towns in Poland.

Historical Hierarchies [Info] [Provenance]

Administrative Ecclesiastical Military Judicial

1490-1521
1521-1526
1526-1742
1742-1806

Holy Roman Empire
1348-1804
Bohemian Crown
1490-1742
(Governorate of) Duchy of Silesia
1422-1521
Duchy of Opole
1281-1521
Opole

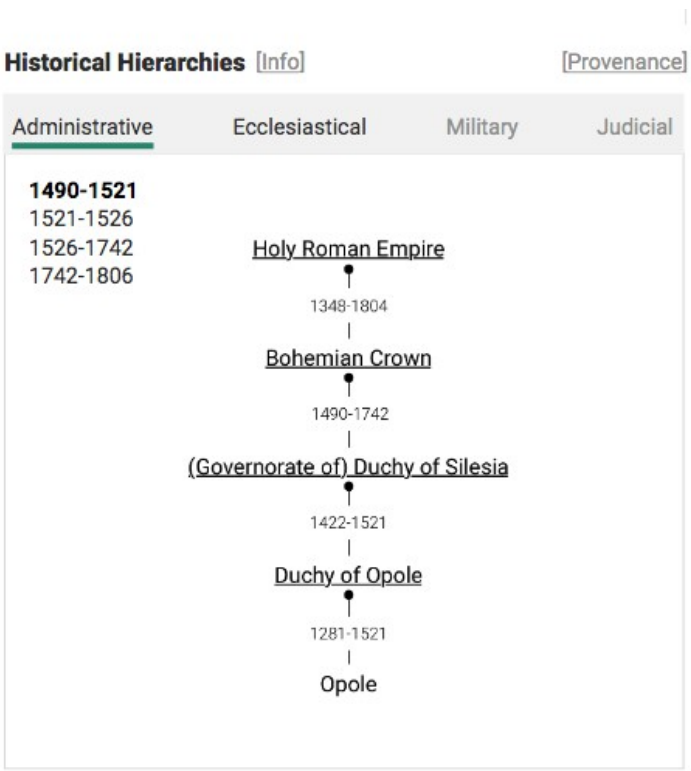
The aim of the EM Places project is initially to create a free-standing web site and service that can be used to collect information about historical places associated with Early Modern correspondence. This may subsequently be expanded to include other places referenced by other projects with similar needs.

Our aim has been to create a resource from which we can generate something like the display mock-up shown here, with temporally-qualified information about names used, calendars used, etc.

But we don't want this to become yet another isolated resource, so it is also important to be able to incorporate and reference existing authority data, where it exists, into EM Places. Our data model has been developed to maintain a separation between data that are obtained from different sources.

///

EM Places require-ments



The historians involved in this project felt that existing gazetteers did not fully meet their needs for a number of reasons:

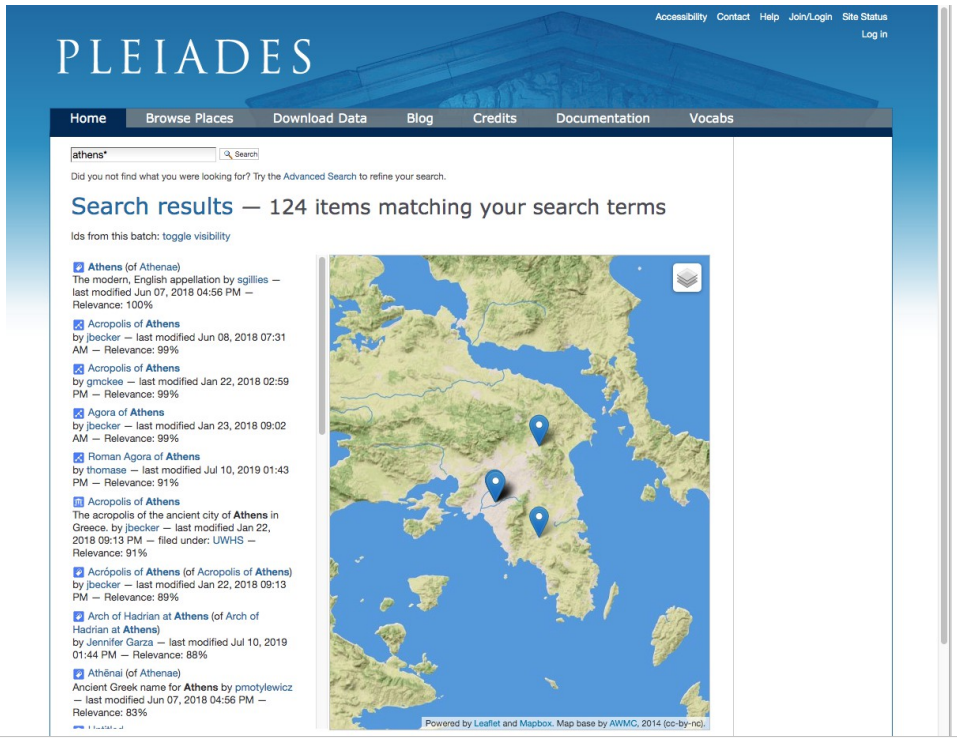
---
1. a need to record information about places that no longer exist.

---
2. a need to capture time-specific information about a place; e.g. different names and different calendars used in different periods.

---
3. a desire to be able to capture multiple time-varying hierarchical relationships between places; e.g. administrative, ecclesiastical, military, etc.

///

Related work: Pleiades



There has already been a fair amount of work on recording information about historical places, from which we have tried to avoid gratuitous deviation.

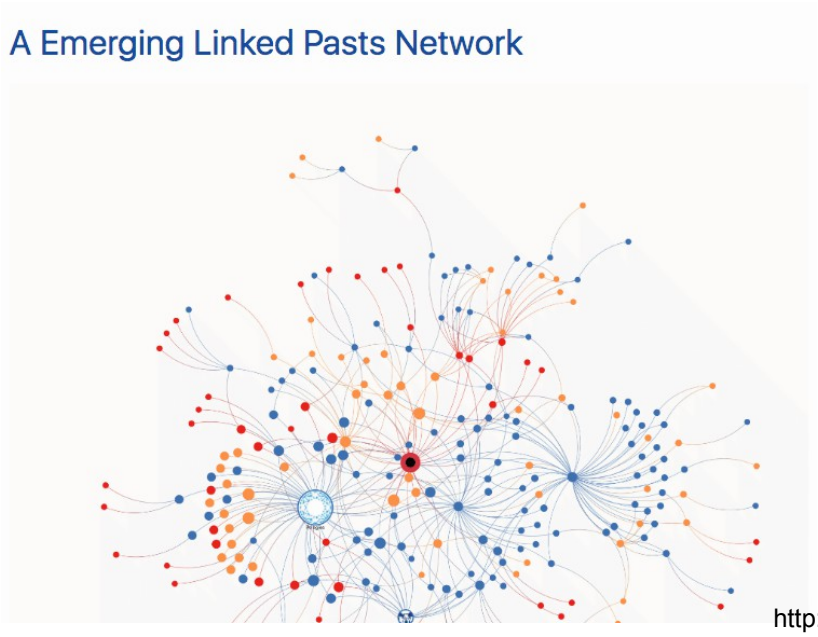
One important related work is the Pleiades project, which has created an online service to access information about historical places.

A starting point for our place modelling work was the [Pleiades data model(https://pleiades.stoa.org/help/pleiades-data-model), which explicitly separates the notions of place, place name, place location and temporal applicability. This approach has been adapted and extended to handle other kinds of temporally qualified information, especially hierarchical relationships between places.

When we started on EM Places, Pleiades did not record time-varying hierarchies. Since our early modelling work, the Pleiades project has started to introduce the notion of a “connection resource”, which closely parallels some of our modeling ideas for temporally qualified relationships, but was not available to us in time for us to use in our work.

(But it is encouraging that the similarity of approach does mean that more faithful data exchange should be possible in due course.)
///

Related work:
Linked Pasts/Linked Places



http://linkedpasts.org/

Another related work is the Linked Pasts Network, which is developing a “Linked Places” data model for interchange of historical place data. The Linked Places model itself draws ideas from the Pleiades project.

This is a quote from the Linked Pasts home page:
“Linked Pasts Network refers to a growing collection of historical datasets published as Linked Open Data and the community of researchers across the globe whose projects are producing and using them.”
-- http://linkedpasts.org/

(Linked Places)[https://github.com/LinkedPasts/linked-places], is a broadly based project that aims to provide a common framework for interchange of historical place data. As such, it covers some of the same requirements as our Cultures of Knowledge work, but aims for broader use while not necessarily being a complete solution for any single application.

We have been in communication with the Linked Places developers to ensure that our modelling remains broadly compatible and capable of supporting good fidelity conversion. As such, our work forms part of a wider effort to publish independently-developed historical datasets as linked data, allowing for information to be shared and cross-referenced between the datasets, thus building an aggregate that is more generally useful than the sum of its parts.
///

What is a place?

“constructed by human experience”

Pleiades content is organized into three types of "information resource": *places*, *locations*, and *names*.

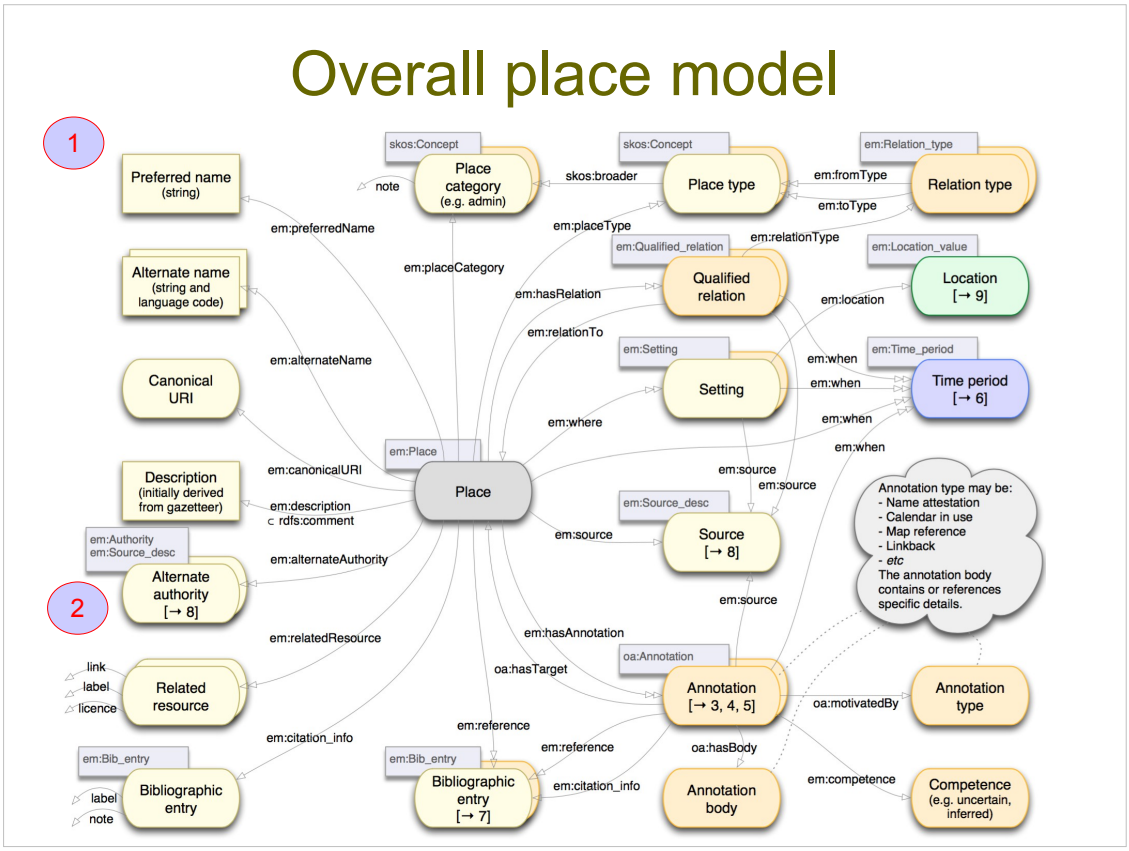
Places

In *Pleiades*, *places* are conceptual entities. They are objects of thought, speech, or writing, not tangible, mappable points on the earth’s surface. Even though they are usually tied closely to some aspect or feature of the physical world — a sea, a bay, a river, a mountain range, a pass, a road, a settlement, or an ethnic region — They have no spatial or temporal attributes of their own. Their primary quality is that, in the words of Yi-fu Tuan [1], they are "constructed by human experience."

-- https://pleiades.stoa.org/help/data-structure

There are many ways in which the idea of “Place” may be conceived. We are particularly interested in capturing an idea of place from a human perspective.

Thus, the idea of a place that we use is not location-specific nor period-specific. We follow the notion adopted by the [Pleiades project](https://pleiades.stoa.org/) that
---
‘[places] are conceptual entities ... They have no spatial or temporal attributes of their own. Their primary quality is that they are "constructed by human experience.”.
///



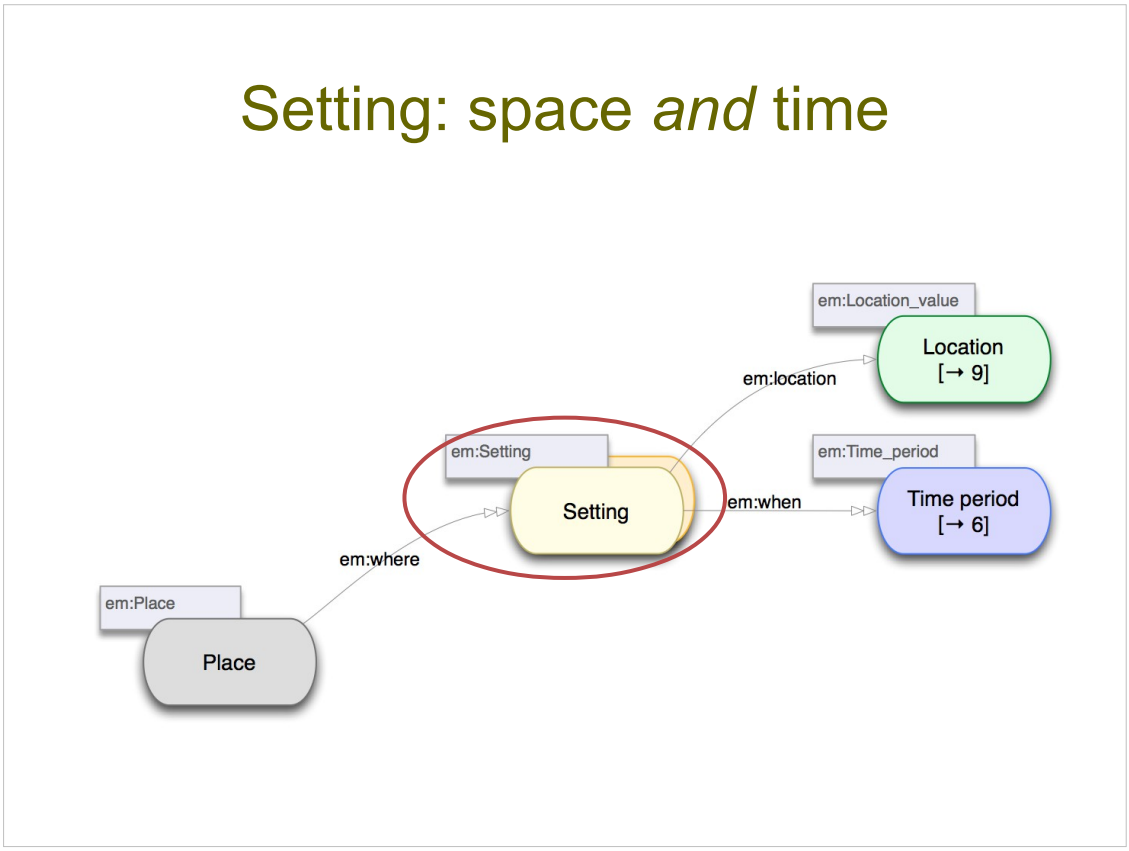
So how does EM Places describe a place?

In the next few slides, I shall describe some specific elements of the EM Places data model, and relate them back to the requirements discussed previously.

In viewing this work, do bear in mind that when modeling data there is usually not a single right way to do it. (Even if there are plenty of wrong ways!) So some of the choices presented here are somewhat arbitrary, and shouldn't be taken as a specific prescription. What I hope to expose is how the previously identified requirements can impact the modeling choices made.

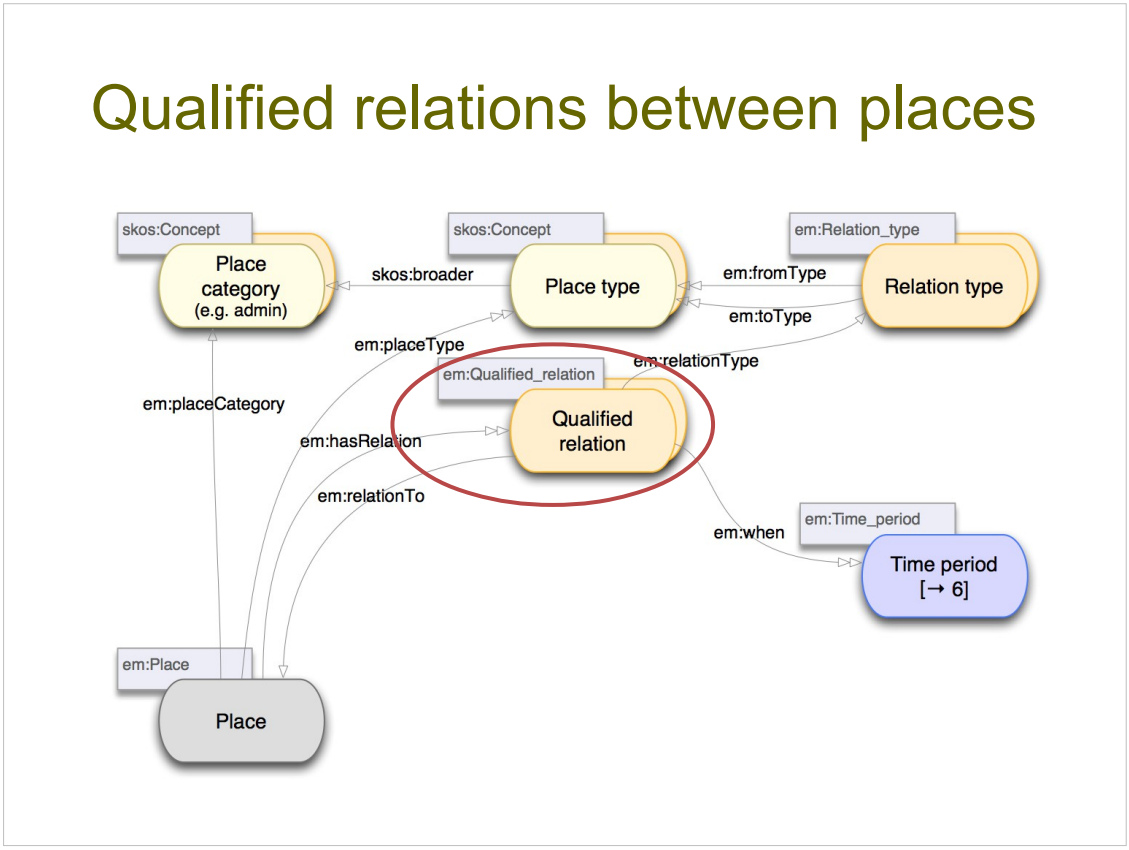
This slide shows the overall place model, which is mostly fairly conventional, but there are some points I'll highlight:
---
1. Much of the left hand side is information extracted from an existing authoritative source (e.g., GeoNames), and as such is designed to be refreshable when an original source is updated. More details about this come later.
---
2. Alternate authority data (i.e. identifiers derived from different authorities) allow cross-referencing to existing gazetteers. In our work, given a GeoNames identifier for a place, we use Wikidata to locate some of these alternative identifiers.
///

https://github.com/culturesofknowledge/emplaces/tree/master/models
https://github.com/culturesofknowledge/emplaces/blob/master/models/PDFs/20180802-EMPlaces-data-model-multisource.pdf



3. In our work, the notion of a “Setting” is used to represent an extent in place and time (this terminology is adopted from earlier work of the Linked Pasts project).

A place location may change over time, so multiple settings may be associated with a place (where place is in the sense of being “constructed by human experience”).
///



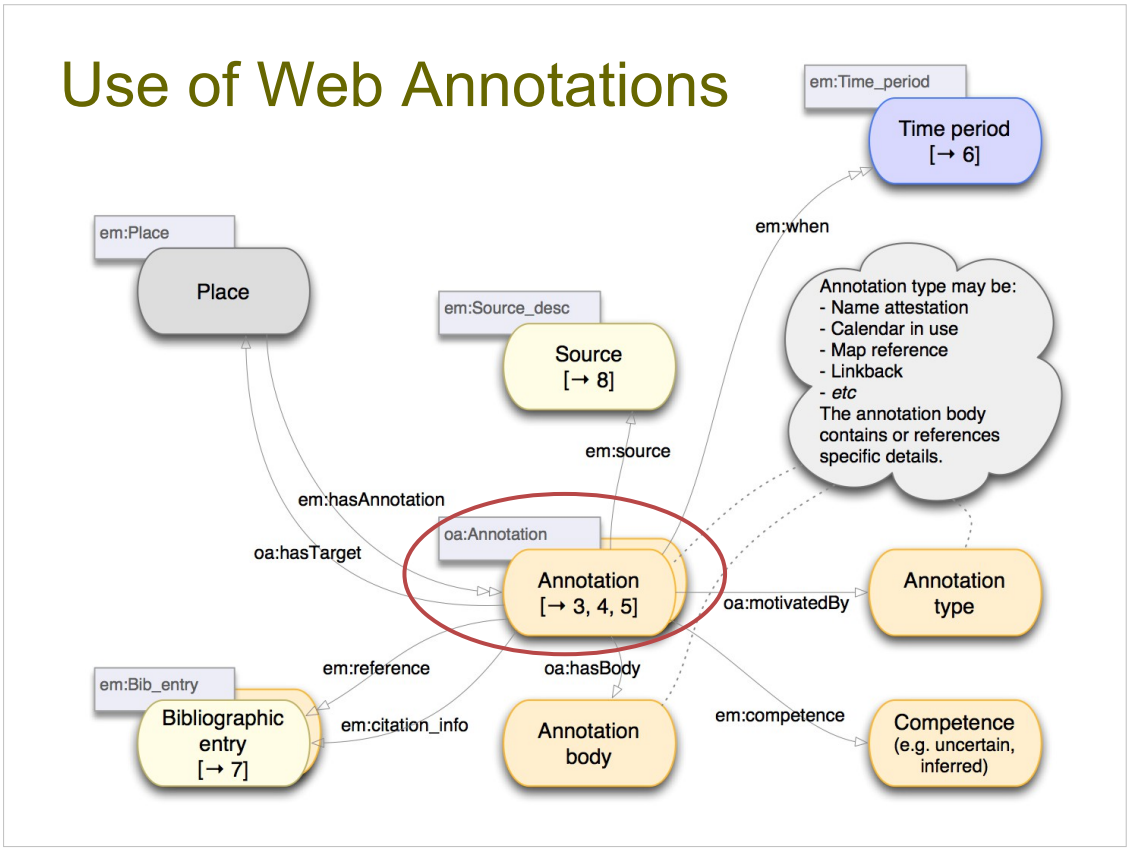
4. We represent a Qualified Relation as a node in the graph, rather than as just a direct relation. This is used to attach additional context for the relations between places, such as a time period, allowing us to capture multiple hierarchies of temporally qualified relations.

The pair of “Place type” and “Place category” values for relation types means that we can isolate different hierarchies (based on “Place category”). E.g, some relations apply to places as administrative entities, while others may apply to ecclesiastical entities (church parishes, etc.). A place may exist in multiple categories.

For example, the City of Opole in Poland is a populated place and an administrative region that is part of Opole Voivodeship, itself an administrative region that is part of the country of Poland. By comparison the parish of the Holy Cross in Opole is in the Deanery of Opole, in turn part of the Diocese of Opole: these ecclesiastical part-of relations would be distinct from administrative part-of relations, and are not necessarily in alignment.

The additional node representing a relation is an example of what I call “selective reification” of a relation, which is something I'll come back to later..

///



5. Other information, such as place names used, calendars used, map references, etc., is captured using Web Annotations, used here as a kind of qualified and contextualized relation between a place and the additional information.

This use of Web Annotations means that we are using an existing standard vocabulary, albeit in a slightly extended form (e.g. with an associated time period). This provides an extension point (via the annotation type value) that can be used to incorporate additional information not defined by the initial data model, yet with a reasonable expectation that at least some of the information will be accessible to applications that know about web annotations.

The provision of a time period value attached to the Web Annotation node is used for recording a temporal constraint on the scope of the annotation.

///

Reflection on recording context

"People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated."

(Dan Brickley and Libby Miller)

- https://book.validatingrdf.com/bookHtml005.html

See also:

https://github.com/gklyne/notes/blob/master/20180307-qualified-property-links.md

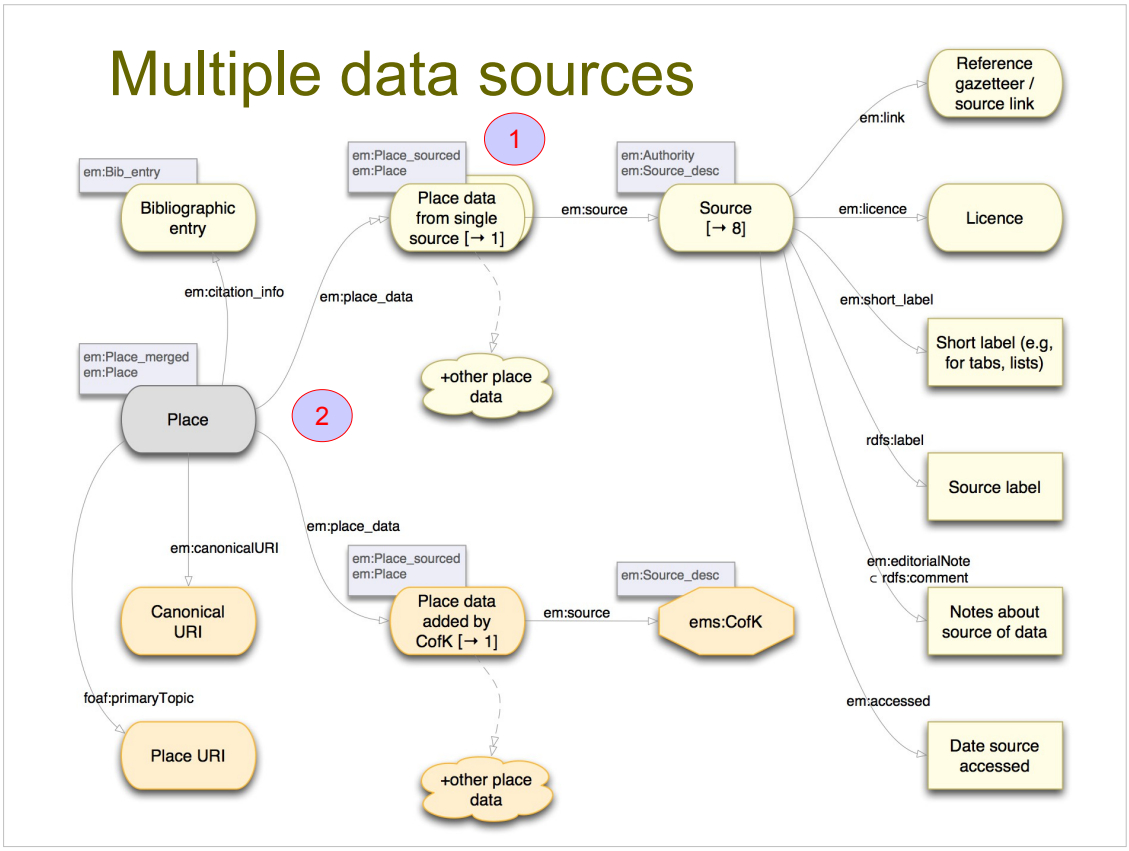
Recall the quote from Dan Brickley and Libby Miller. Note how our place data model has been formulated to allow additional context information to be explicitly recorded.

In this respect, the introduction of Web Annotation and Qualified Relation nodes reflects a limitation of RDF compared with some other representations (such as property graphs). RDF does not allow additional information to be directly associated with a relation between nodes, so, to incorporate such information, we expand the application data model.

I have prepared [a small survey of modeling techniques] (https://github.com/gklyne/notes/blob/master/20180307-qualified-property-links.md for what I call "qualified properties", which can be found at the link shown here. Further, there is work in progress in the RDF community to address the representation of qualified relations, some of which is referenced in the survey document.

In this matter, we have adopted an approach that I refer to as selective reification. The approach is similar to that taken by the W3C provenance ontology (PROV-O), and also one that is being developed as a recommendation for representing role-in-activity information in CIDOC CRM. (Where CIDOC CRM is an ISO standard data model for representing museum data and more.)

There are also some theoretical considerations indicating that reified relations may allow simpler and more predictable performance for certain kinds of inference based on the relations (such as being specifiable using a particular restricted form of the web ontology language, OWL-QL, which is designed to be implemented using conventional, fast query mechanisms). For example, these considerations were an immediate motivator for introducing "Qualified associations" and related elements, in the W3C provenance ontology, these being another example of selective reification.
///



Referencing multiple authority data sources has been mentioned previously. Using the original place model as presented, this is not achieved in a way that easily allows data from a single source to be refreshed.

The problem here is that when new data is pulled in from an authoritative source, the data previously obtained from that source should also be removed. But this is only possible if we have some way of identifying which data came from that source.

The solution we have adopted is to introduce an additional layer of place information, so we have:

---

1. A “Place_sourced” node representing information about a place from a single source, which uses the place data model described previously. Additional information about the source, when it was retrieved, and licensing information about the source data is associated with the “Source” node. Data that doesn’t come from an existing authority, but which is generated by EM Places editorial activity, is associated with a special “local-source” identifier (e.g., ems:CofK), but otherwise the modeling pattern is essentially the same as for any other authority source.

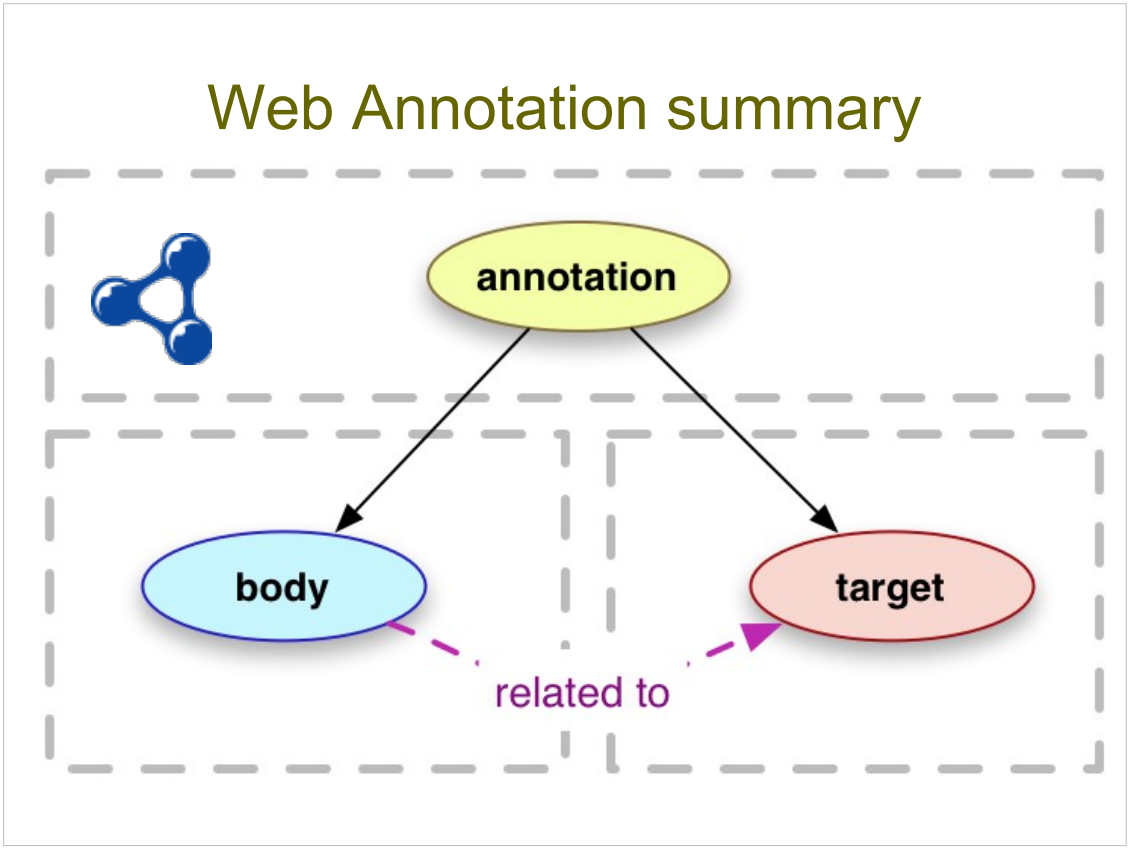
---

2. We also use a “Place_merged” node representing all of the available information about a place, which in turn references one or more single-sourced place descriptions.

Using this structure, all the information obtained from a particular source is associated with a “Place_sourced” node that references that source, and as such can be located and removed when data is refreshed.

The trade-off here is that applications needing to see the overall place data must take steps to traverse and aggregate data from each of the single sources.

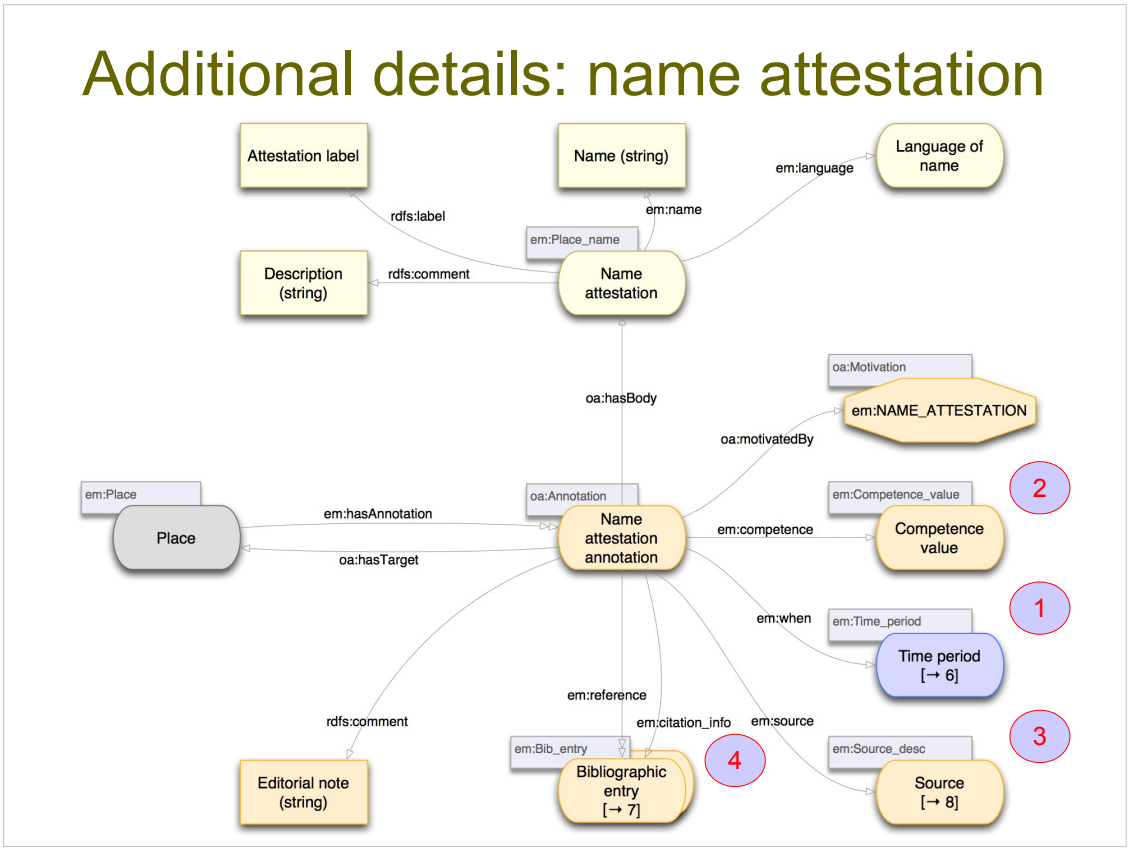
///



Next we'll look at additional details about a place that are included using Web Annotations.

The basic structure of a Web Annotation (at least, as used in EM Places) is an “annotation” node that references a “target” (what the annotation is about), and a “body”, which is what the annotation is claiming about the indicated target.

An additional “motivation” value associated with the annotation node is used to indicate the kind of relation between the target and body it is that the annotation is being used to represent.

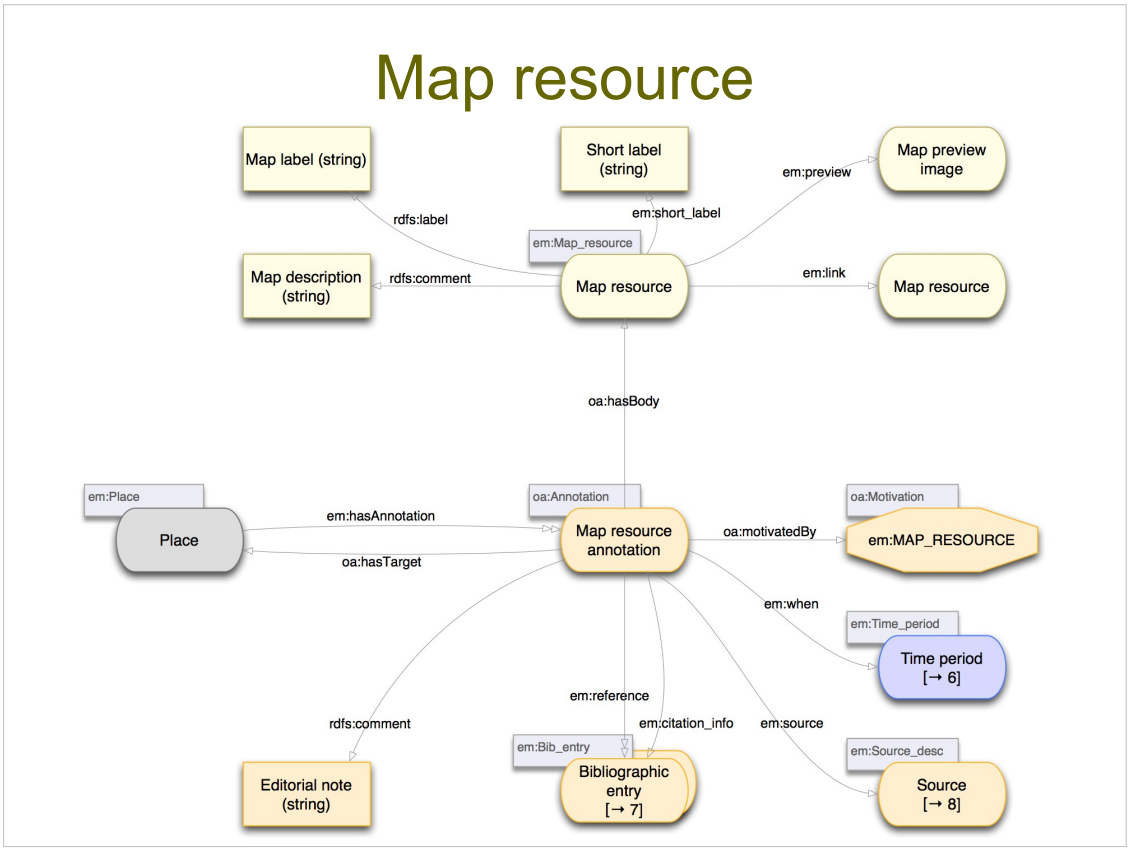


A “Name attestation” is our first example of place detail using Web Annotations. This is an evidenced description of an alternative name that has been used for a place.

This diagram shows how the attested name and contextual metadata are associated with the place using a Web Annotation.

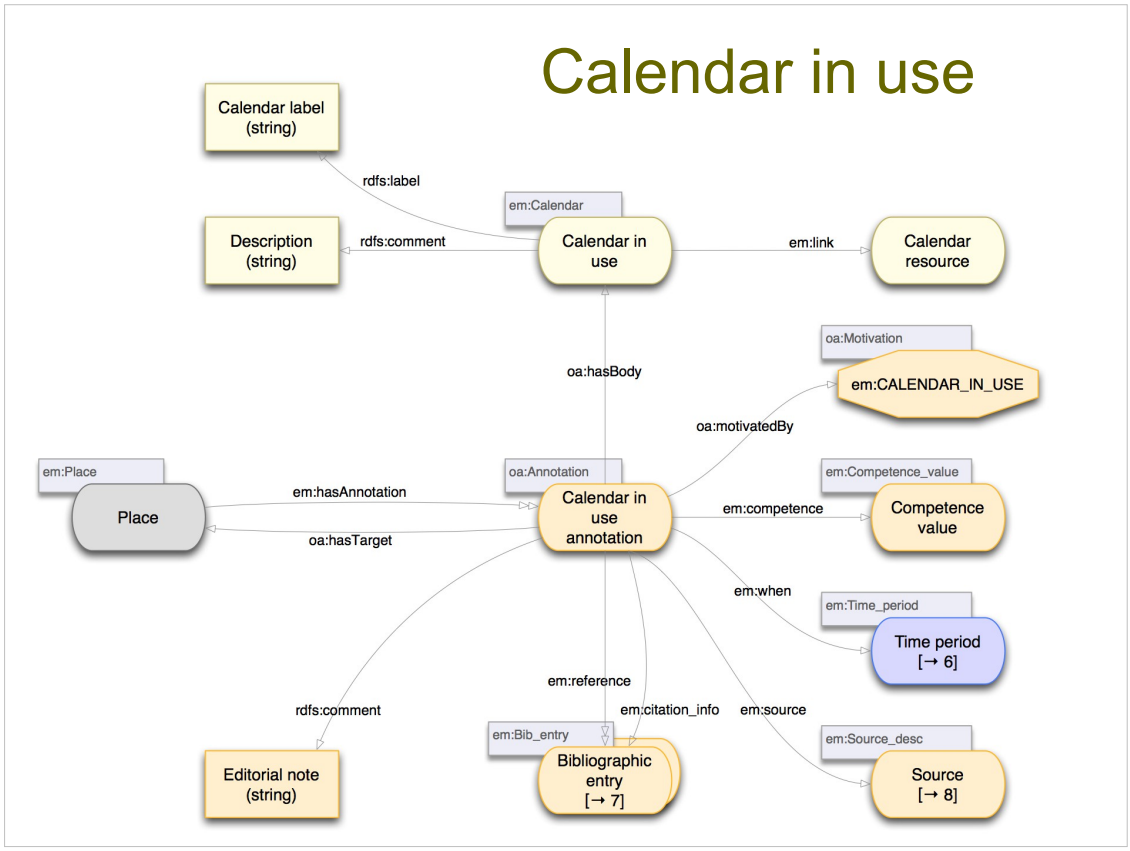
The name itself is described by a “Name attestation” node, which is referenced as the annotation body. The information specified here is mainly determined by requirements of the intended user interface for EM Places.

Elements used here that are additional to the the Web Annotation specification include:
---
1. temporal constraint on the applicability (i.e., when, historically, the name was used)
---
2. an indication of reliability or accuracy (referred to jointly as “competence”) of the information provided
---
3. an indication of the source of the name attestation information
---
4. bibliographic and citation information relating to the name attestation
---
All of these additions are intended to help scholars to assess the applicability of a given name attestation, by providing additional context to aid interpretation of the data.
///



The map resource annotation allows historical maps corresponding to different time periods to be associated with a place.

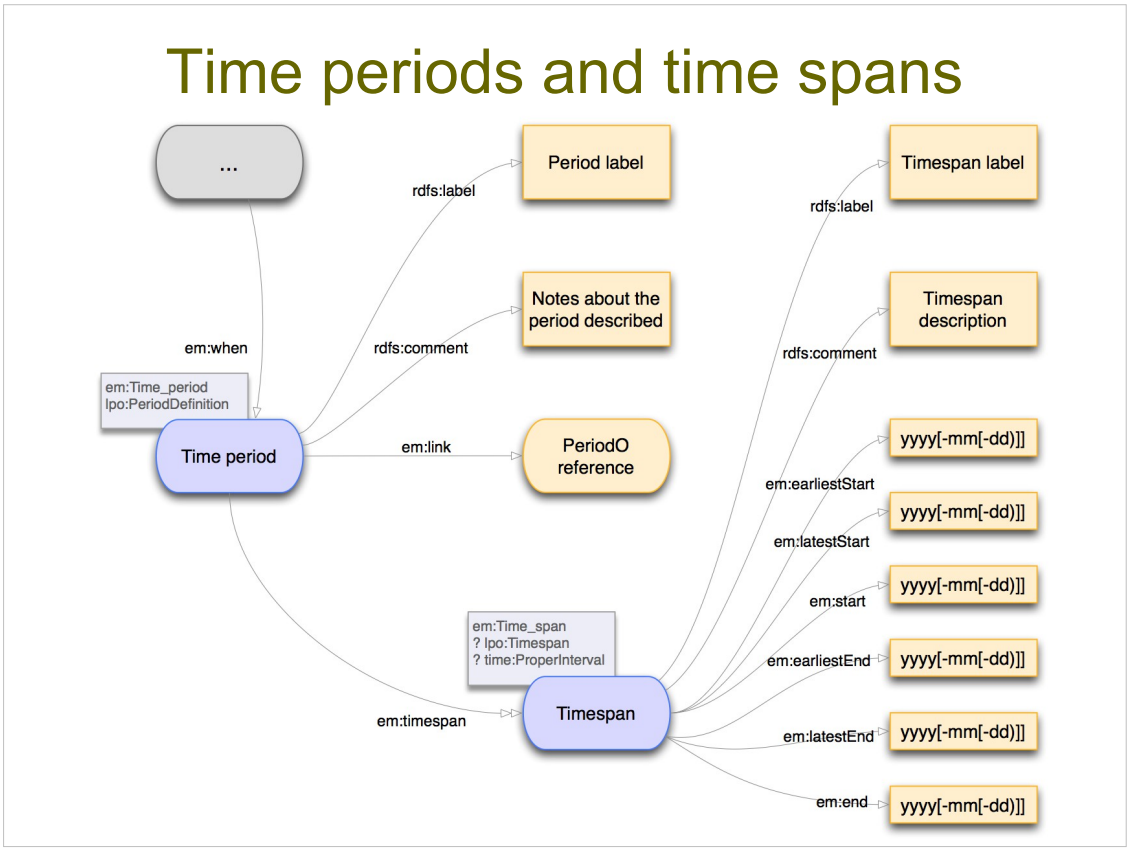
This follows a similar pattern to the name attestation annotation, but with a different value for the motivation, and a different structure for the annotation body.
///



Similarly, a calendar-in-use annotation allows calendars used locally during different time periods to be associated with a place.

Again, the pattern is similar, but the details vary.

///



I've made several references to temporal constraints on information in the EM Places data model. These are all captured through using a “time period” description.

A time period may be associated with a place, a place setting, a qualified relation between two places, or an annotation providing additional qualified information about a place.

Some key points to note here are:
---
1. there is a distinction between a “time period” and a “time span”. A “time period” refers to some identified historical period without necessarily indicating when it actually occurred; e.g. “Early Modern period”. This follows the notion of time periods catalogued by the PeriodO project authority data, where each identified period is based on a cited document that refers to that period by (some) name.
---
2. a time period MAY have an associated “time span”, which provides an indication of when in history a period actually occurred, and whose exact extent may be only approximately known. We have settled on 6 terms to allow capture of uncertainties in the starting and ending dates of a time span.

This time span structure loosely follows work in the “Linked pasts” project, which was attempting to build upon the OWL Time ontology, but which we found was difficult to use for our purposes without introducing semantic inconsistencies. Semantically consistent mapping from the terms used here to OWL Time terms is possible, but the resulting data structure becomes somewhat more complex.
///

EM Places model summary points

Context and provenance

- Expanded data model

Ontology sharing/reuse: parallel efforts

- Tyranny of timescales

Ontology sharing/reuse: local needs

- Specific application needs vs concept sharing

In summary, the following points can be made about our EM Places data model.

Context and provenance are captured by enlarging the application data model compared with what would be required to capture just the raw information.

The existence of parallel efforts in our domain of interest means that the ideal of vocabulary sharing conflicts with project timescales, particularly for some humanities projects where the technical data model design may be a part-time effort conducted over an extended period.

Local project needs are not necessarily reflected by common vocabularies for information sharing, so rather than trying to force-fit into an existing vocabulary we have tended to create new terms, but remaining cognisant of how these can be mapped to the common vocabularies. In some cases (e.g. for time spans), local terms have been used to keep the overall data model simpler.
///

Data model references

Place model description files:

- https://github.com /culturesofknowledge/emplaces/tree/master/models

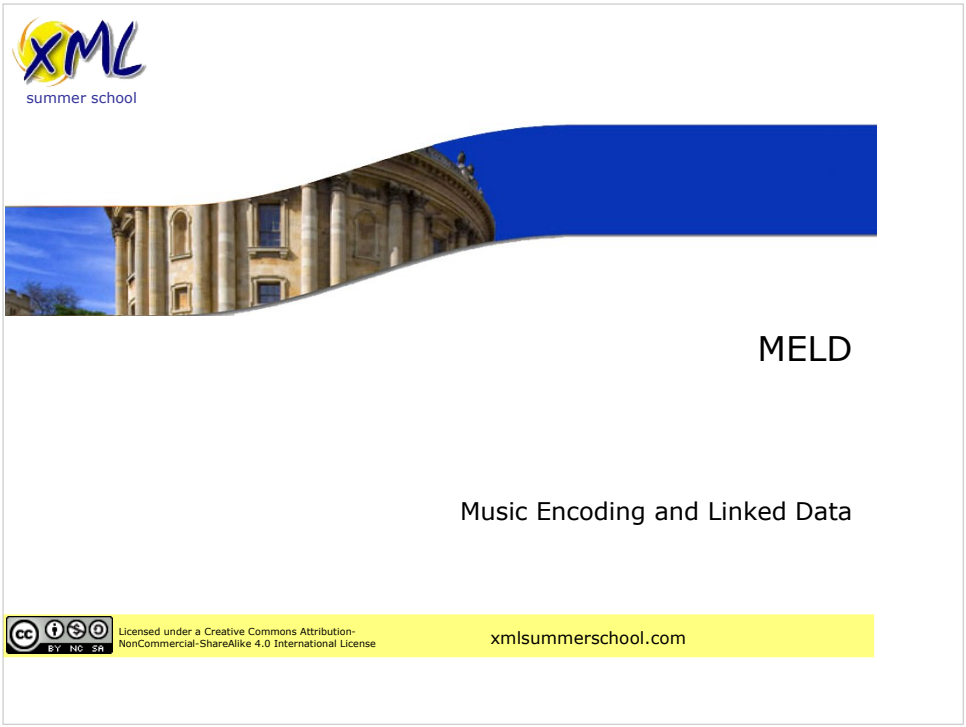
Place model diagrams:

- https://github.com /culturesofknowledge/emplaces/blob/master/models /PDFs/20180802-EMPlaces-data-model-multisource.pdf

Place model sample RDF data:

- https://github.com /culturesofknowledge/emplaces/blob/master/models /20180802-opole-example-multisourced.ttl

(This slide is included for the course notes to provide references to more details about our design.)
---
https://github.com/culturesofknowledge/emplaces/tree/master/models
---
https://github.com/culturesofknowledge/emplaces/blob/master/models/PDFs/20180802-EMPlaces-data-model-multisource.pdf
---
https://github.com/culturesofknowledge/emplaces/blob/master/models/20180802-opole-example-multisourced.ttl
///

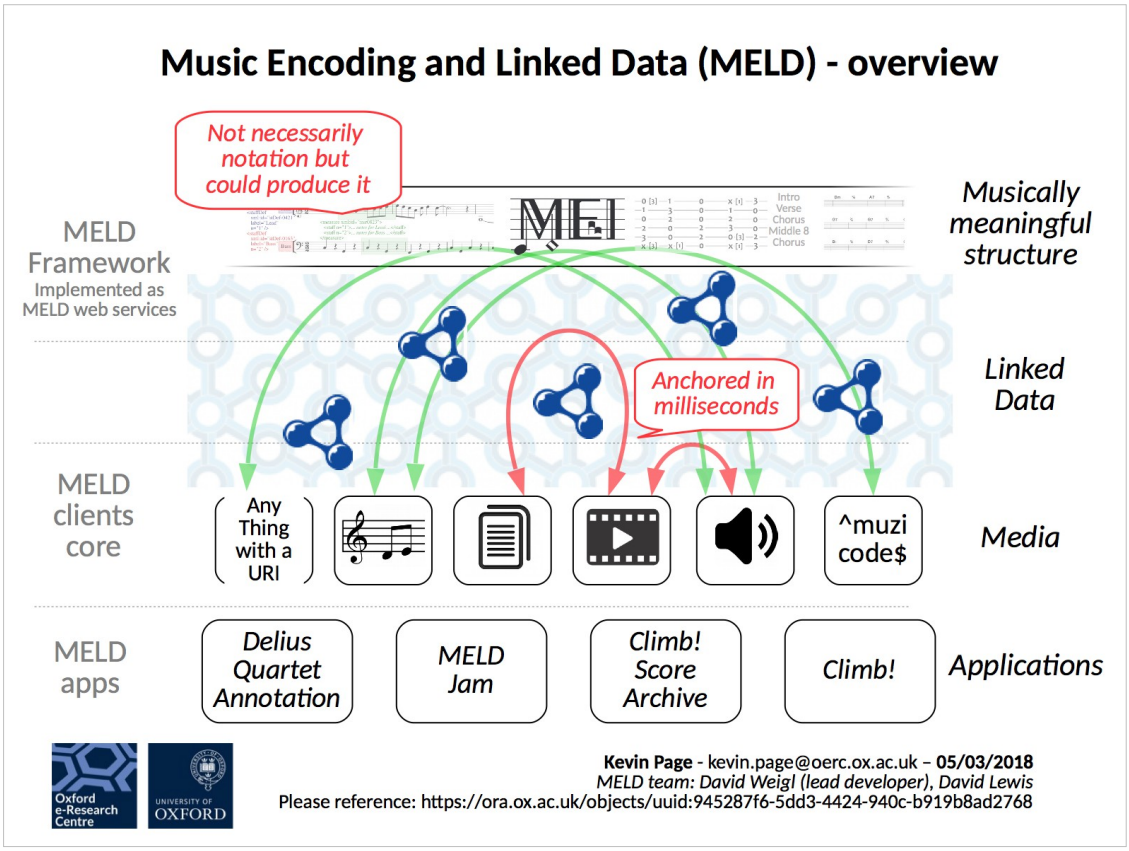


I now turn to the second project I shall describe: Music Encoding and Linked Data (MELD).

///

MELD

Music Encoding and Linked Data



Historically, a body of work dealing with computer analysis of music has focused on an audio stream of a performance as a primary record of the music, for which a common way of identifying particular musical elements is by time offset within the performance. This raises a number of problems:

---

1. Different performances will generally have different timing, so a reference is performance-dependent.

---

2. References are separated from any notion of musical structure.

---

3. Research to determine musical structure from audio performances is still early-stage, and is far from complete and reliable across all potential applications.

---

4. Similar problems arise when an image of (say) a musical score is used.

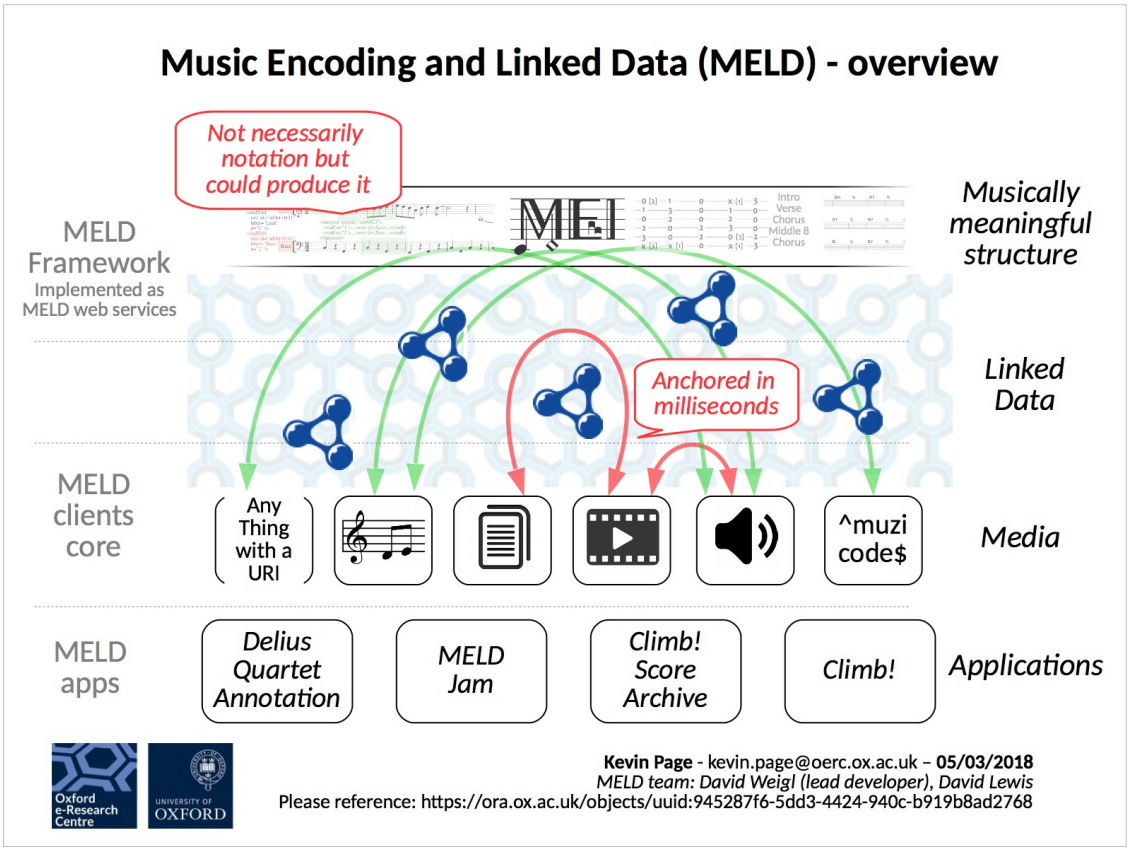
In many cases, there exists a more structured digital description of the music (e.g. Midi) that can be used as a performance-independent description of the musical content. One such format, that is finding application in music-related research, is the Music Encoding Initiative (MEI), which is somewhat like TEI but for musical rather than textual source material.

MEI is additionally supported by the Verovio renderer, which can present suitable MEI data as a musical score that is suitable for a sight-reading musician to follow.

///

-- https://github.com/oerc-music/meld

-- https://github.com/oerc-music/meld/blob/master/meld-overview.png



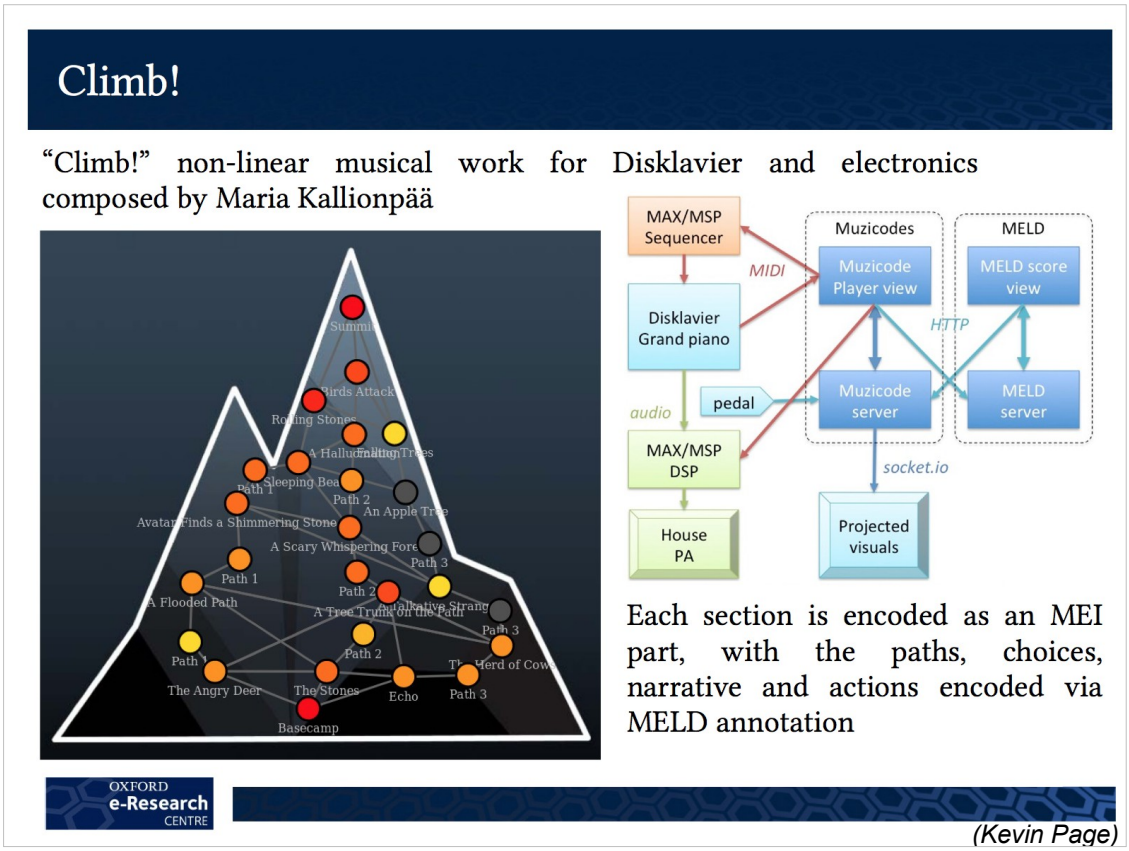
MELD uses linked data to connect musically meaningful structure with with other related resources (e.g., musical score images, documents, performance audio, videos - indeed, any data that can be located on the web using a URI.)

MELD is not a single application, but a framework that has been used for a diverse range of applications, a few of which are covered in the following slides.:

(To follow: 1. Climb!, 2. Climb archive, 3. SOFA remixer)

It is not my purpose here to describe the various applications in detail, but to highlight that MELD can support a range of very diverse applications, including musical composition, musical performance and musical review activities, based on a small number of common underlying elements.

///



(Kevin Page)

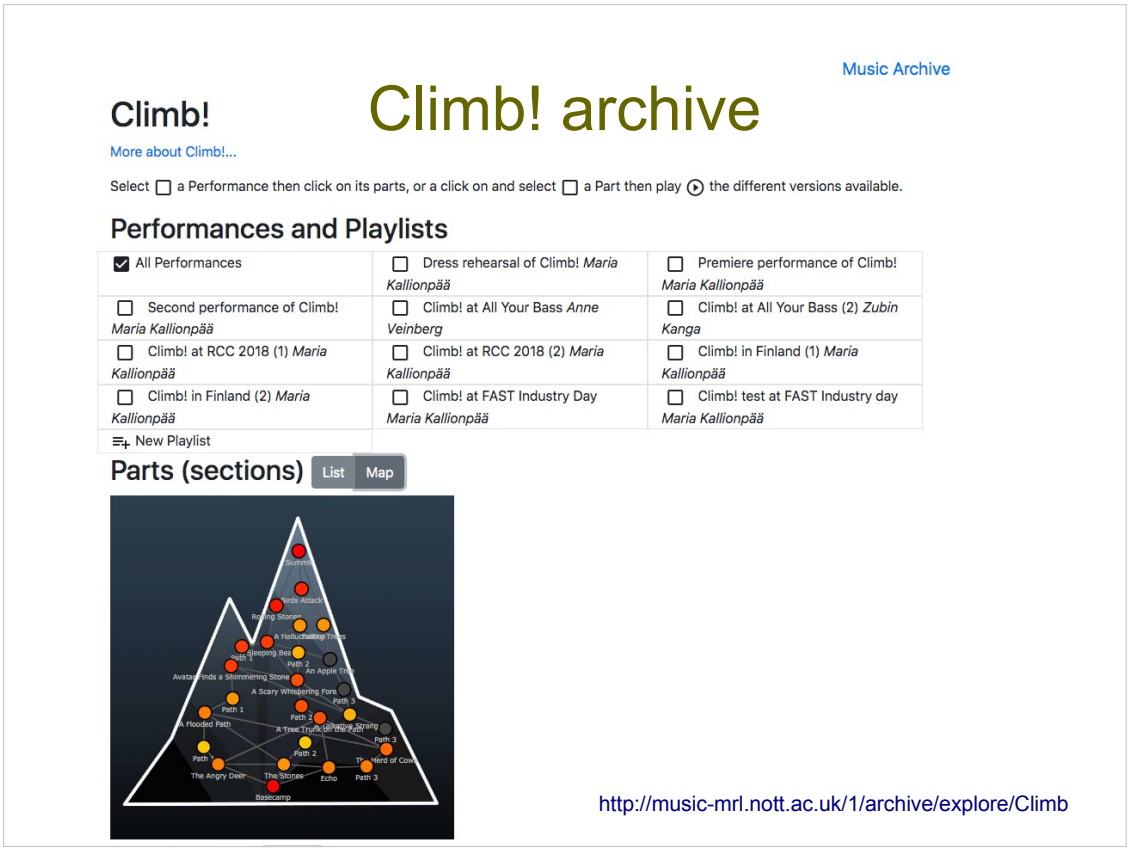
The first of our MELD example applications is "Climb!".

"Climb!" is a non-linear musical work for Disklavier and electronics, composed by Maria Kallionpää.

The overall piece represents a passage up a mountain, with various obstacles to be traversed. Each performance may follow a different route, so, typically, no two performances follow the same static score.

Each passage in the music is encoded as an MEI file, with choices for the next passage to play determined by events triggered by the musician's actual performance. The score presented to the performer is updated in real time, reflecting the actual path up the mountain.

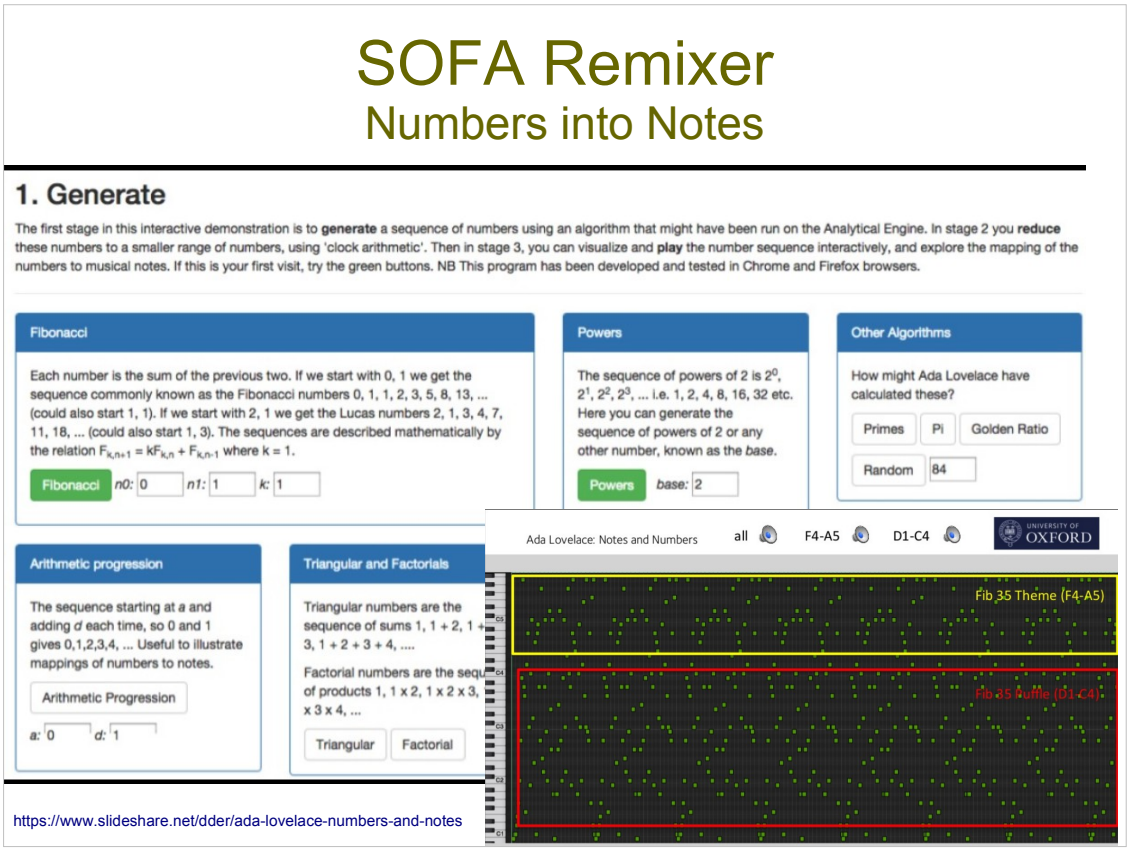
///



Our second example of MELD, an application related to Climb!, is the “Climb!” archive.

With no two Climb! performances being the same, retrospective study by musicologists or reviewers becomes more challenging.

Using MELD to connect archive user interface events to audio playback and score display allows navigation and comparison of different performances that follow different paths, but still contain some passages in common.
///



Anf finally, a very different use of MELD is in a guided music composition application, the SOFA remixer:

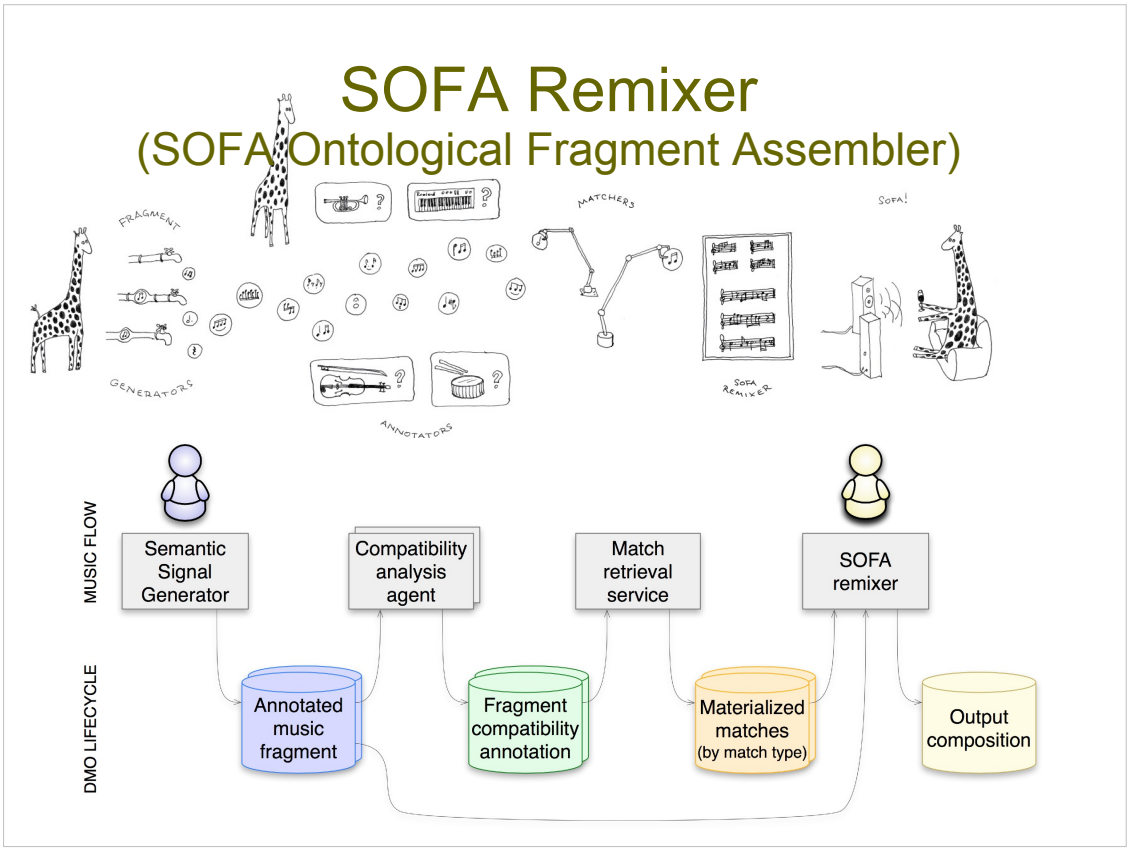
The SOFA remixer builds upon “Numbers into Notes”, an application written by Prof. David De Roure, which was inspired by this speculation nearly 200 years ago by Ada Lovelace:

“Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent.”

The “engine” here is a reference to Babbage’s Analytical Engine.

The [Numbers into Notes](https://www.eng.ox.ac.uk/people/david-de-roure/ @@down?) application uses a combination of numerical manipulation and user selected parameters to generate fragments of music algorithmically.

///



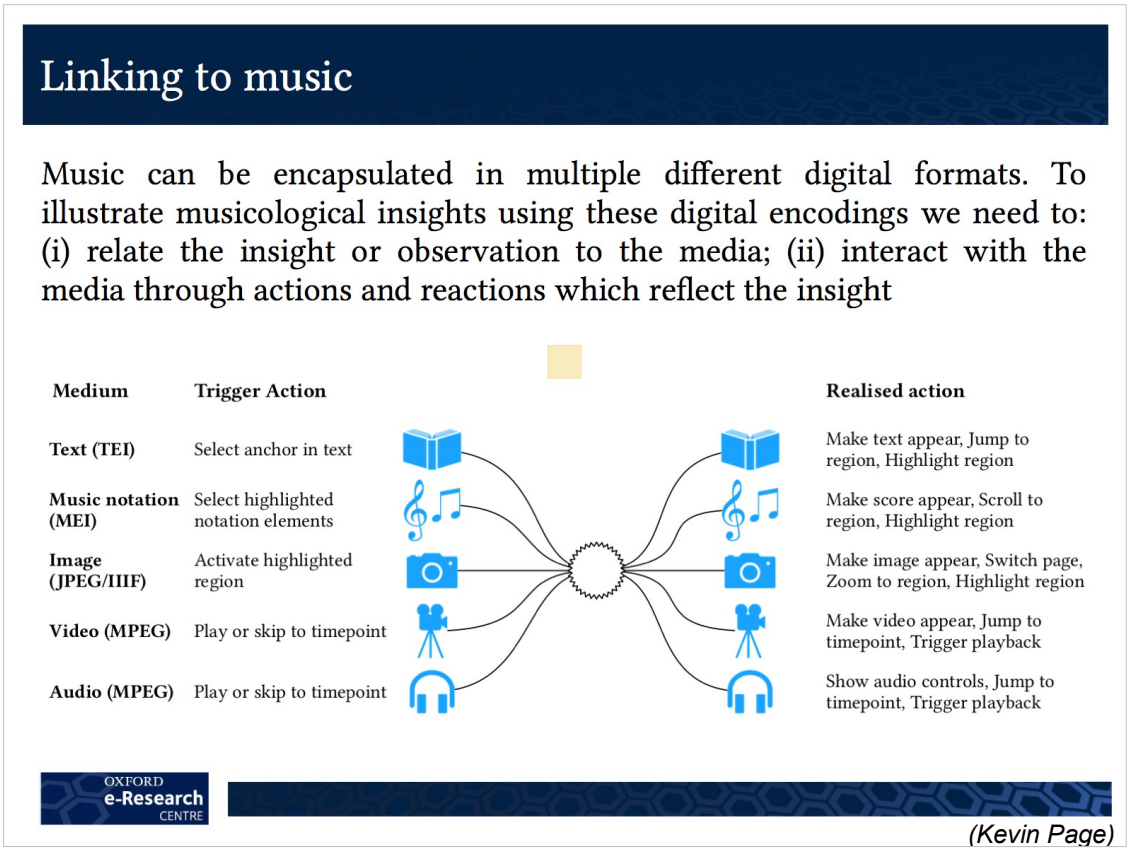
The SOFA Ontological Fragment Assembler (SOFA), reads a collection of generated fragments (such as those generated by Numbers into Notes) and provides an interactive interface for assembling these into a multi-part musical composition that meets certain constraints, such as musical key compatibility. Satisfaction of these constraints is determined using MELD annotations created by agents that analyse the music fragments.

Each fragment includes a playable audio sequence, and also an MEI description that can be used to render score elements for that fragment.

The elements of MELD allow the analysis agents to be independent and separate from the remixer application. Thus, new forms of fragment matching analysis can be added by writing a new agent, without having to change the rest of the system.

The MELD linked data model is used to provide the structure that the remixer needs to locate and incorporate the various available analyses.
///

SOFA Remixer: sample interface



(Kevin Page)

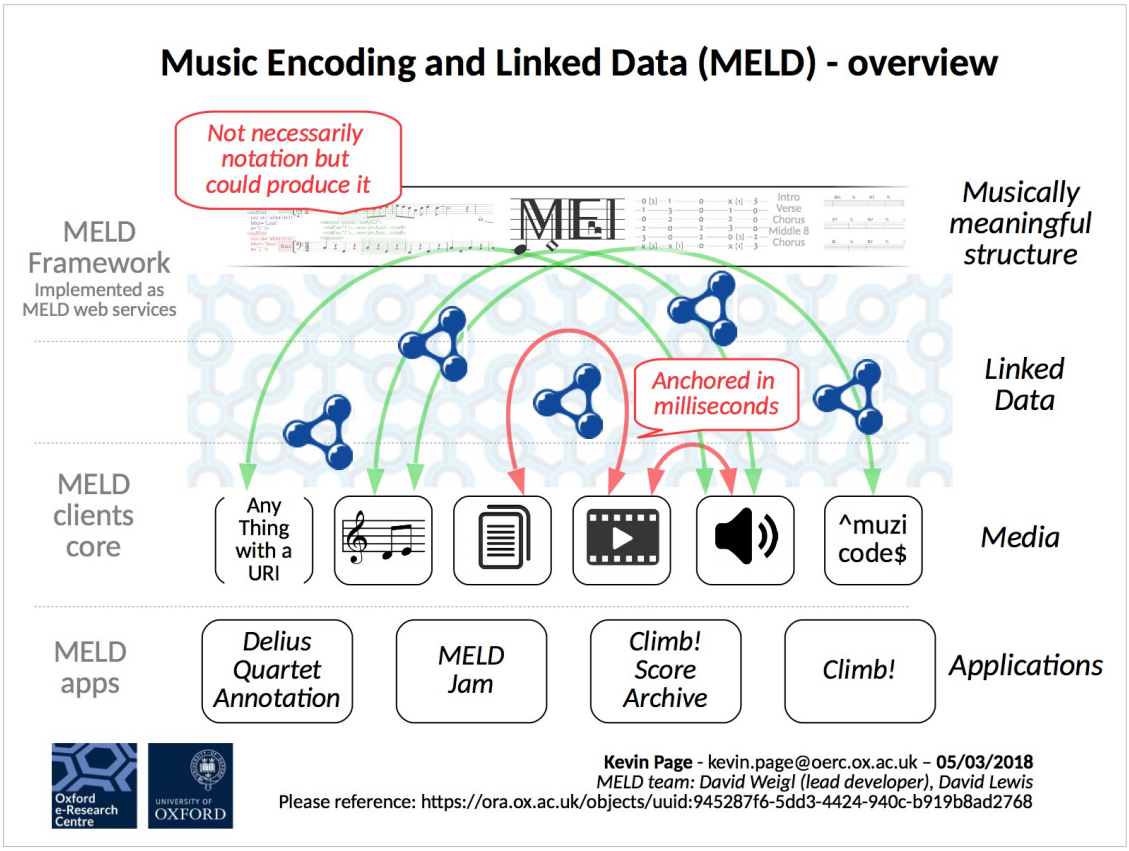
The common theme across these MELD applications is the use of linked data to capture musical insights through connecting musical elements and media, and to use these connections to enable a diverse set of interactions with the musical content.

In the case of Climb!, it is actions by the musician and detection of patterns in the music as it is played, that trigger updates to the musician's displayed score and also to audience displays about the progress of the performance.

In the case of the Climb! archive, it is user interactions with the archive web interface that trigger playback of audio from a past concert, alongside a presentation of the score and other displays as they appeared during the live performance.

In the case of the SOFA remixer, it is user interactions with the remixer interface that trigger updated display, playback of composed audio, and eventually the creation of a new musical work.

///



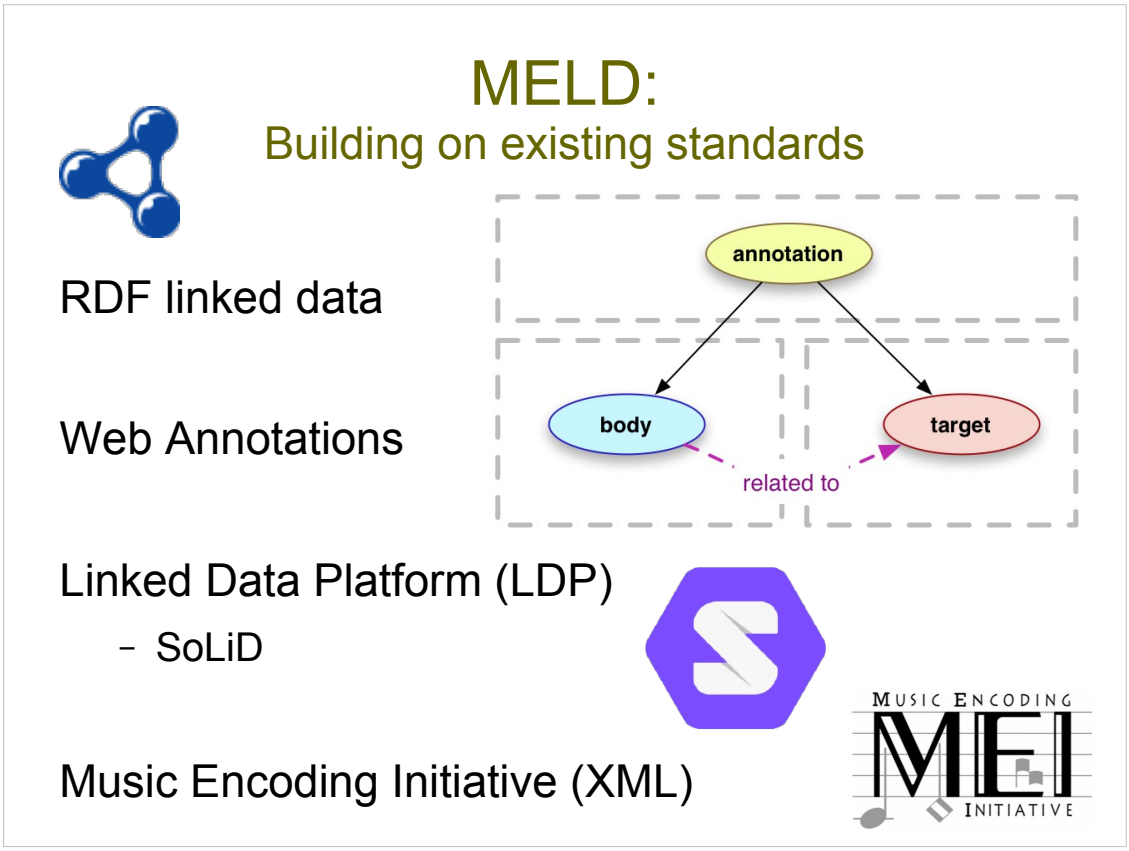
The role of MELD in these applications is to provide a framework that connects corresponding elements across diverse media.

It does this using linked data to make the connections, along with data that represents musically meaningful structure in a computationally accessible fashion that raw audio or score images can not. Typically, this is provided by MEI, but it could be any data format that represents musically meaningful elements, and allows instances of these elements to be individually identified by a URI.

In the case of MEI, addressing of musical elements is handled by using @xml:id attributes in the XML data, which can be referenced by the MELD linked data using URIs with fragment identifiers. (For the purposes of linked data, a URI can include a fragment identifier, even though when accessing the resource over the web, the fragment identifier is not usually sent to the origin server.)

The “Verovio” engraving software can be used to render a musical score image from MEI data. In so doing, element identifiers in the MEI are persisted to the resulting SVG image elements, allowing user interactions with the image to be connected back to the musical elements in the MEI.

///



MELD itself defines very little that is new. Rather, it brings together a number of established web standards and uses them in innovative ways. As such, it can benefit from re-use of existing software platforms and libraries.

One element of MELD not mentioned previously is the Linked Data Platform, or LDP. LDP is a simple HTTP-based protocol that allows both reading and writing of arbitrary resources, allowing parts of the web to be used like a simple file system.

We have recently started using a Solid data pod server as our implementation of LDP. This is a product of Tim Berners-Lee's Solid project at MIT to “re-decentralize the web”. As well as supporting the LDP interactions, it also provides a framework for user authentication and access control to data. The idea of decentralized operation enabled by separation of data storage and web application logic is quite important to some of our ideas for web applications based on MELD.

An important idea here is that linked data is being used, not as a stand-alone technology, but as an integral member of the technology family that makes up the World Wide Web. The strength of linked data here is its ability to connect arbitrary resources on the web (including things like people and musical works that are not themselves web documents), and as an open-ended framework which can accommodate new and unanticipated information about the related entities.

This approach uses linked data as something that is truly ***of*** the web, rather than just another graph database that is accessed over the web.

///

Questions / discussion

(Reminders)

- Linked data connecting other formats (esp. XML)
- MELD
  - LDP + Web Annotations + musical structure...
- SoLiD (re-decentralizing the web)
  - Separation of data custody from web applications
- EM Places
  - Time-constrained attributes and relations
  - Extending basic data model to capture context and provenance

This concludes my main presentation.

I've hopefully left plenty of time for questions and discussion, which may be directly about the applications discussed, or more broadly about linked data applications, particularly when dealing with data that are uncertain, or call for some degree of interpretation.

The topics shown here are intended as reminders, or prompts for possible avenues of discussion... but don't feel constrained in the questions you might ask!
///

summer school

Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

xmlsummerschool.com