

# Interactivity, distributed workflows, and thick provenance:

## A review of challenges confronting digital humanities research objects

Katrina Fenlon  
RO-2019  
2019-09-24  
IEEE eScience, San Diego

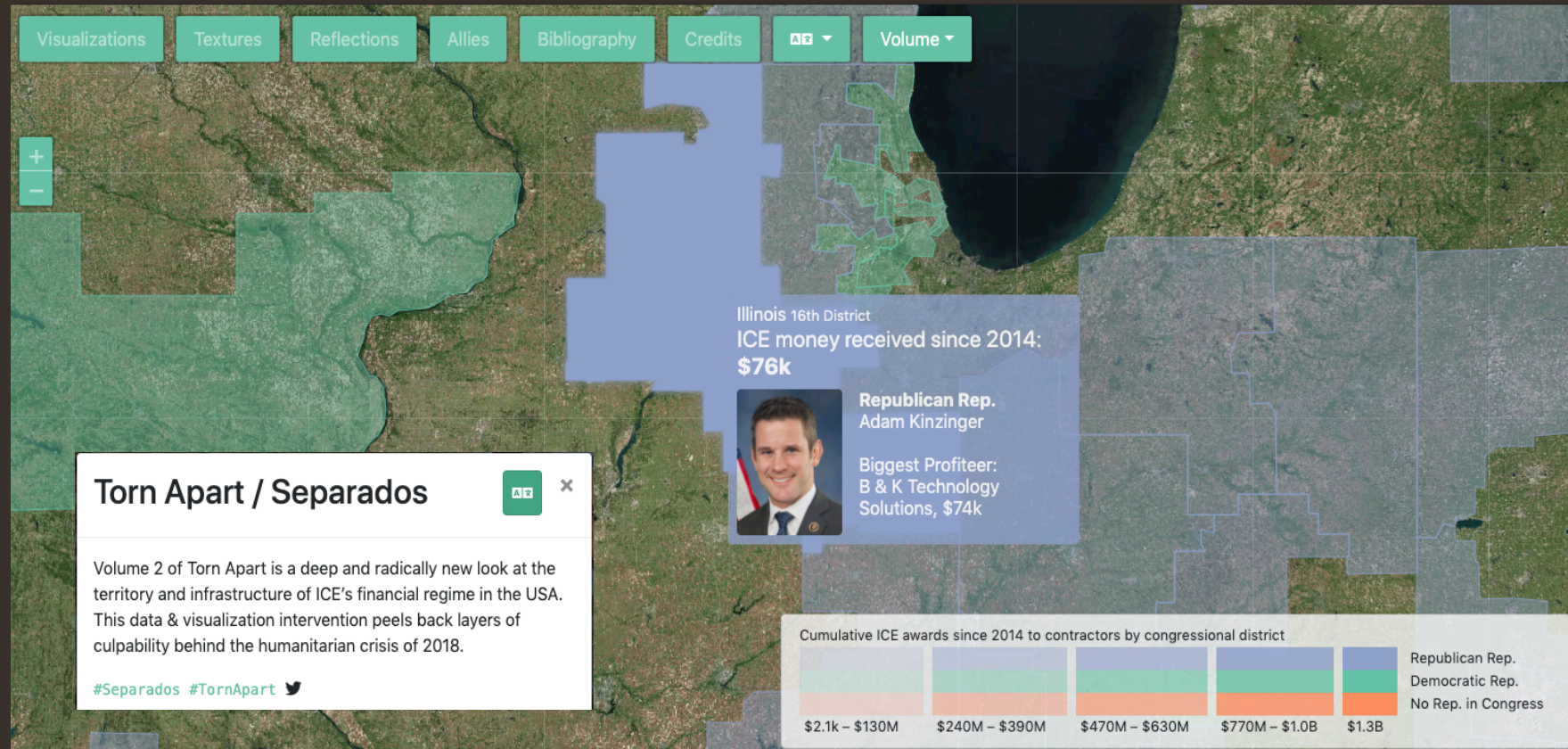


COLLEGE OF  
INFORMATION  
STUDIES

MARYLAND'S iSCHOOL



# Digital scholarship in the humanities



<http://xpmethod.plaintext.in/torn-apart/volume/2/index>

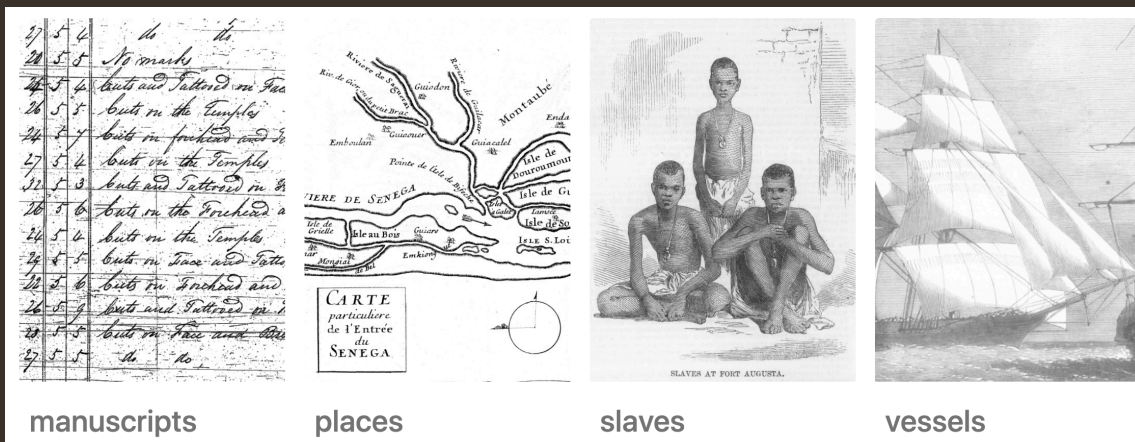


# Digital scholarship in the humanities

## Slave Voyages

### EXPLORE THE DISPERSAL OF ENSLAVED AFRICANS ACROSS THE ATLANTIC WORLD

This digital memorial raises questions about the largest slave trades in history and is available to answer them. European colonizers turned to Africa for enslaved labor resources of the Americas. They forced millions of mostly unnamed Africans across the Atlantic from one part of the Americas to another. Analyze these slave trades and view interactive animations to see the dispersal in action.



## Slave Voyages

### Trans-Atlantic Slave Trade - Database

Results Summary statistics Tables Data visualization Timeline Maps Timelapse

Showing 1 to 15 of 36,002 entries

Voyage ID	Vessel name	Place where voyage began IMP	Principal place of purchase IMP	Principal place of slave landing IMP
1	Pastora de Lima	Rio de Janeiro	Mozambique	Bahia, port unspecified
2	Tibério	Bahia, port unspecified	Mozambique	Bahia, port unspecified
3	Paquete Real	Bahia, port unspecified	Cabinda	Bahia, port unspecified
4	Rom Caminho	Bahia, port unspecified	Quilimane	Bahia, port unspecified

<https://www.slavevoyages.org/>

@kfenlon



# Digital scholarship in the humanities

**Frankenstein Volume I Draft in Chapter Sequence**

Author(s) : Mary Wollstonecraft Shelley  
Date Written : [August or September]-[?December] 1816  
Title/Literary Work : Frankenstein

State : draft  
Institution : Bodleian Library, University of Oxford

Hand(s) : Mary Shelley, Percy Shelley  
Shelfmark : MS. Abinger c. 56  
Folio : 4r

Transcription Status: ● ● ● ●  
Metadata Status: ● ● ● ●

Q Q Q ↺ ↻ Search LIMIT VIEW: ● All ● Mary Shelley ● Percy Shelley ⏪ ⏩ ⏴ ⏵



<http://shelleygodwinarchive.org/>



# Digital scholarship in the humanities

- Media-rich
- Data-centric
- Interactive
- Dynamic
- Interlinked
- Always evolving

**Slave Voyages**  
Trans-Atlantic Slave Trade - Database

Results | Summary statistics | Tables | Data visualization | Timeline | Maps | Timelapse

Showing 1 to 15 of 36,002 entries

Voyage ID	Vessel name	Place where voyage began	Principal place of purchase	Principal place of slave landing
1	Pastora de Lima	Rio de Janeiro	Mozambique	Bahia, port unspecified
2	Tibério	Bahia, port unspecified	Mozambique	Bahia, port unspecified
3	Paquete Real	Bahia, port unspecified	Cabinda	Bahia, port unspecified
4	Rom Caminho	Bahia, port unspecified	Quillimane	Bahia, port unspecified

Slave Voyages

## EXPLORE THE DISPERSAL OF ENSLAVED AFRICANS IN THE WORLD

This digital me  
available to ans  
resources of the  
from one part o  
animations to s

**Frankenstein Vol. 1**

Author(s): Mary Wollstonecraft Shelley  
Date Written: [August or September]-[?December] 1816  
Title/Literary Work: Frankenstein

State: draft  
Institution: Bodleian Library, University of Oxford

Hand(s): Mary Shelley, Percy Shelley  
Shelfmark: MS. Abinger c. 56  
Folio: 4r

Transcription Status: ●  
Metadata Status: ● ●

Search:  LIMIT VIEW: ☒ All ☐ Mary Shelley ☐ Percy Shelley

41 4

X yet there was an harmony in that very dissimilitude-

sevr servants had any request to make it always through <sup>her</sup> the intercession of Elizabeth For me I loved he We agreed perfectly although there were many <sup>There For, although</sup> was a great dissimilitude in our



The collage features a variety of logos and icons related to digital humanities. On the left, there are icons of a classical building, a book, a magnifying glass over a document, and an open book. The logos include Europeana Collections, MUKURTU, Open Archives, neo4j, Twitter, Hathi Trust, Perseus Digital Library (Gregory R. Crane, Editor-in-Chief, Tufts University), EEB (Early English Books), OJS (Open Journal Systems), Elsevier, Library of Congress, arXiv.org, Zotero, DPN (The Digital Preservation Network), Fulcrum, and Natural Language Analyses with NLTK. On the right, there are icons of a person at a computer, a person at a desk with a speech bubble, a person at a desk with a speech bubble, and a person at a desk with a speech bubble. The logos also include ATLAS.ti (Qualitative Data Analysis), hastac (Changing the Way We Teach + Learn), WordPress, and a network diagram.

@kfenlon



# Humanities research infrastructure

How to build infrastructure that can *sustain* digital scholarship broadly without attenuating its diversity?

Looking for (partial) solutions at the level of the **data model**

**RQ: To what extent can (scientific) research object data models accommodate digital humanities collections?**



# Prior work:

## Strengths and challenges of RO model for DH

- **Strengths of the model for DH collections are substantial:**

- ROs readily perform the most essential function of collections: aggregation
- ROs have capacity to accommodate rich, domain-specific relationships
  - Making hidden and vulnerable relationships among components explicit, usable, and therefore preservable
- ROs accommodate aggregations of linked data

(We'll come back to challenges)



# RO-like things in the humanities

- **Data-sharing architectures:** Perseids (Almas, 2017); Hedges and Blanke (2013)
- **Digital preservation and archive serialization:** Tarrant et al. (2009); Kouper et al. (2014)
- **Digital libraries for musicology:** Page, Lewis, and Weigl (2017)
- **Computational text analysis worksets** (HTRC)
  
- Suggest need for study of how the model could serve wider range of digital scholarship



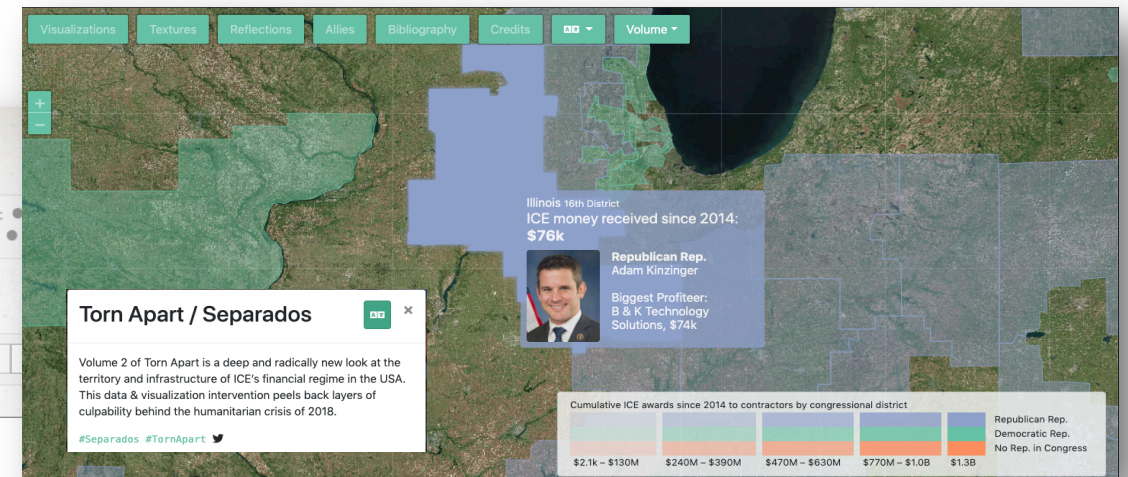
# Three challenges

1. Digital humanities scholarship requires specialized interactive use
2. Digital humanities scholarship relies on idiosyncratic, distributed workflows
3. Provenance description in the humanities is “thick”, complex



# Challenge 1: Essential interactivity for specialized use

- Much of humanities digital scholarship is *essentially* interactive
  - dynamic, responsive, participatory
- The scholarly contribution is realized by interactive components





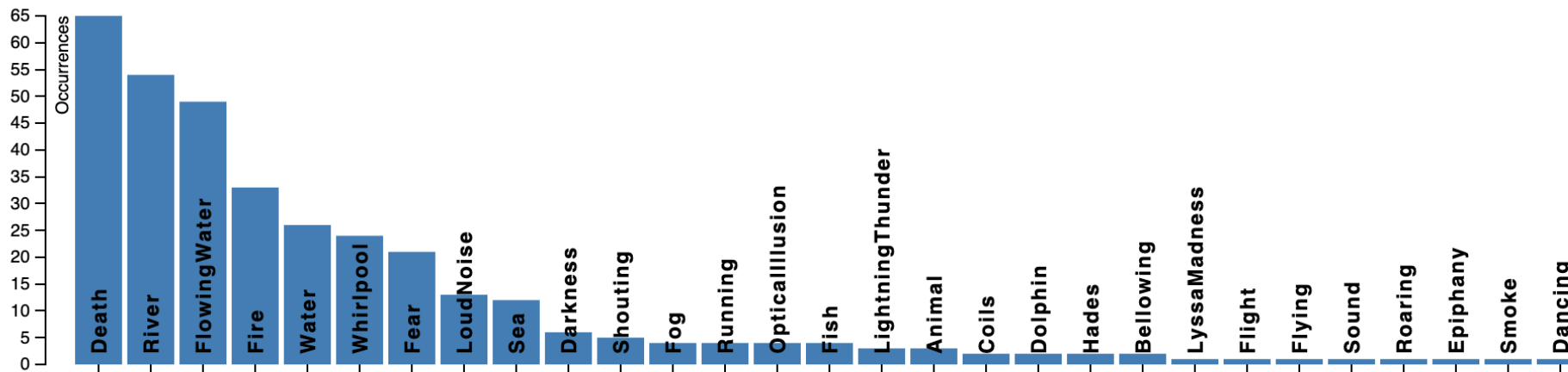
# Challenge 1: Essential interactivity for specialized use

- Much of humanities digital scholarship is *essentially* interactive
  - dynamic, responsive, participatory
- The scholarly contribution is realized by interactive components
- These components not usefully captured at level of RO model
  - But RO model could underpin platforms for custom development



# Challenge 1: Essential interactivity for specialized use

## Example: Perseids project



ὄχθαι δ' ἀμφὶ περὶ μεγάλ' ἴαχον· οἱ δ' ἀλαλητῶ  
ἔννεον ἔνθα καὶ ἔνθα ἐλίσσόμενοι περὶ δῖνας·  
ὥς δ' ὅθ' ὑπὸ ρίπῃς πυρὸς ἀκρίδες ἡερέθονται  
φευγέμεναι ποταμὸν δέ· τὸ δὲ φλέγει ἀκάματον πῦρ  
ὄρμενον ἐξαίφνης, ταὶ δὲ πτώσσουσι καθ' ὕδωρ·  
ὥς ὑπ' Ἀχιλλῆος Ἕάνθου βαθυδινήεντος  
πλῆτο ῥόος κελάδων ἐπιμίξ ἵππων τε καὶ ἀνδρῶν·  
αὐτὰρ ὁ διογενὴς δόρυ μὲν λίπεν αὐτοῦ ἐπ' ὄχθῃ  
κεκλιμένον μυρίκησιν, ὃ δ' ἔσθορε δαίμονι ἴσος  
φάσγανον οἶον ἔχων, κακὰ δὲ φρεσὶ μῆδετο ἔργα,  
τίπτει δ' ἐπιστροφάδην· τῶν δὲ στόνος ὄρουσ' ἀεὶ κέρε

But when they were now come to the ford of the  
fair - flowing river, even eddying Xanthus that  
immortal Zeus begat, there Achilles cleft them  
asunder, and the one part he drave to the plain  
toward the city, even where the Achaeans were  
fleeing in rout the day before, what time glorious  
Hector was raging — thitherward poured forth  
some in rout, and Hera spread before them a thick  
mist to hinder them; but the half of them were  
pent into the deep - flowing river with its silver





# Challenge 1: Essential interactivity for specialized use

## Implication:

Need RO implementations that create platforms for experimentation and collaborative development within distributed communities



# Challenge 2: Humanities research workflows

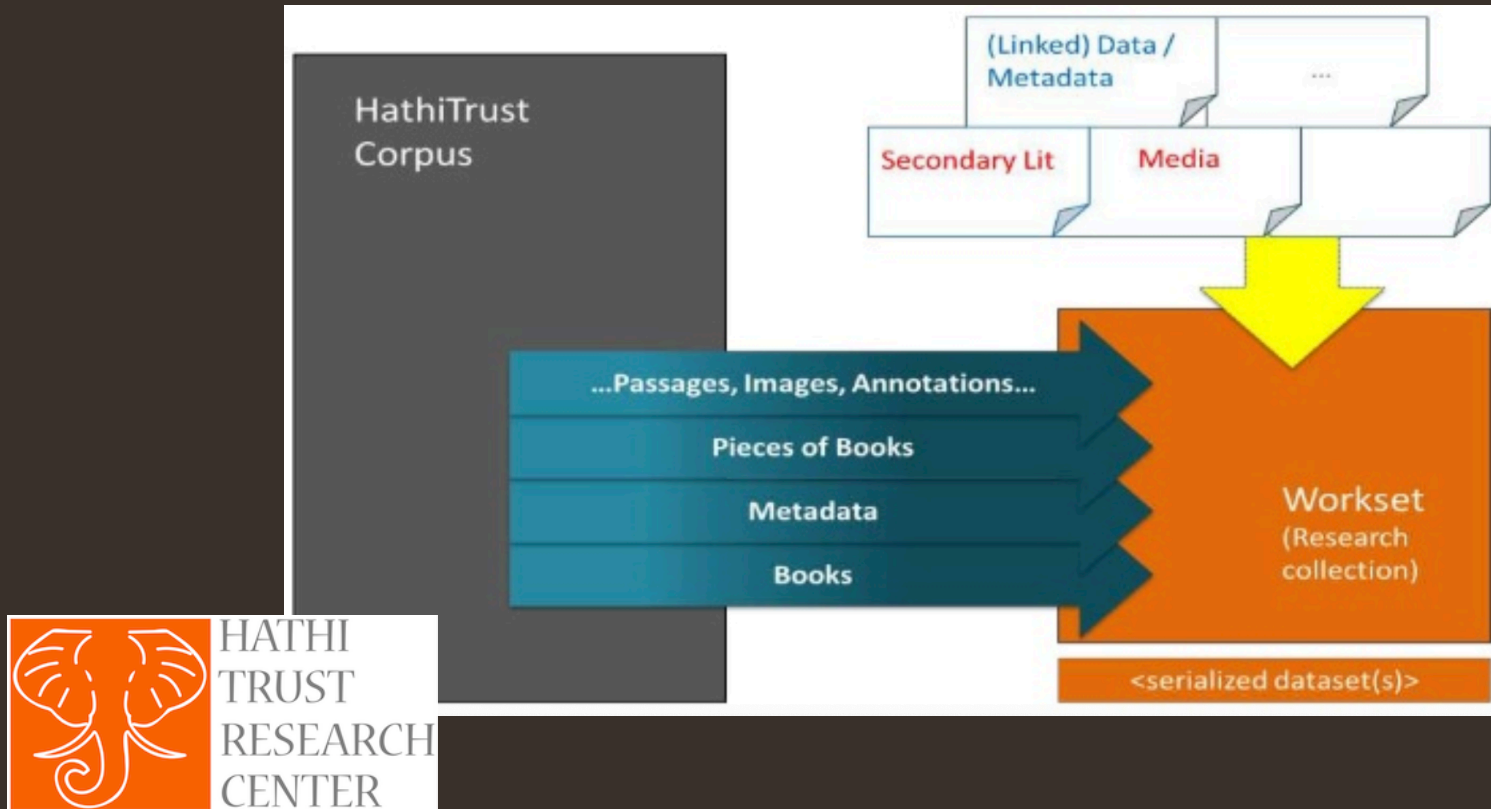
- Pretty well established *scholarly primitives* in the humanities
  - Discovery, collecting, reading, annotating, etc.  
but how these get cobbled together into workflows is...
- Idiosyncratic
- Collaborative
- Distributed
  - Sources / Collections
  - Analytic spaces
  - Teams
- Digital scholarship emerges from AND plays host to scholarly workflows!



# Challenge 2: Humanities research workflows

## Examples:

Constrained, task- or domain-specific workflow capture





# Challenge 2: Humanities research workflows

## Implication:

What is meaningful workflow capture/documentation in the humanities?

In other words,

What constrained kinds of / aspects of workflows in different domains are important to capture

for validation (authority and integrity of evidence), data reuse, reproducibility, evaluation of interpretive decisions, etc.?

Later:

What model extensions and implementation features do these require?



# Challenge 3: Thick provenance

- Fully documenting the provenance of digital scholarship would include documenting:
- Provenance of...
  - Data / digital objects as primary sources or evidence
    - Including multiple and conflicting provenances
    - Including expressions of uncertainty



# Challenge 3: Thick provenance

- Fully documenting the provenance of digital scholarship would include documenting:
- Provenance of...
  - Data / digital objects as primary sources or evidence
    - Including multiple and conflicting provenances
    - Including expressions of uncertainty
  - Annotations / claims made about the evidence
    - Including metadata, which is a site of controversy



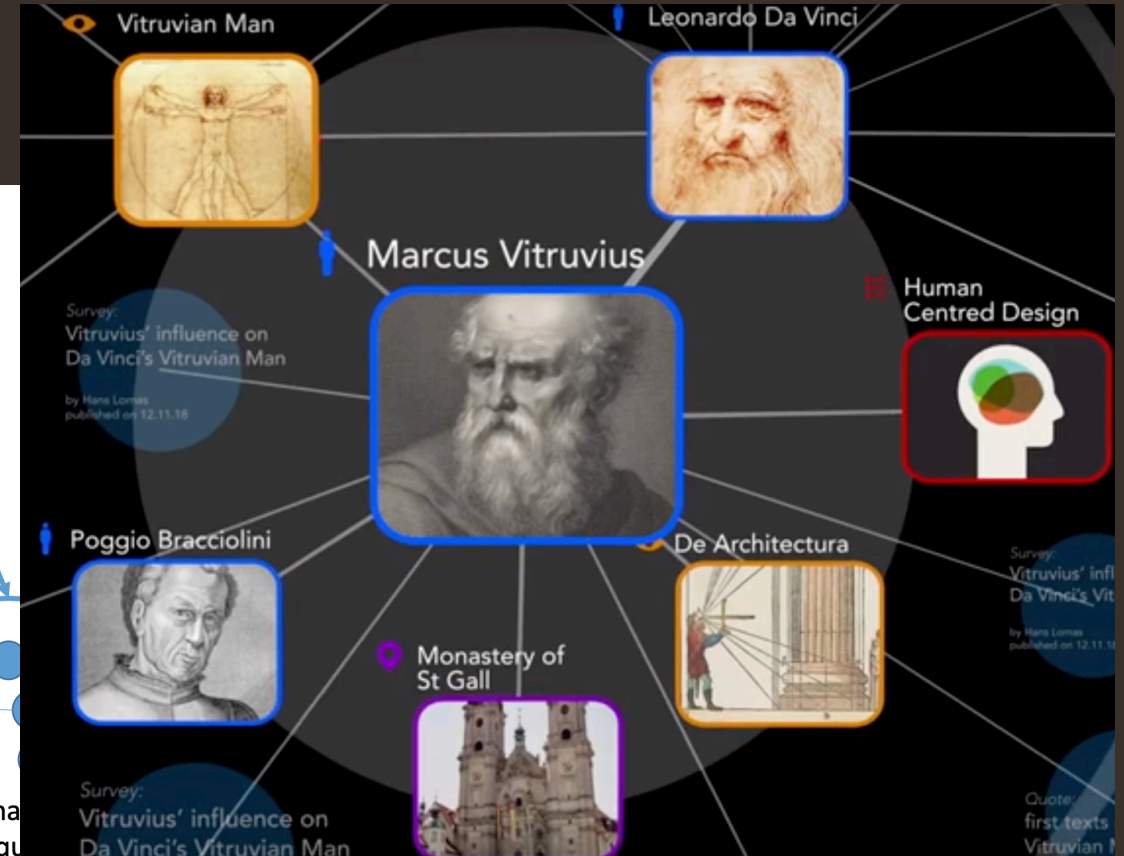
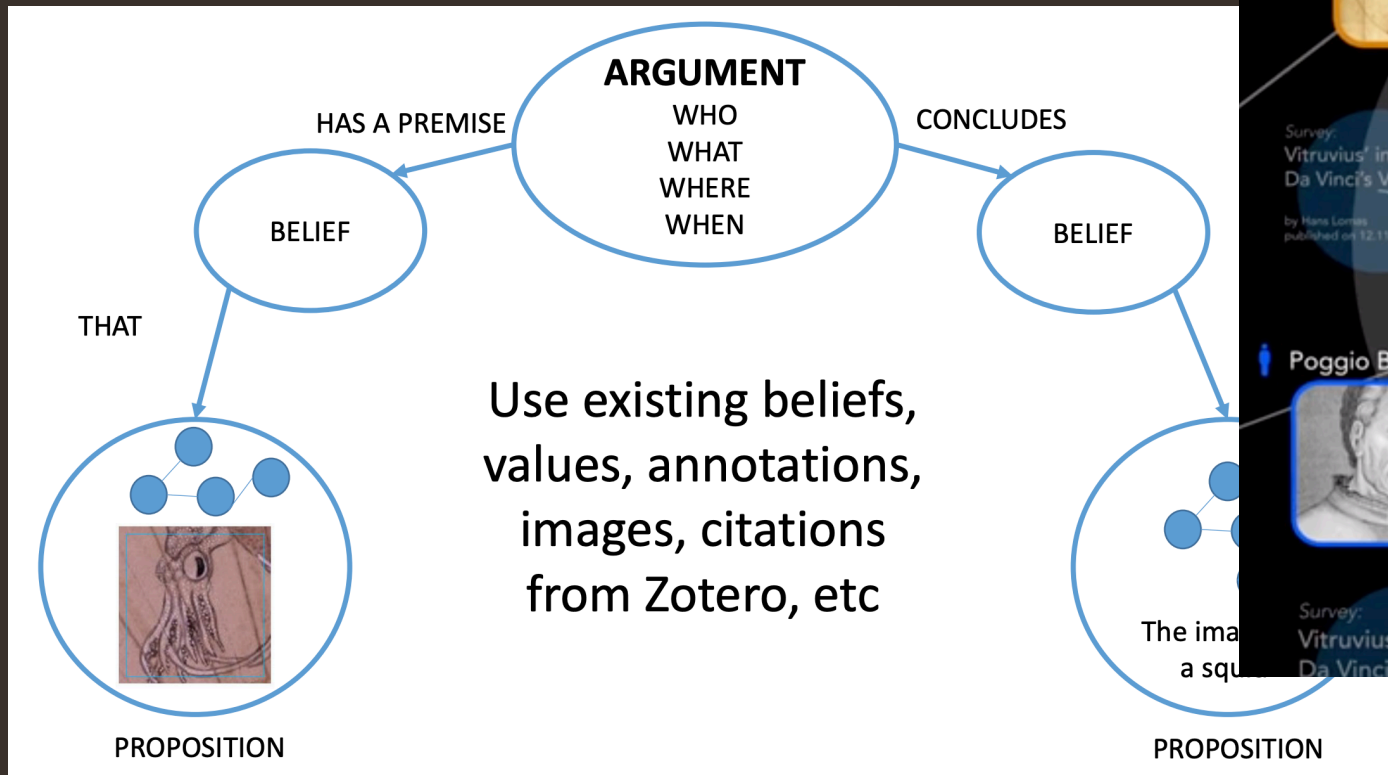
# Challenge 3: Thick provenance

- Fully documenting the provenance of digital scholarship would include documenting:
- Provenance of...
  - Data / digital objects as primary sources or evidence
    - Including multiple and conflicting provenances
    - Including expressions of uncertainty
  - Annotations / claims made about the evidence
    - Including metadata, which is a site of controversy
- Provenance claims themselves



# Challenge 3: Thick provenance

**Example:** ResearchSpace and CRMInf





# Challenge 3: Thick provenance

## Implications:

Provenance model *and implementations* must accommodate conflict assertions and formalize uncertainty

May need domain-specific provenance schemas and standards

The RO model has the capacity, but what would this look like in terms of implementation?



# Strengths and challenges of RO model for DH

## ○ Strengths of the model for DH collections are substantial:

- ROs readily perform the most essential function of collections: aggregation
  - ROs have capacity to accommodate rich, domain-specific relationships
    - Making hidden and vulnerable relationships among components explicit, usable, and therefore preservable
  - ROs accommodate aggregations of linked data
- In summary: **virtual aggregation and semantic annotation** are potentially very beneficial for a wide variety of humanities scholarship



# Strengths and challenges of RO model for DH

- **Strengths: virtual aggregation and semantic annotation** are potentially very beneficial for a wide variety of humanities scholarship
- **Challenges are significant but surmountable:**
  - Need RO implementations that create platforms for experimentation and collaborative development within distributed communities
  - Need better understanding of humanities research workflows
  - Need creative implementations to support thick provenance description



# Future work

- How to build RO implementations that establish dynamic platforms for experimentation, participation, and co-creation?
- How to extend the RO model for DH domain-specific relationships among primary, derived, secondary sources and workflows, people, and contextual entities?
- Potential intersections between ROs among other emergent publication models and linked data representations of library, archive, and museum collections
- More complete study of DH scholarly forms and workflows