

BOLOGNA 20 JAN 2020

# Mining Authoritativeness in Art History

Marilena Daquino, Francesca Tomasi  
/DH.arc, University of Bologna, Bologna, Italy

# outline

---

- **connoisseurship in art historical photo archives**

questionability of artwork attributions

- **measuring authoritativeness**

of secondary sources of attribution

- **recommending attributions**

leveraging Semantic Web technologies

# connoisseurship in art historical photo archives

scholars need to gather and compare plenty of documentation for supporting questionable statements

- 1 providers of heterogeneous sources**  
photos, books, catalogues, correspondence
- 2 witnesses of different scholars' methodologies**  
annotations on the back of photos (including attributions, provenance, bibliographies), consultancies
- 3 archivists' consultancy**  
old fashion recommending systems
- 4 online catalogues**  
possibility to query massive collections

# problems

what information technologies have not solved yet

- 1 unknown overlap between archival collections**  
lack of reconciliation services, hence overlaps are not evident
- 2 error-prone retrieval of sources**  
missing, partial or outdated information
- 3 questionability of statements**  
methodologies for validating attributions are not reproducible
- 4 hermeneutical aspects not explicit in data**  
no agreement on the definition of authoritativeness
- 5 data quality issues over time**  
updating data is expensive and time-consuming for institutions

# (negative) impact



- **lack of competence**

current aggregators cannot cope with **scholars' difficult tasks** and the **decision-making process** is demanded to the final user

- **affected reliability**

**poor data quality affects reliability** of cultural institutions over time

we need smart systems, based on real-world competencies, to cope with real-world problems

# the objective

- **reproduce the hermeneutical approach of second-hand knowledge providers when validating attributions**

make it reproducible and quantifiable

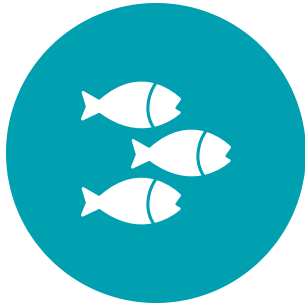
- **measure textual authority of secondary sources and cognitive authority of scholars**

dimensions characterizing authoritativeness of sources and people

- **support scholars and cataloguers in complex tasks by leveraging cutting-edge technologies**

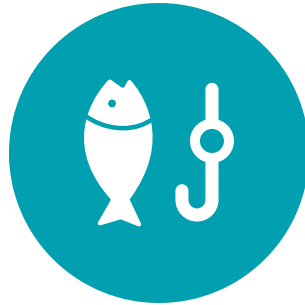
automatic methods for improving cataloguing metadata quality and recommend attributions to scholars

# approach



## ontology

- **formal representation** of argumentations around attributions
- facilitate **data integration**



## hermeneutics

- **formal definition** of authoritativeness
- **measurement and reasoning** on argumentations around attributions



## recommendation

- **user-tailored tools** to accomplish complex tasks
- support **decision-making process**

# approach



## domain analysis

review of definitions and guidelines in content and metadata standards

## data analysis

domain experts' consultancy and computational analysis over a corpus of cataloguing records for extracting features characterising hermeneutical aspects

## implementation

mAuth - a semantic crawler and recommender system of artwork attributions

## ontology development

HiCO - a task ontology for describing provenance and hermeneutical aspects related to attributions

## IQ measures design

definition of a priori (always valid) and non a priori measures and weight in a ranking model



# domain analysis

## 1 collect content standards, guidelines, metadata standards

CDWA, Getty Vocabularies, ICCD cataloguing rules, Spectrum (UK), LIDO, CIDOC-CRM

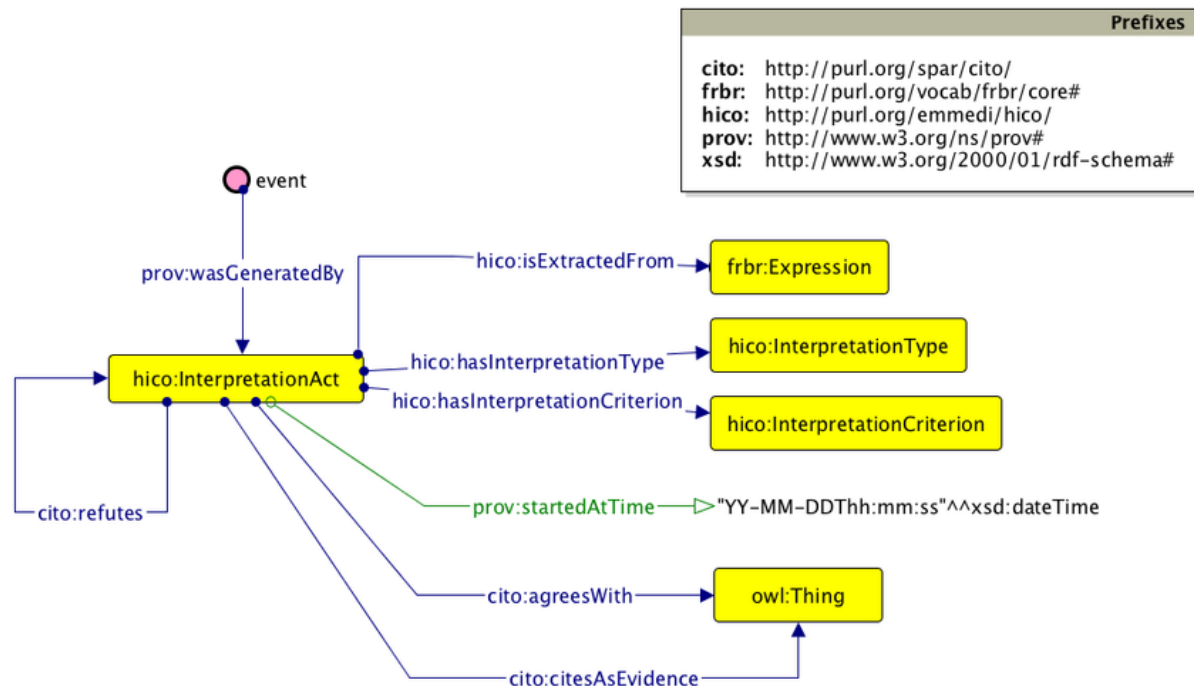
## 2 deduce which context information is fundamental to validate an attribution

e.g. an author, a criterion, a date, a source

## 3 extract controlled vocabularies for representing criteria used to justify an attribution

e.g. bibliography, museum attribution, auction attribution, note on a photograph

# ontology development



- **HiCO ontology**

extension of the PROV Ontology  
partially aligned to CIDOC-CRM

- **allows to describe**

- source where the attribution is stated
- classification of the attribution
- criteria
- time span
- sources supporting the statement
- relations between attributions

# domain analysis

however...

Guidelines do not include **how to weight criteria** when contradictory statements are available

**E.G. IS AN ATTRIBUTION CLAIMED IN A PEER-REVIEWED ARTICLE MORE RELIABLE THAN AN AUCTION ATTRIBUTION?**

---

Criteria may not be valid **a priori**

**E.G. A RECENT DISCOVERY IN THE ART MARKET MAY BE MORE RELIABLE THAN AN OUTDATED ARTICLE**

---

The final (subjective) decision **is up to cataloguers**, which are generally also art historians

# data analysis

## 1 domain experts' initial rating of criteria extracted from content standards

rating of reliability of criteria [scale 1-10]

## 2 validation of the rating over a representative LOD dataset

Analysis of the distribution of criteria in the Federico Zeri photo archive, Villa I Tatti - Berenson Library, and Frick Art Reference Library

# distribution of criteria

compare criteria supporting accepted and discarded attributions recorded in cataloguing records (e.g. Zeri photo archive)

tot.	ACCEPTED	DISCARDED	tot.												
				documentation	artist's signature	scholar's attribution	bibliography	archival classification	scholar's note on photo	museum attribution	inscription	sigla	auction attribution	collection attribution	
34	documentation	15	108			21	12	13	3				3		4
26	artist's signature		8	3	5	13	2	3				1	1		
2629	scholar's attribution	69	783	3	527	973	194	144	42	4	3	436	76	14	
1697	bibliography	17	2547	2	253	1288	199	81	29			96	32	16	
5322	archival classification	108	334	8	795	2585	328	315	88	6	4	700	153	2	
471	scholar's note on photo	6	318	1	49	218	53	110	5	1		58	12	2	
1	museum attribution		88		1		1		1						
3	inscription		6			3									

# IQ measures design

## 1 review of existing IQ measures

2 a priori measures:

**Reliability** (rating of criteria)

**Reputation** (list of trusted data providers, scholars' citation indexes)

2 variable measures:

**Relevance** (sources in agreement)

**Timeliness** (date of the attribution)

## 2 design of the ranking model

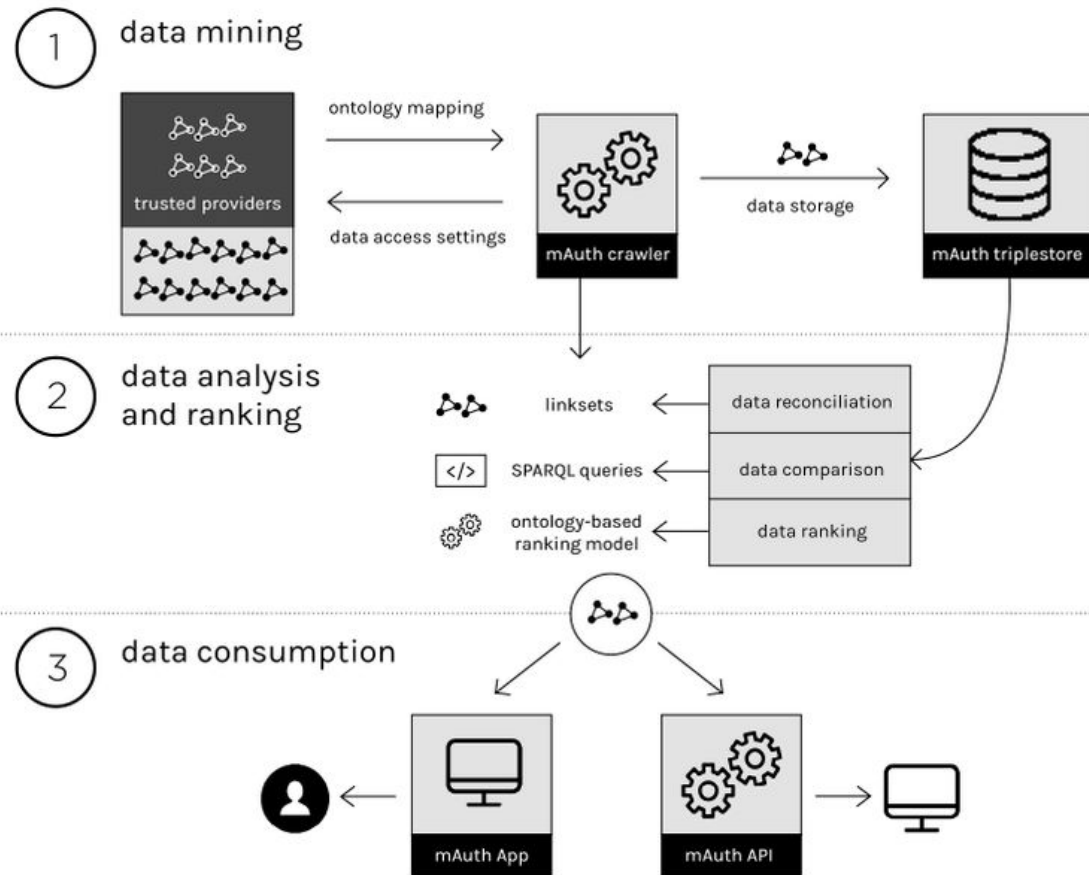
weight and interaction of measures in a ranking model

## 3 design of policies to automatically update cataloguing records on the basis of the ranking model

revise less recent bibliographic references, add criteria used by other providers, review questionable attributions

# implementation

Semantic crawler of artwork attributions + data analysis and ranking + recommendation



- **mAuth - mining authoritativeness in art history**

<http://purl.org/emmedi/mauth/search>

# an example



- **the Three Graces**

- records are available in 3 datasets (Zeri / I Tatti / Frick)
- different levels of reliability of contradictory attributions





Peruzzi Baldassarre

16.27

Zeri Foundation (University of Bologna) accepted attribution

1 ♥

**artwork title:** *Peruzzi Baldassarre , Tre Grazie*

**date(s) of attribution:** 1990;

**criteria:** archival classification; bibliography;

**source of attribution:** none;

**bibliography:** bibliografia specifica Morelli G. Della Pittura italiana: studii storico-critici ; le Gallerie Borghese e Doria-Pamphili in Roma 1991 144;  
bibliografia specifica Berenson B. Italian Pictures of the Renaissance - Central Italian and North Italian Schools 1968 vol. I, p. 334;  
bibliografia di confronto Freedberg S.J. Painting of the High Renaissance in Rome and Florence 1961 vol. I, p. 135; vol. II, p. 182;  
bibliografia specifica Venturi A. Storia dell'Arte Italiana 1901-1940 vol. IX, 5, p. 412;

[see more](#)

Peruzzi, Baldassare (1481-1536,  
Italian) artist

10.24

I Tatti - Berenson Library (University of Harvard) accepted attribution

1 ♥

**artwork title:** *Three Graces (fragment)*

**date(s) of attribution:** 1968;

**criteria:** archival creator's bibliography;

**source of attribution:** Bernard Berenson ( h index: 32.0; );

[see more](#)

Italian School (16th century),  
16th C, attributed to.

8.22

Frick Art Reference Library (NY) accepted attribution

0 ♥

**artwork title:** *The Three Graces.*

**date(s) of attribution:** 1952;

**criteria:** archival classification;

**source of attribution:** none;

[see more](#)

Luini Bernardino scuola

4.27

Zeri Foundation (University of Bologna) discarded attribution (1)

0 ♥

**artwork title:** *Peruzzi Baldassarre , Tre Grazie*

**date(s) of attribution:** 1994;

**criteria:** auction attribution;

**source of attribution:** Asta Christie's ( h index: 4.0; );

[see more](#)

# evaluation: user study

Results: <https://doi.org/10.6084/m9.figshare.7409384.v2>



## 31 users (20 domain experts)

- scholars, teachers, connoisseurs, art historians
- collection data managers
- DH and CS researchers



## 3 research scenarios

- all sources in agreement, only one domain expert
- only two sources in disagreement, not well-documented, both relying on scholars' attributions only (citation indexes contradict the ranking model)
- three sources, two in agreement and well-documented (citation indexes support the ranking model)



## benchmark of 3 systems

- PHAROS images
- Wikipedia
- Zeri catalogue



## Fleiss' kappa

fair agreement between domain experts

# conclusion

---

- **Semantic Web technologies fit for the purpose**

- support common tasks (harvesting and data integration)
- support more complex tasks (ranking, decision-making process, improve data quality)

- **authoritativeness on the web is still an open issue**

- **textual authority** works only in simple-medium tasks
- **cognitive authority** is a key element in Humanities, but it is hard to grasp since there are no representative citation databases for historical scholars in the Humanities

# thanks!

QUESTIONS?