Semantic Web 0 (2019) 1
IOS Press

Trust-based retrieval of artwork attributions in digital libraries

Using Semantic Web Technologies for Validating Argumentations around Attributions

Marilena Daquino [a,*] and Enrico Daga [b]
[a] *Digital Humanities Advanced Research Centre (/DH.arc), University or Bologna, Bologna, Italy*
E-mail: marilena.daquino2@unibo.it
[b] *Knowledge Media Institute, The Open University, Milton Keynes, UK*
E-mail: enrico.daga@open.ac.uk

Abstract. Validating argumentations around attributions is a well-known issue in the cultural heritage domain, where competing sources offer contradictory information on the same artefacts. To date, data aggregators allow users to retrieve heterogeneous information faster. However, contradictory information is rarely handled and argumentations are unlikely to be processed due to a number of limitations, namely: arguments are usually recorded in non machine-readable formats, attributions are not integrated with other sources on the web, there is no shared mechanism for ranking attributions, and data may suffer of Information Quality (IQ) issues over time. In this article we argue that Semantic Web technologies can effectively facilitate data harmonisation tasks, can support users' decision-making process when appraising online secondary sources recording artwork attributions, and can avoid expensive curatorial efforts to cultural heritage institutions. In detail, we introduce an ontology for representing argumentations around attributions, methods for measuring Information Quality in the Arts domain, and an ontology-based recommending system of artwork attributions. The aim is to demonstrate the suitability of Semantic Web technologies for solving trust-related problems in the Arts field, and highlight the portability of developed methods to near fields.

Keywords: Information Quality, Argumentations, Semantic Web, Trust, Art History

1. Introduction

To date, the Web offers the possibility to publish opinions and arguments on everything. Different viewpoints, based on different levels of knowledge, intentions, and reliability, populate the network. To this extent, trust-based retrieval mechanisms have recently became fundamental in several communities for evaluating and recommending reliable information to users.

Among the others, cultural heritage institutions are deemed high-quality information providers, and their digital libraries support scholars in a number of enquiries. Still, providers may publish competing and contradictory information on the same artefacts (e.g. different authors, dates, sources). A prominent case is the validation of competing artwork attributions, a well-known problem in the Arts field. Art historians collect existing attributions and compare argumentations around those so as to position their work in the state of the art and support their claims with authoritative references [1]. Despite several methodologies for validating the veracity of an artwork attributions have been proposed over time [2–4], these are not reproducible and many attributions are still debated. Secondly, collecting online information is time-consuming (contradictory sources are not linked with each other), error-prone (sources may not be updated), and potentially misleading (sources may be incomplete, can be biased by market interests, or may reference sources characterised by diverse degrees of reliability). However, the aim of cultural heritage institutions is to provide users with a comprehensive account on information related to artefacts (e.g. including both discarded attributions and currently accepted ones) and

*Corresponding author. E-mail: marilena.daquino2@unibo.it.

support users in their decision-making process. To this extent, both institutions and final users would benefit from reusing existing knowledge on the web (e.g. other sources of attribution) and means for validating their reliability, avoiding respectively curatorial efforts and time-consuming activities.

One the one hand, existing methods for publishing argumentations (e.g. forums, blogs, news websites) rarely address the structure of argumentative statements in depth and their semantics can only be loosely represented and leveraged in smart applications for supporting data integration and decision-making process [5]. On the other hand, art historical data aggregators, like Europeana[1] and Pharos[2], collect data that rarely include machine-readable data on argumentations. For instance, looking for the artwork called *Three Graces*[3] in images.Pharosresearch, three competing attributions are retrieved. Motivations are recorded in original data sources in natural language, that is, (1) the data structure is shallow, (2) there are no links between attributions, and (3) there are no insights on how providers respectively validated their recommended attribution among the existing ones.

Crucially, since there are no reproducible methodologies or bespoke metrics for validating attributions, no shared rating mechanisms are available and therefore, sophisticated questions such as "what is the most documented, shared, and authoritative attribution for the artwork at hand?" or "what is the most reliable information provider?" are completely demanded to the user's judgement.

In this article we argue that the Semantic Web can support users' decision-making process when gathering and evaluating argumentations around artwork attributions recorded in online secondary sources by providing the methods, models, and techniques for tackling the aforementioned issues. First, we introduce the Historical Context Ontology (HiCO), i.e. a task ontology for representing features characterising argumentations. We evaluate its fitness for purpose in facilitating users' tasks such as (1) gathering attributions along with relevant arguments for comparative purposes and (2) support their decision-making process. Secondly, we describe a number of metrics for enabling (1) content-based trust mechanisms, formally defining features characterising authority of secondary sources, (2) context-based trust mechanisms, selecting domain-independent measures that apply to the Arts field, and (3) reputation-based trust mechanisms, investigating features characterising scholars' authority. The result is a framework of Information Quality (IQ) measures implemented in an ontology-based recommending system for ranking artwork attributions. The recommending system is integrated in a proof-of-concept semantic crawler, called *mAuth*, that harvests Linked Open Data including information on artwork attributions. The recommending system has been evaluated by means of a user study. The evaluation allowed us to draw conclusions on domain-specific limitations and to discuss general opportunities and drawbacks derived from the usage of Semantic Web technologies to handle Trust in digital libraries.

This article extends previous work on the HiCO ontology [6, 7] and the computational analysis of art historical Linked Data for defining IQ measures [8] with novel contributions, namely: (1) we evaluate the HiCO ontology along with the application that leverages it for recommending attributions; (2) we present components and resources of the mAuth framework, that can be accessed, customised, and reused in different contexts; (3) we extend the evaluation of the recommending system in order to show benefits derived from the usage of Semantic Web technologies in the Cultural Heritage domain. Finally, we discuss results in the light of general considerations related to the Web of Trust, and we report limits and lessons learned.

The article is organised as follows. In Section 2 we provide insights on prior work on technologies for enabling trust-based retrieval mechanisms in Semantic Web and digital libraries. In Section 3 we describe the use case, including domain-specific terminology and preliminary considerations so as to facilitate the reader's understanding. In Section 4 we detail the research approach adopted to realise the ontology (further described in Section 5), the framework of IQ measures (described in Section 6), and the semantic crawler (described in Section 7). In Section 8 we present the user study performed to evaluate the ontology, the ranking model, and the IQ framework. In Section 9 we discuss results of the user study, and finally we conclude in Section 10 with lessons learnt and considerations on the portability of the approach in near research fields.

[1] https://www.europeana.eu/portal/en
[2] http://pharosartresearch.org/
[3] http://images.pharosartresearch.org/artworks/itatti/8000852449?compare

2. Related Work and Materials

In this work we present a feasibility study of Semantic Web technologies applied to trust-based retrieval tasks in the Arts field, and we contribute with resources for enabling content-based trust mechanisms in Arts digital libraries. In this section we present the stack of technologies implemented in this work, and we acknowledge prior work in measures for addressing trust-related problems in digital libraries.

Representing provenance and trust in the Semantic Web The architecture design of Semantic Web includes a layer dedicated to Proof and Trust [9], which should allow information recommendation on the basis of automatic reasoning and proof-checking. Such a layer is currently rather speculative, since it has not envisioned yet pragmatic solutions to allow agents to decide which claims are trustworthy or not due to the diversity of trust-based mechanisms that should be applied case-by-case. Researches in several fields resulted in a number of approaches for representing and measuring trust in Linked Open Data. Studies focus on context-based, reputation-based, and content-based trust mechanisms [10–13]. We acknowledge these approaches in our solution, however, there is no one-size-fits-all approach to the problem of trust in the Semantic Web.

On the one hand, provenance-based mechanisms have been widely investigated and applied in the Semantic Web. Technologies such as the PROV Ontology [14], and Named Graphs [10, 15] allow reputation-based and context-based mechanisms to be applied to a variety of domains. We build on top of such approaches for representing provenance information of harvested attributions and store them in observation graphs.

On the other hand, content-based mechanisms, i.e. mechanisms addressing how the nature of information affects trust judgment, are not formalised in the Semantic Web architecture design. This is due to several reasons, including data variety and data integration issues [16]. In fact, data (1) are stored in different collection systems, (2) present heterogeneity at schema and instance level, and (3) are not linked with each other.

Linked Data are served on the web in several ways. Services such as SPARQL endpoints and APIs, content negotiation, Linked Data Fragments, and data dumps are the most common ways to access and query data. SPARQL federated queries can be adopted to access Linked Data sources live and overcome vocabulary heterogeneity in the retrieval phase [17]. However, SPARQL endpoints may have limited availability, sometimes resulting in query timeout and downtime. Secondly, data dumps are not handy solutions for live query, nor are affordable in case of large datasets. In this work we implement separated queries to different types of access points when harvesting sources of attribution to speed-up the retrieval. Specifically, whenever available we preferred Linked Data Fragments servers [18], followed by SPARQL endpoint APIs, and content negotiation. However, querying heterogeneous data keep being an expensive task. Since sources in scope do not change often, harvesting data can be a once-in-awhile task. We store snapshots of data in a dedicated triplestore for versioning purposes and we sort them live when recommending attributions.

Semantic heterogeneity at schema level can be overcome by means of automatic alignment algorithms [19] or semi-automatic alignment methods and user validation [20]. Due to the limited amount of data harvested in this work, we manually align stakeholders' vocabularies to a crawling schema so as to ensure precision of retrieved metadata. Heterogeneity at instance level can be overcame in several ways. Data providers often reconcile named entities (e.g. people, places) to community and widely-adopted datasets (e.g. VIAF[4], DBpedia[5], Getty vocabularies[6]) so as to facilitate interlinking. Mapping languages [21], Named Entity Recognition (NER) services [22], and data cleaning tools [23] make such operation easier. However, for those entities that do not have a corresponding entity in such datasets (e.g. an niche artist, an artwork), the creation of a traversal link to similar less known datasets is often postponed, since selecting and accessing a broad number of data sources is time-consuming. We adopt existing reconciliation services[7] and fuzzy string matching[8] to interlink entities mentioned in sources of attribution that do not have any *sameAs* link to any established authority. Secondly, when such methods fail to detect a good candidate match (e.g. because metadata are incomplete or contradictory), we adopt existing image recognition tools (i.e. Pastec[9]). Finally, we use linksets [24–26] to

[4] https://viaf.org/
[5] https://dbpedia.org
[6] https://www.getty.edu/research/tools/vocabularies/
[7] https://github.com/ehanson8/linked-data-reconciliation-python-scripts
[8] FuzzyWuzzy, https://github.com/seatgeek/fuzzywuzzy
[9] http://pastec.io/

store *sameAs* links between entities so as to speed up the reconciliation of harvested sources of attribution.

Approaches for measuring trust in Arts digital libraries Metadata standards [27–30] and thesauri [31] for art historical data provide instructions on which sources should be used and which information should be included when recording attributions in cataloguing records [32]. However there are no guidelines on how to rate attributions when contradictory claims are available. As a consequence, it's not clear what trust mechanism is in place when cataloguers reviewed existing attributions and recorded a recommended attribution (e.g. whether they preferred to rely on attributions provided by auction firms, i.e. reputation-based, or on a well-documented expertise provided by a art historian, i.e. reputation-based and content-based).

Existing vocabularies [33–35] allow to represent aspects peculiar of the Arts domain in machine-readable formats and a number of projects (e.g. Linked.art[10], CORDH[11]) aim at harmonising art historical data for data exchange and reusability purposes. However, terms for representing argumentations around attributions are still lacking. We extend the PROV Ontology [14] to formally represent features characterising argumentations around attributions (the HiCO ontology), and we align the proposed model to golden standards (CIDOC-CRM) in order to foster its reusability.

Existing Information Quality measures [36, 37] address domain-independent context-related measures for validating whether information fits for purpose. Among the metrics for measuring reputation, citation indexes (e.g. h-index [38]) are extensively used. However, traditional reputation-based trust mechanisms are not easy to be implemented in the Humanities, since citation databases and indexes are not representative (e.g. do not include citations of verbal communications) or comprehensive (e.g. do not include citations of historians of the last century).

So far, there are no studies addressing content-based metrics for Art digital libraries. Indeed, methods for modelling and reasoning over argumentations [39] have not been applied to cataloguing data. While quantitative methods for assessing Information Quality have been applied in cataloguing practices to evaluate aspects such as completeness and consistency [40–42], these have never been applied to validate veracity of contents.

[10] https://linked.art/model
[11] https://www.cordh.net/

The main contribution of this work is a portable approach for reproducing the hermeneutical methodology in the Humanities. In particular, (1) we rely on knowledge acquisition with domain experts and data-driven validation to understand which content-related variables affect users' judgement, (2) we select few context-related measures from Naumann et al. [37] that apply to the Arts domain, (3) we tune existing metrics [38] and we develop bespoke ones for evaluating reputation-based variables.

3. Case Study

Art historians collect massive amounts of documentation to support their statements. When validating the veracity of attributions, historians evaluate a number of (1) content information – e.g. reliability of argumentations –, (2) information provider's reputation – both the reputation of historians that first ascribed the artwork and the reputation of the cultural heritage institution [43] –, and (3) context information – e.g. the number of cultural institutions in agreement on an attribution, whether information is recorded in updated, scholarly, and peer-reviewed evidences. Cataloguing data provided by art historical photo archives are privileged sources, since these (1) are the result of cataloguers' research on primary sources, (2) include prior attributions for historiographic purposes, and (3) include argumentations around attributions (both accepted and discarded) for validation purposes.

For instance, a historian looking for existing attributions related to the artwork called *Three Graces* will find records in catalogues of three art historical photo archives, namely: the Zeri photo archive[12], Villa I Tatti - Berenson Library[13], and the Frick Art Reference Library of New York[14]. The Zeri photo archive catalogue includes (1) the attribution currently accepted by the institute supported by plenty of bibliographic references, and (2) a prior discarded attribution claimed by an auction firm. Villa I Tatti - Berenson Library and the Frick Art Reference Library include only their currently accepted attribution, supported respectively by few bibliographic references and the cataloguer's statement of responsibility. Along with attributions, records include (1) the date of the attribution, and (2) the source of attribution (articles, books, auction

[12] http://www.fondazionezeri.unibo.it/en
[13] https://itatti.harvard.edu/berenson-library
[14] https://www.frick.org/research/library

catalogue) and (3) the type of source (bibliography, auction attribution, cataloguer's decision). Arguments around attributions are expressed in natural language. Neither cited evidences nor other existing attributions recorded in other archive catalogues are linked to the record, and types of cited sources are not classified according to any scheme for defining their reliability.

That is, a user has to deduce that (1) two attributions out of three are supported by the same type of source, i.e. bibliography, (2) there is an overlap between cited sources (Villa I Tatti cites a subset of sources mentioned by Zeri), and (3) The Zeri Foundation relied on some internal rating of sources (i.e. scholarly literature was preferred over an auction attribution). Moreover, overlapping cited sources include Bernard Berenson's work, which is deemed an authoritative art historian.

It's worth to notice that trust mechanisms are key aspects in the Arts and Humanities. On one hand, reputation-based trust mechanisms affect users' perception of contents. Information providers' authority is often deemed sufficient for judging the goodness of the information itself [4]. As a consequence, contents are said to inherit authoritativeness from its providers, namely (1) the first-hand provider, e.g. the scholar that first ascribed the artwork to an artist, and (2) the second-hand knowledge provider, i.e. the cultural heritage institution that ensures the quality of the issued information [43]. In the prior example, scholars' attributions recorded in peer-reviewed articles were preferred by two out of three providers.

On the other hand, content-based trust mechanisms may hinder reputation-based trust mechanisms. Maintaining cataloguing data is an expensive task for cultural heritage institutions, and data are likely to be affected by information quality issues over time. For instance, biased information may have been recorded (e.g. due to market interests) and not being updated lately, or may include partial information (e.g. not fully documented attributions). This has a negative impact on research results and may affect data providers' reliability over time. Following the prior example, the third provider (Frick) records the oldest attribution, not supported by any source.

Lastly, some context-related aspects influence the final decision. For instance, a common belief is that the most recent attribution can be deemed reliable, and the number of cultural institutions in agreement on an attribution can influence the user's judgement. Following the prior example, domain experts from the third institution (Frick) that participated to the user study presented in Section 8, acknowledged the possibility to change their attribution after the comparison with Zeri and I Tatti.

In conclusion, a formal definition of authority in the Arts and Humanities domain is still an open issue that regards both secondary sources (cataloguing records) and cited primary sources of attribution (scholars' claims). Several gaps and challenges can be highlighted, namely:

- Different types of arguments and evidences supporting conclusions are unlikely to be formally represented. Argumentations are expressed in natural language and include implicit knowledge (internal trust-based mechanism).
- Reliability is both a domain-independent and domain-dependent variable. The latter mainly depend on content-based trust mechanisms (e.g. preference of certain types of sources), which are hard to grasp and to reproduce.
- Sources that present contradictory viewpoints are generally not integrated with each other. Existing services do not provide semantic links among argumentations and cited sources, people, and the period of validity of a claim.

4. Approach to the Research

This work aims at assessing authoritativeness of artwork attributions recorded in online secondary sources by means of Semantic Web technologies. Specifically, we (1) address features characterising authoritativeness of argumentations around attributions that affect users' judgement, and we (2) provide effective means for supporting data harmonisation, sense-making, and knowledge discovery in the Arts domain. Gaps in the representation and reasoning on argumentations around attributions can be summarised as follows:

- Lack of a coherent approach in documenting argumentations around attributions.
- Lack of metrics for measuring (1) Information Quality of contents, along with (2) information providers' reputation and (3) context variables.
- Lack of effective means for supporting users' decision-making process when comparing and validating reliability of argumentations.

We cope with the three aforementioned problems by leveraging Semantic Web technologies whenever applicable, so as to test their suitability in solving complex tasks. For this purpose we set up a use case in the Arts domain. In detail:

- We address the formal representation of argumentations around artwork attributions by leveraging well-grounded formal languages and technologies, i.e. ontologies.
- We formalise a set of IQ dimensions by reviewing cataloguing rules, leveraging domain experts' consultancy, and validating their opinions analysing a representative RDF dataset organised according to the aforementioned ontology.
- We implement a semantic crawler and an ontology-based ranking model for retrieving Linked Open Data and recommending artwork attributions.

The ambition is to reproduce the methodology adopted by cataloguers and scholars in the Humanities (meaning the decision process) by means of a portable, domain-independent approach. Indeed, in our approach we combine knowledge acquisition and data-driven validation for addressing domain-dependent variables, and we reuse existing quantitative measures for addressing domain-independent variables. In summary, seven actions (from A1 to A7) were undertaken to achieve the final ranking model.

1. **(A1) Ontology development.** Cataloguing standards are reviewed so as to identify aspects characterising argumentations around attributions. Existing ontologies are reviewed so as to map requirements to a formal representation. An extension of existing ontologies is proposed.
2. **(A2) Knowledge extraction of content-related variables.** A representative corpus of cataloguing records is created. A controlled vocabulary of terms identifying types of cited sources and criteria motivating attributions is extracted from the corpus. Terms are reconciled to discursive arguments in the corpus.
3. **(A3) Knowledge acquisition with domain experts.** Domain experts review the controlled vocabulary of terms (A2) and provide a rating of those according to their perceived degree of reliability.
4. **(A4) Data-driven validation of content-related IQ measures.** The consistency of the rating (A3) is validated by analysing the distribution of terms in the corpus, so as to confirm whether arguments are consistently used for supporting accepted attributions over discarded attributions with a lower degree of reliability.
5. **(A5) Selection of context-related IQ measures.** Context-related IQ dimensions affecting users' judgement are selected from literature according to domain experts' consultancy.
6. **(A6) Development of reputation-based measures** Two bespoke metrics are developed for measuring scholars' reputation.
7. **(A7) Ranking model.** IQ measures are weighted and combined in the ranking model.

Results of above actions include:

1. The HiCO ontology for representing argumentations (A1).
2. A framework of IQ measures for measuring content/context/reputation-related variables (A2-A6).
3. A ranking model for recommending sources of attributions (A7).

In order to demonstrate the fitness of the vocabulary and the soundness of the ranking model, we developed an artefact for harvesting and consuming art historical data. The artefact leveraging findings here presented is called *mAuth - mining authoritativeness in art history*[15]. mAuth is a proof-of-concept semantic crawler that harvests attributions and arguments from Linked Data sources and recommends attributions according to the ranking model.

To assess the validity of our approach, an user study with domain experts and scholars in near research fields is set up. Users are asked to perform researches on the corpus of art historical data by means of a web application implementing the semantic crawler and the ranking model and to fill in a form.

In the next sections we present results of this research grouped by actions, namely: the ontology resulting from A1 is illustrated in Section 5; the IQ measures and the ranking model resulting from A2-A6 are shown in Section 6; and the semantic crawler implementing the ranking model (A7) is presented in Section 7.

5. The HiCO Ontology

The cataloguing of cultural objects is the result of a hermeneutical activity made by one or more cataloguers. Cataloguing records can be seen as complex assertions on intrinsic and extrinsic aspects of objects they describe. Information included can be questionable, may change over time, be incomplete, or super-

[15] http://purl.org/emmedi/mauth/search

seded by new findings. The validity of pieces of information is bound to a number of constraints, here called *context layers*. Such layers can be defined as follows:

- The *context of an object* includes any statement on entities involved in its life-cycle (e.g. people, places, dates). Statements on the context of a cultural object answer questions such as: Who is the author of the artwork? When was the artwork created? Where was it created? How was it created? Who influenced the realisation of the artwork?
- The *context of a statement* includes provenance information on the aforementioned statements, such as dates, sources, and criteria supporting the statements. These statements answer questions such as: What type of claim is it? Who claims that? When was it claimed? What is the primary source of the statement?

Moreover, when catalogue information comes in a machine-readable format, a third context layer applies:

- The *meta-context of a statement* includes provenance information on the machine-readable version of aforementioned statements so as to verify their extent of the validity. It answers the questions such as: Who is responsible for the machine-readable version of the statement? Where is it extracted from? When was it extracted?

Such three layers of context all together define the provenance of an assertion related to an artefact as recorded in a specific source. Information at the first level are usually represented by means of domain ontologies, e.g. CIDOC-CRM [34]. The third layer can be represented by means of annotations to Named Graphs [15]. For instance, the statement *The "Three Graces" is created by Perruzzi Baldassarre* (context of an object), is a claim made by a cataloguer, recorded in a cataloguing record, stored as a metadata document (context of an object), and transformed into RDF statements (meta-context of a statement).

The Historical Context Ontology (HiCO)[16] [6], is an OWL 2 DL ontology developed for representing the second layer of context, i.e the provenance of a statement. HiCO is a task ontology, meaning it addresses aspects related to a single, domain-independent, representational issue. In particular, it addresses features characterising hermeneutical activities generating a piece of information.

Crucially, HiCO addresses all the features that contribute to define the degree of reliability of a statement. Following prior example, the cataloguer cites bibliographic sources to support the statement, and records the date of the cited attribution.

As a good practice, existing ontologies have been directly reused in HiCO (prefix `hico`) so as to represent specific aspects, namely: the PROV-O ontology (prefix `prov`) [14] was used to describe the provenance of a statement and it was extended so as to describe features of argumentations around statements, such as motivations and primary sources; the CiTO ontology (prefix `cito`) is imported to describe relations between attributions, sources of information, and involved agents; an OWL DL 2 formalization of the FRBR model (prefix `frbr`) [44] was considered for describing sources of information, such as cataloguing records or cited sources.

Figure 1 shows classes (rectangles), object properties (solid lines beginning with a solid circle and ending with a solid arrow), and assertions among classes (solid lines ending with a solid arrow). The main class of HiCO is `hico:InterpretationAct`. An interpretation act is a situation in which a statement about an event is linked to all the pieces of information necessary to validate its reliability (i.e. the context of the statement). This includes the following aspects:

- The classification of the interpretation, e.g. being an artwork attribution.
- The description of the type of sources or criteria motivating the statement, e.g. bibliography, scholar's attribution.
- Cited sources of information, e.g. a bibliographic source, an oral communication.
- The temporal extent of the attribution, i.e. when it was claimed.
- The document where from RDF statements are extracted, e.g. the cataloguing record.

For instance, statements about the creation of an artwork are related to individuals of the class `hico:InterpretationAct` by means of a property of the PROV-O Ontology, i.e. `prov:wasGeneratedBy`. Individuals of `hico:InterpretationAct` class are defined by means of a number of object properties, namely:

- `hico:hasInterpretationType`. The arbitrary classification of the interpretation, such as being an authorship attribution or a date attribution.

[16] http://purl.org/emmedi/hico

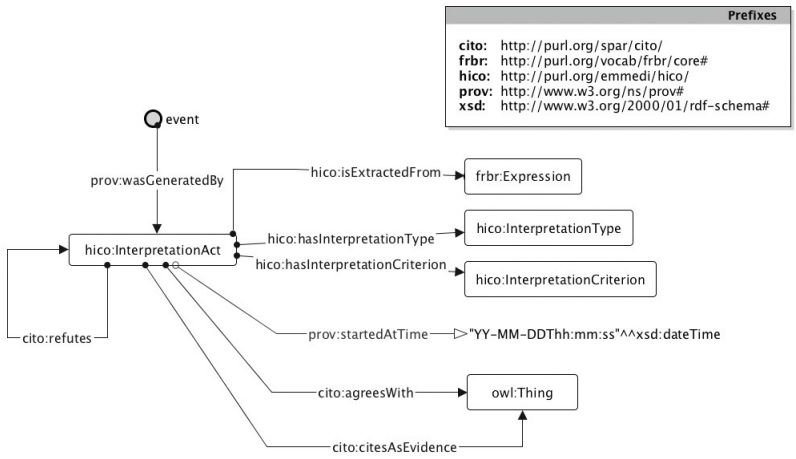

Fig. 1. The HiCO Ontology main classes and properties

- `hico:hasInterpretationCriterion`. The criterion or type of source supporting the interpretation, e.g. usage of bibliography, quotation of a scholar's opinion. Terms should be taken from a controlled vocabulary.
- `cito:citesAsEvidence`. Sources of information.
- `cito:agreesWith`. Statements or people in agreement with the statement.
- `cito:refutes`. Contradictory attributions on the same subject.
- `prov:startedAtTime`. The date of the attribution.
- `hico:isExtractedFrom`. The content of a metadata document (`frbr:Expression`) including the attribution. The property is defined as a subproperty of `prov:wasInfluencedBy`. The responsible entity for the attribution is the creator of the metadata document.

In this research, the HiCO Ontology is applied to represent argumentations around artwork attributions. It is used along with a controlled vocabulary of terms (i.e. individuals of the class `hico:InterpretationCriterion`) describing sources and criteria adopted by cataloguers. The controlled vocabulary is the result of the data analysis performed over a representative corpus of art historical data presented in the next section.

The ontology is evaluated as a component of a web application leveraging it (i.e., mAuth). The aim is to (1) assess the fitness of the vocabulary for representing argumentations around attributions, and to (2) confirm the suitability of the ontology for enabling further reasoning on the goodness of attributions. The evaluation is discussed in Section 8.

6. IQ Measures and Ranking Model

In order to effectively rank contradictory attributions, we designed a portable methodology for (1) addressing factors affecting the judgement of contradictory attributions, and (2) defining a shareable set of measures.

First, we perform a corpus analysis in order to address domain-dependent and content-related factors that influence cataloguers' decision. Secondly, we select context-related and reputation-related metrics from literature that contribute, along with content-related factors, to reproduce the hermeneutical activities of a cataloguer. We acknowledge the lack of metrics for reputation-based trust mechanisms that apply to the Arts domain, due to (1) the lack of citation indexes for art historians in the last century, and (2) the heterogeneity of citation forms. Hence, we develop bespoke metrics for the purpose.

The methodology is applied to a representative use case, namely a corpus of cataloguing records belonging to art historical photo archives. Results of this work are: (1) a rating of types of sources or criteria adopted by cataloguers when recording argumentations around attributions (Section 6.1), (2) a number of IQ measures selected from literature (Section 6.2), (3) two bespoke IQ measures realised for measuring scholars' authoritativeness (Section 6.3), and (4) the final ranking model including all of the IQ measures (Section 6.4).

6.1. The rating of types of sources and criteria

The objective of the corpus analysis is to identify, classify, and rate, types of sources and criteria adopted by cataloguers when supporting attributions. Here we summarise the key components of the approach that is detailed in [8].

Records collected belong to three relevant art historical data providers, namely: the Federico Zeri Foundation (19.061 records), Villa I Tatti - Berenson Library (12.256 records), and the Frick Art Reference Library (10.207 records). The corpus is gathered on a subject base, i.e. records on artworks of the Modern Art. Records include both accepted and discarded attributions. Argumentations appear in the form of discursive text fields, which can be reconciled to (one or more) terms identifying types of cited source or criteria. For instance, an attribution motivated with the statement “Federico Zeri’s attribution (1979)" is classified as *Scholar’s attribution*; “Christie’s attribution (1928)", is classified as *Auction attribution*.

Cataloguing metadata are transformed into RDF according to the HiCO Ontology, and the dataset is queried for knowledge extraction purposes[17]. The analysis was performed over the Linked Data version of the corpus rather than the original XML collections since (a) data cleansing and data reconciliation techniques have been applied, and (b) the semantic interoperability makes easier the comparative analysis.

We extracted a controlled vocabulary of terms classifying types of cited sources (e.g. “Bibliography", “Scholar’s attribution", “Auction attribution") and we reconciled arguments to such terms. Terms were initially deduced from the literature review of cataloguing standards and extended during the corpus analysis.

Domain experts’ reviewed the controlled vocabulary and provided a first rating of terms based on their perceived reliability. The rating is achieved by using a 1-10 scale.

We validated the rating provided by domain experts by analysing the distribution of terms in the dataset. In detail, given a subset of records including both accepted attributions and discarded attributions for the same set of artworks, the rating associated to types of sources supporting accepted attributions is compared one-by-one to the rating of types of sources that support discarded attributions. The aim is to highlight potential inconsistencies in the usage of the rating when contradictory attributions are available for the same artwork[18]. The rating provided by cataloguers is revised and extended according to results.

N.	Term	Score
1	Documentation	10
2	Artist’s signature	10
3	Archive creator’s attribution	9
4	Archive creator’s bibliography	8
5	Bibliography	7
6	Archival classification	7
7	Archive creator’s note on photograph	7
8	Scholar’s attribution	6
9	Museum attribution	5
10	Scholar’s note on photograph	5
11	Inscription	5
12	Sigla	5
13	Auction attribution	4
14	Collection attribution	4
15	Market attribution	4
16	Traditional attribution	4
17	Stylistic analysis	3
18	Anonymous note on photograph	3
19	False signature	2
20	Caption on photograph	2
21	Other	2
22	None	1

Table 1
The controlled vocabulary of types of sources and their rating

Table 1 shows the final vocabulary of terms identifying types of sources accompanied by their rating (see table 1).

6.2. Identification and selection of IQ measures

The corpus analysis showed that the rating is not always consistent. Reliability of sources may be affected by external factors. For instance, more recent scholars’ verbal communications (i.e. “Scholar’s attribution") may be judged more reliable than attributions stated in older bibliographic references (“Bibliography"), despite the latter is generally deemed a more reliable type of source.

To understand what variables may influence the user’s judgement and therefore affect the reliability of sources and criteria, we reviewed existing IQ measures. We selected a number of existing domain-independent dimensions and measures from Naumann

[17] Data sources and RDF dumps of the datasets are available at https://github.com/marilenadaquino/mauth/tree/master/data

[18] Results of the corpus analysis are available in [45]

et al. [37]. Dimensions are pruned so as to include only measures that apply to the Arts domain. The selection is made according to online guidelines [28, 32], domain experts' consultancy, and aspects highlighted by the corpus analysis, which shows that recentness of an attribution, authoritativeness of information providers and authoritativeness of cited sources are the main factors influencing the rating of criteria.

For each dimension we define an assessment method. In detail, we selected two Subject criteria (i.e. features that depends on the observer's perspective), namely *relevance* and *reputation*, and two Object criteria (i.e. features that characterise the information source), namely *reliability* and *timeliness*.

Relevance is the extent to which information is applicable and helpful for the task at hand. To assess it, we rely on a list of data providers which are likely to include artwork attributions. In this research, the list of data providers includes the three aforementioned art historical data providers and three multipurpose datasets, namely: DBpedia, Wikidata and VIAF. A common belief in the Arts field is that the more sources agree on a certain attribution, the more such an attribution is likely to be the most relevant among the contradictory ones. We measure relevance by counting the number of sources in agreement on a certain attribution.

Reputation is the extent to which information is highly regarded in terms of source or content. We assume that reputation of information can be inherited by second-hand knowledge providers and first-hand knowledge providers' reputation [43]. Second-hand knowledge providers' reputation (i.e. data providers) can be evaluated by relying on third party opinions. In particular, providers that are part of the aforementioned list of data providers are flagged as domain experts or non experts. Secondly, reputation of cited scholars is measured by means of two bespoke metrics for measuring their authority.

Reliability is the extent to which information is correct and trusty. According to domain experts, reliability of criteria and types of sources supporting an attribution are the most important means to validate its reliability. We measure reliability of an attribution by using the rating of twenty two terms extracted from the corpus analysis.

Timeliness is the distance between the date of the information and the retrieval date. A common belief in the Arts domain is that the latest recorded attribution - assuming it is also well-documented - is likely to be the most reliable. Timeliness of an attribution is measured by calculating the difference between the date of retrieval and the date of the attribution.

6.3. Metrics for measuring scholars' reputation

Authoritativeness is a key aspect when reviewing contradictory attributions in the Arts and Humanities. Metrics for measuring how scholars are perceived (1) in the community, and (2) with regard to specific artists, are fundamental aids for final users.

Traditional metrics include citation-based indexes and networks [38]. Unfortunately, there is a lack of such measures for the Arts and Humanities field, due to several factors including: (1) the absence of bibliometric indicators for publications in this field, (2) a significant amount of Humanities journals and books are not indexed in citation indexes, (3) there are no indexes for those art historians that worked at the beginning of the 20th Century, and (4) the diversity of citation forms in the Humanities (e.g. citation of a note on the back of a photograph) is not well-represented by indexes. Moreover, not only scholars but organizations and institutions (e.g. auction firms) are cited as sources of information.

To fill this gap, we developed two bespoke flexible metrics, on the basis of the corpus of cataloguing records. In particular, existing citation indexes for representing scholars' authoritativeness are selected and tuned so as to measure the likelihood of scholars to be reliable sources of information. The resulting metrics are the *artist-related index* and the *acceptance-rating*.

The artist-related index is inspired by the h-index metric [38]. The H-index is a metric that uses the number of an author's publications along with the number of times those publications have been cited by other authors in an attempt to gauge an author's perceived academic authority in a given field of research. In order to apply a similar metric on art historians, the following two parameters are taken into account:

- The number of artists to whom the scholar ascribed some artworks. This is currently limited to the number of artists retrieved in the three data sources, whose artworks were ascribed by the scholar at least once (discarded attributions that cite the scholar are not considered).
- The number of artworks that the scholar ascribed to a certain artist correspond to the number of the scholar's citations. The number includes all the scholar's accepted attributions retrieved in the three photo archives.

For instance, in the course of his activities Mary Berenson ascribed artworks to 8 artists. For each of these artists he has been cited as favourite source of attribution respectively 10, 9, 9, 8, 8, 3, 2, 1 times. In details, he has been cited 10 times for having ascribed 10 different artworks to the first artist, 9 times for 9 different artworks to the second artist, and so on. His artist-related index is 5, because he has been cited at least five times with regard to 5 artists. Limits of the metric are evident. Connoisseurs that work on a narrow group of artists, or artists that were not particularly productive, are somehow penalised.

The acceptance-rating is a scoped measure that uses the number of a scholar's accepted attributions with regard to a certain artist, along with the total number of possible attributions for that artist (i.e. the total number of artworks surveyed in the three photo archives). Precisely, given a list of tuples `(scholar, artist)` the rating is calculated for each tuple as the proportion between the number of scholar's citations for that artist over the three data sources (`numberOfCitations`) and the number of artworks that are ascribed to the latter in the three sources (`totalNumberOfArtworks`).

For instance, Bernard Berenson has been cited 10 times with regard to Titian's artworks (i.e. 10 of his attributions were accepted by data providers). The three data sources include in total 20 Titian's artworks. Therefore, the acceptance-rating of Bernard Berenson with regard to Titian is 50%.

6.4. The ranking model

The ranking model elaborates a number of steps and incrementally associates partial scores to attributions. A final score is associated to both accepted and discarded attributions. Different units of measure apply to the definition of partial scores, hence these lie on different ranges of values. Table 2 summarises the aforementioned dimensions, related scores, and ranges.

Relevance is addressed by the *agreement score (g)*, that counts the number of providers in agreement with the attribution at hand minus the selected source. For instance, having six data providers, the range of the agreement score is between 0 (no other sources in agreement) and 5 (all the sources agree with the attribution at hand).

The domain *expert score (a)* is a boolean measure that is 1 when attributed to domain experts and 0 when attributed to non-experts. The score is intentionally low so as to not penalise less scholarly sources, such as DBpedia, Wikidata, and VIAF. Indeed, the latter contribute to highlight the broad acceptance of an attribution, despite these cannot be be deemed authoritative by themselves.

Reliability is measured by relying on the rating of types of sources and criteria that motivate the attribution. According to domain experts' opinions, the *criteria score (b)* is the one that mostly affects the ranking of results, hence it must weight significantly more than others. The score is cumulative, meaning that if several sources support an attribution, the score is given by the sum of the ratings of all the types of sources supporting the attribution at hand.

Finally, timeliness is measured by the *date score (f)*, obtained by comparing the dates of retrieved attributions. The score is normalised between 1 and 0 so as to balance the rating of criteria with a lower rating, e.g. the most recent scholar's attribution should weight as much as the archival classification.

7. mAuth. A Semantic Crawler of Artwork Attributions

As aforementioned, a proof-of-concept web application called *mAuth - Mining Authoritativeness in Art History* is developed to evaluate the soundness of the ranking model. Specifically, mAuth relies on a focused crawling process to guide the crawler through relevant information, discarding immediately irrelevant resources, and saving time. Focused crawlers [46] are also called preferential or heuristic-based crawlers. The heuristic we use in the proposed solution is based on ontology mapping. All the harvested resources are semantically annotated, served as RDF data, and represented according to one or more vocabularies that are mapped to a crawling schema. Fetched data are stored in a central triplestore. The three main objectives of mAuth are:

- Discover relevant attributions with regard to an input artwork.
- Analyse and rank observed attributions on the basis of a framework of IQ measures.
- Enable users and applications to perform trust-based attribution retrieval in the web of data.

The URI identifying an artwork minted by a certain data source is the starting point of the crawling process. We restrict the number of sources to be fetched by relying on a list of trusted providers. This allows us to have a high accuracy of retrieved information,

IQ Group	IQ Measure	Score	Range
Subject criteria	Relevance	agreement (g)	[between 0 and (n-1)] where n is the total number of retrieved information sources minus the one in scope
	Reputation	domain expert (a)	[0 or 1] boolean
Object criteria	Reliability	criteria (b)	$\sum_{i=1}^{n} x$ where x is the rating associated to a criterion and n is the number of criteria recorded for the attribution at hand
	Timeliness	date (f)	[between 0 and 1]

Table 2
IQ dimensions, scores and ranges

and reduces time-consuming and error-prone activities related to the customisation of the crawler for many non-relevant websites. Precisely, six data sources are used for the sake of the evaluation of the framework, namely: the Zeri photo Archive, Villa I Tatti Berenson Library, the Frick Art Reference Library, DBpedia (the Italian, English and French datasets), Wikidata, and VIAF.

The ranking model highlights the most documented and well-researched information sources, that are returned to users accompanied by context information and, eventually, scholars' metrics. The list of ranked attributions is served to applications by means of an API, and to users by means of a Web application - that shares the same logic of the API.

In terms of requirements, the crawler responds to the following tasks:

- The crawler is started from a command-prompt, with a number of given components, namely: (1) a linkset of URIs identifying artworks, (2) a settings file including instructions on how to access data sources (whether by content negotiation, SPARQL endpoint, or Linked Data Fragments server), (3) a list of trusted providers to be harvested, and (4) an ontology mapping document, including triple patterns for query rewriting.
- The crawler queries a linkset including a collection of URIs identifying artworks, it parses the URI bases, and looks for matches in the list of trusted providers.
- The crawler looks into a settings file providing instructions on how to access the data sources.
- The crawler looks into a mapping document to collect triple patterns to be parsed, rewrites a query to be performed against some endpoint, and returns results annotated according to the crawling schema.
- The crawler stores retrieved triples in a dedicated named graph in the local triplestore.

Data stored in the triplestore are queried and analysed by a set of functions that return a sorted list of results, which are grouped by data provider and artist, and are associated to a score. Such algorithms respond to the following requirements:

- assign the domain expert score to the attribution;
- sort attributions by date, calculate their timeliness, and associate the date score;
- query the controlled vocabulary of terms and related rating to associate the criterion score;
- if an attribution cites a scholar, perform a statistical analysis to return scholar indexes;
- query a linkset including equivalence statements on artists, group attributions by artist, and calculate the agreement score, and
- sum all the scores and associate a final score to each attribution.

Data can be consumed in two ways according to the nature of the request. An API accepts as input the persistent URI identifying an artwork and returns the ranked list of results as a JSON object. A Web application provides a web interface for querying data and accepts as input the URL of the cataloguing record or web page describing an artwork at hand.

7.1. Components of the framework

mAuth is made out of three components, namely: (1) a crawler for discovering and mining relevant data sources describing artworks, (2) a stack of technologies for analysing and ranking data sources, (3) bespoke software solutions - an API and a web app - to serve ranked data according to the request.

Figure 2 shows how components of the framework interact with each other. The framework consists of the following components:

- Settings file. A JSON file including (1) the URI base of resources belonging to a domain, (2) the data access strategy (content negotiation, SPARQL endpoint, or Linked Data Fragments),

and (3) the access point (the URI of the endpoint, the rewriting rule to fetch RDF documents by content negotiation).
- List of trusted providers. mAuth relies on a list of six trusted data sources to be fetched and analysed. The list is used by the domain filter to prune the URI stack from not relevant domains in the mining process.
- Image similarity index. Pastec[19] search engine is used to compare images belonging to the trusted data providers and create the index of matching images. If the similarity score is greater than 30.0, the URI of the artwork identified by the image is included in the URI stack, and the link between similar artworks is included in the linkset of artworks.
- Equivalence lookup service. The lookup service retrieves equivalences for the URIs identifying (i) artworks, (ii) artists, (iii) organisations and scholars retrieved in the six trusted data providers. The lookup service queries the linksets and uses instructions detailed in the settings file to access data sources of all the URIs.
- URI stack. The URI stack is the initial collection of URIs belonging to six trusted data providers that identify artworks, artists, organisations and scholars. These are stored in mAuth triplestore in dedicated named graphs. Each URI identifying an artwork is sent in a *first in first out* order to the Domain Filter for further processing.
- Domain filter. The domain filter checks whether URIs included in the URI stack belong to a domain in scope or not, so as to restrict the scope of the crawler to the six specified domains.
- Mapping rules. The mapping document is a JSON file including triple patterns useful to retrieve: (1) artist, (2) title of the artwork, (3) criteria motivating the attribution, (4) date of the attribution, (5) sources of information, (6) cited scholar and institutions, and (7) images. Triple patterns are used by the Data miner for rewriting SPARQL queries.
- Data miner. The data miner is the core component of the crawler, which integrates all the previous components. It accepts in input (1) a URI identifying an artwork taken from the URI stack, (2) a settings file, and (3) a document of mapping rules. The data mining algorithm is iterated over all the URIs in the URI stack. Data are accessed by means of rules specified in the setting file. For each property path listed in the mapping document the miner rewrites a SPARQL query to be performed against the data source, by using the method specified in the settings file. Data fetched are stored in the Observation graph, further analysed by the Data analyser. Results stored in the Observation graph are represented according to a Crawling schema.
- Crawling schema. The crawling schema is based on the Observation ontology pattern[20] and the PROV Ontology.
- Observation graph. The Observation graph represents snapshots of fetched attributions. Data are queried and further elaborated by the Data analyser. Storing data rather than querying data sources on-the-fly every time a user asks for a URI allow us (1) to speed up the query phase and (2) to preserve changes in attributions over time (i.e. versioning).

The following listing in turtle syntax shows an example of data included in the Observation graph organised according to the Crawling schema. The example describes the observation of an attribution fetched from the cataloguing record n. 39459 of the Zeri photo archive. The record describes an attribution made by Everett Fahy of “San Pietro Martire in preghiera e sante" to Francesco Granacci, in 1990. Links to photographs depicting the artwork are also fetched. See listing 3

Listing 1: Example of data stored in the Observation graph

```
mauth:39459-artist-granacci-francesco-obs
  rdfs:label "Zeri Foundation (University of
    Bologna) accepted attribution";
  mauth:hasType mauth:accepted;
  mauth:hasObservedArtist zeri:granacci-
    francesco;
  mauth:hasObservedArtwork zeri-artwork
    :39459;
  mauth:hasObservedCriterion criteria:scholar
    -attribution;
  mauth:agreesWith zeri:e-fahy;
  mauth:hasAttributionDate "1990-01-01T00
    :00:00.001Z"^^xsd:dateTime;
  mauth:hasSourceOfAttribution <http://w3id.
    org/zericatalog/artwork/39459>;
  mauth:image <http://catalogo.fondazionezeri
    .unibo.it/foto/120000/82800/82571.jpg> ,
```

[19] http://pastec.io/

[20] http://www.ontologydesignpatterns.org/cp/owl/observation.owl

```
    <http://catalogo.fondazionezeri.unibo.it/
    foto/120000/82800/82572.jpg> ,
     <http://catalogo.fondazionezeri.unibo.it
    /foto/120000/82800/82573.jpg>;
  prov:atTime "2018-07-22T22:35:48.767Z"^^xsd
    :dateTime;
  prov:wasAttributedTo mauth:md;
```

- Data Analyser. The Data analyser consists of a number of scripts that query the Observation graph, sort results according to the ranking model, calculate scholars' citation indexes, and send the final list of attributions to the API and the web application. In order to associate attributions with a score it performs four operations, namely: (1) checks data provenance against the List of trusted providers and attributes a *domain expert score* to the attribution; (2) extracts the rating associated to a type of source or criterion from the Controlled vocabulary, and computes the *criterion score*; (3) sort attributions by date (when available), calculates the timeliness of the attribution, and associates the *date score*; (4) groups attributions by artist, and calculates the *agreement score*.
- Ontology-based ranking model. The ontology-based ranking model provides ranges of scores. The Data analyser weights retrieved information accordingly. The ranking model takes as input a number of property values, namely: the name of the data provider, the label of criteria, the position of the attribution date in the list of sorted attributions, the number of agreements.
- Controlled vocabulary. The Controlled vocabulary of rated criteria is a named graph stored in the mAuth triplestore describing the twenty-two types of sources and criteria that can motivate an attribution. For instance, the criterion "documentation" is described as follows (in turtle syntax).

Listing 2: Example of type of source and related rating in the controlled vocabulary

```
:documentation
  a hico:InterpretationCriterion;
  rdfs:label "documentation" ;
  dbo:rating "10.0"^^xsd:float .
```

- Statistics graph. The Statistics graph is the result of the analysis performed over the Observation graph in order to extract information on scholars' indexes. For instance:

Listing 3: Example of indexes stored in the Statistics graph

```
:bernard-berenson-franciabigio
  mauth:hasArtistIndex "0.78"^^xsd:float;
  mauth:hasIndexedArtist :franciabigio;
  mauth:hasIndexedHistorian :bernard-
    berenson.

:bernard-berenson
  mauth:hasHIndex "32"^^xsd:float.
```

The Data analyser queries the Statistics graph to retrieve indexes to be associated to the attributions, and includes them in the list of results to be sent to the API/app.

- mAuth API. The mAuth API provides functionalities that go beyond data access. Indeed, it is a means for relationship discovery and data integration. It is accessible through HTTP calls, and accepts in input the persistent URI identifying an artwork included in one of the six aforementioned providers. It reuses the described components so as to look into the Observation graph for a match, retrieves attributions and indexes, ranks and sorts results, and serve the list of results as a JSON file[21].
- mAuth Web application. The mAuth web application shares the same logic of the mAuth API, i.e. all the aforementioned components and auxiliary files, and serves ranked data to users that look for the history of attributions related to a single artwork. It works as an aggregator of results, and it is used for the user-centered evaluation of the IQ framework.

mAuth comes as a toolkit for art historians that want to retrieve the history of attributions of pieces of art related to the Modern Era. It offers services to access and compare data (the web app), and for integrating data in other applications (the API). The toolkit includes a number of flexible python modules that can be reused and extended in different contexts, and all the resources realised for the evaluation of the framework,[22] namely:

- Three linksets including results of data reconciliation on artists, artworks and historians related to

[21] For instance, http://163.172.177.79:8000/full/http://www.wikidata.org/entity/Q727875 returns all the attributions related to the Wikidata entity Q727875, i.e. "Venus of Urbin".

[22] The source code of mAuth and all related resources are stored in https://github.com/marilenadaquino/mauth

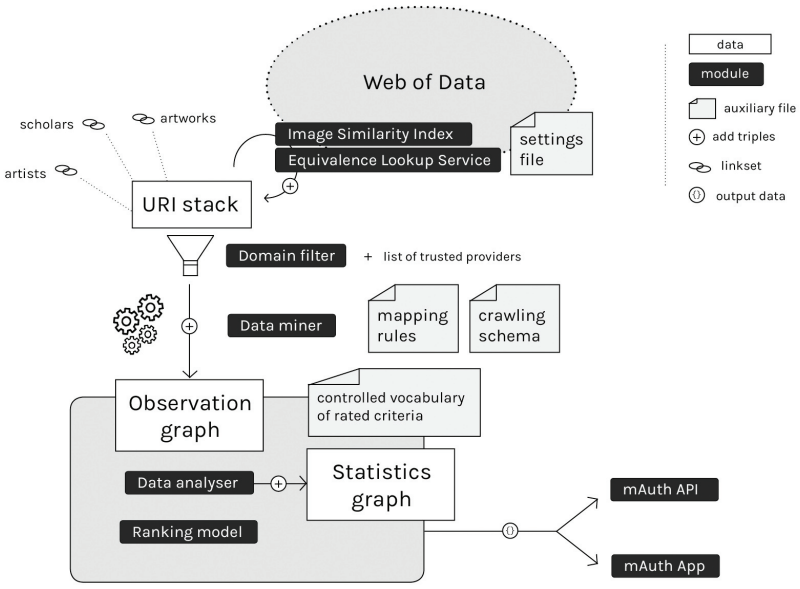

Fig. 2. Components of the mAuth framework

paintings of the Modern Art. The linkset of artists includes around 37.386 equivalences links between 12.227 individuals. The linkset of artworks includes 7.284 links between 2.474 artworks described by the six aforementioned providers. The linkset of historians includes 33.676 links between 11.996 individuals.

- A dataset of multiple attributions (the Observation graph) related to 1.269 unique individuals.
- A dataset on citation indexes (the Statistics graph) related to 11.996 scholars.
- An ontology mapping document including triple patterns for retrieving attributions in six data sources.
- A settings file with instructions on how to access a number of art historical data sources.
- An instance of Blazegraph triplestore storing the aforementioned graphs.

8. User-centered Evaluation

The evaluation aims at demonstrating that Linked data applications can efficiently support trust-based information retrieval tasks in Arts digital libraries (H1), and that the developed approach can effectively support users' decision-making process (H2).

In order to demonstrate H1, we evaluate the following subhypotheses: (H1.1) users can find relevant information faster when using mAuth than other online catalogues and web portals, and (H1.2) users can find relevant information by accessing a less number of pages when they use mAuth rather than pharosresearch and online catalogues and web portals. The two hypotheses are based on the idea that Semantic Web technologies positively impact users' expectations when performing common tasks such as collecting relevant information on a subject at hand. Therefore, our evaluation is focused on the task. We do not evaluate the efficiency of the crawler itself or human computer interaction aspects such as user experience, which are out of scope and left to future work.

We set up a task-based evaluation where we compare the usage of mAuth with other systems and we measure quantitative and qualitative aspects to validate our hypotheses. Since there are no similar tools available, we compare mAuth to (1) an aggregator of images and metadata of artworks – images.pharosresearch[23] – and (2) three online sources of attributions, namely: the Zeri photo archive online catalogue,[24], Wikidata,[25] and Wikipedia.[26].

[23] http://images.pharosartresearch.org
[24] http://catalogo.fondazionezeri.unibo.it
[25] http://www.wikidata.org
[26] http://www.wikipedia.org

To demonstrate (H2), we assess the soundness of the framework of IQ metrics and the ontology-based ranking model. So doing we aim at demonstrating that Semantic Web technologies can respond to users' sophisticated information needs, such as supporting the validation of questionable information. To assess it, we demonstrate the following subhypotheses: (H2.1) users' satisfaction when validating internal grounds of attributions is higher when using mAuth rather than pharosresearch and online catalogues and web portals, (H2.2) users' satisfaction when validating the ranked list of attributions in mAuth is high. Scholars' indexes are presented to users but these do not affect the ranking model. Therefore we evaluate the satisfaction with regard to implemented content-based trust mechanisms, while we do not fully evaluate reputation-based trust mechanisms. Lastly, we evaluate the fitness for purpose of the HiCO ontology when representing features characterising argumentations around attributions (H2.3).

8.1. The user study

Description of research scenarios. Three research tasks are assigned to users, that have to perform researches on mAuth and other systems, and fill in an evaluation form. Tasks are designed to evaluate the aforementioned hypotheses, and represent common research scenarios in the Arts and Humanities namely:

- Gather information on a well-known artwork avoiding time-consuming researches.
- Gather information on a less-known debated artwork whose attributions are not sufficiently documented.
- Gather information on a debated artwork whose attributions are well-documented.

In the first scenario, named "Retrieve attributions and assess their acceptance", a user is required to complete the same task in the five different systems in order to gather enough information on artwork attributions. Users are asked to search for a given artwork, browse related web pages, and gather information on recorded attributions. The chosen artwork is the well-known painting *La Schiavona*, currently ascribed to Titian, although it was formerly attributed to Giorgione. Users are introduced to the artwork, by showing a picture of it at the beginning of the evaluation test.

In the second scenario, named "Choose and motivate the most reliable attribution" the focus is on mAuth only. Users are redirected to the results of a research, including attributions related to a less known artwork. The scenario here presented is the most complex one. Two contradictory attributions are provided by the Zeri photo archive and Villa I Tatti, and both the domain experts rely on scholars' opinions to support their statements. Users are asked to evaluate metadata associated to attributions and to comment on the goodness of attributions. In this case the artist-related index and the acceptance rating of the two scholars are presented to the user but these do not affect the ranking. Shown indexes contradict the actual ranking, i.e. higher indexes are associated to the scholar supporting a less scored attribution.

In the third scenario the user finds attributions related to the painting *Three Graces* in mAuth. Attributions are provided by three trusted domain experts, namely: the Zeri photo archive, Villa I Tatti, and the Frick Art Reference Library. A discarded attribution is shown too, for a total of four attributions.

Evaluation measures. In every scenario, a number of questions allowed us to perform a quantitative and qualitative analysis on users' satisfaction. In particular, we measured four variables in the first scenario (two quantitative variables and two qualitative variables), and two (qualitative) variables in the second and third scenario. Variables are described as follows.

CT. Completion time. The completion time is the time span between the moment a user begins a task and the moment when the retrieval task is accomplished. The CT metric is widely used to measure users' satisfaction with regard to the performance of the retrieval process [47].

TPV. Total pages visited. The TPV metric measures the number of pages visited by a user in order to get the information seeked. It is measured for the first retrieval task performed with the given systems [48]. The TPV metric measures the efficiency of the crawling system and the user satisfaction with regard to the retrieval information system.

Such quantitative measures apply to the first described scenario, i.e. retrieval of information sources related to a well-known artwork. However, these might not be sufficient to evaluate users' satisfaction, since the users' perception may vary according to their experience and background, and the difficulty of the task at hand [49].

Two qualitative assessments, described below, aim at filling the gap related to the evaluation of user's satisfaction. Users are asked to provide a subjective feedback on their experience with the three systems, and secondly on the ranking of results.

Scenario	Online Catalogues	images.pharosresearch	mAuth
1	CT, TPV, US	CT, TPV, US	CT, TPV, US, RSS, PAS
2			US, RSS, PAS
3			US, RSS, PAS

Table 3

Metrics used in the user-center evaluation grouped by scenario

US. User satisfaction wrt the information retrieval process. The User Satisfaction of Information Retrieval Results (US) measure quantifies the user's satisfaction with regard to the results of the information retrieval. Specifically, it measures whether retrieved information are useful and sufficient to assess the goodness of an attribution. Participants provide the measure by using a Likert scale from 1 to 5 (Strongly disagree to Strongly agree).

USR. User satisfaction wrt the ranking of results. The User Satisfaction of Ranking measure (USR) allows to quantify the user's satisfaction with respect to the ranking model and the suggested attribution. In particular, two scores contribute to define the USR measure, namely:

- *RSS. Rank Satisfaction Score.* The RSS measure provides a feedback on the user's satisfaction with respect to the order of presented results and the score associated to each information source.
- *PAS. Perception of Authoritativeness Score.* The PAS measure provides a feedback on the user's acceptance of the suggested attribution, i.e. the attribution scored more than the others. It is based on the Net Promoter Score [50] for measuring the likeliness of a user to prefer, and eventually suggest and cite, a certain attribution over the others available.

Like the US measure, participants provide the RSS and PAS measures by using a Likert scale from 1 to 5 (Strongly disagree to Strongly agree).

Table 3 summarises the usage of metrics in the three scenarios. As aforementioned, the two quantitative metrics (CT and TPV) apply to the first scenario only, so as to compare the user satisfaction with respect to the three evaluated systems. The two qualitative metrics (US and USR) apply to all of the three scenarios. It is worth to notice that US applies to the three systems in the first scenario, and to mAuth only in the second and third scenarios. The USR measures apply to the evaluation of mAuth only in the three scenarios, since the other systems do not rank results.

Background	N.	Affiliation
Art historian	10	Warburg Institute, Max Planck Inst. for Art History, Frick Art Reference Library, Getty Research Institute, University of Padua, University of Bologna, Italian Public Education System, University of Rome
Collection manager	5	Getty Research Institute, Yale Center for British Art, Italian Ministry of Cultural Heritage and Activities (MiBACT), Paul Mellon Centre for Studies in British Art, Not specified
Photo archivist	5	Federico Zeri Foundation, Kunsthistorisches Institut in Florenz, Bibliotheca Hertziana - Max-Planc Institut, Italian Ministry of Cultural Heritage and Activities (MiBACT), University of Trieste
DH scholar	2	University of Bologna, University of Lausanne
Computer Scientist	4	University of Bologna, Vrije Universiteit Amsterdam, Knowledge Media Institute - Open University
Other	4	University of Milan, University of Florence, University of Bologna

Table 4

Population of the User study

Fleiss Kappa. For US, RSS, and PAS measures we calculated the inter-raters agreement by means of the Fleiss Kappa measure [51].

Lastly, we collected feedbacks on users' preferences for improving the ranking model.

Data collection and population of the study. The data collection was conducted by using a survey online application, i.e. Google Form.[27] Users filled in the form remotely and submitted their answers to be analysed. Data collected from the survey are published online [52].

We collected feedbacks from 31 users. The background of participants is the key element of the evaluation. Users mainly belong to some of the most im-

[27] See the form at https://goo.gl/forms/xDLwvCCaEFWm4D5h2

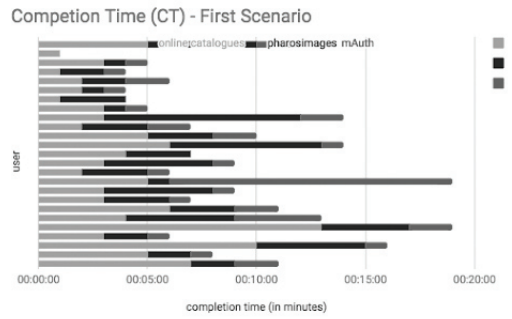

Fig. 3. Completion time required to complete the first task in online catalogues, pharosresearch, and mAuth

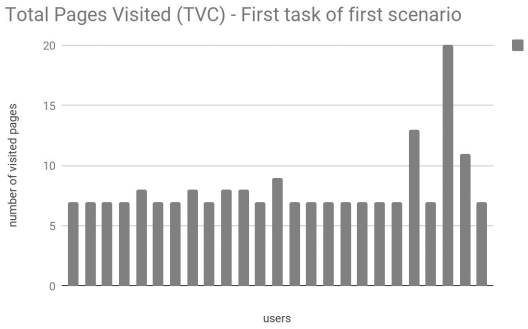

Fig. 4. Total number of pages visited to complete the first task in online catalogues, pharosresearch, and mAuth

portant cultural institutions dealing with art historical data. Other stakeholders in the Humanities and Computer Science were involved to get feedback from different points of view. Domain experts are expected to evaluate the goodness of ranked attributions, while non-domain experts are expected to provide feedback on the soundness of the conceptual framework as applied to any kind of pieces of information, and show whether there are similarities between the art historical research approach and other fields. Table 4 shows users grouped by background and affiliation.

8.2. Results of the user-centered evaluation

Results of the user-centered evaluation are grouped by metric and scenario. Results of the survey are available in [53].

8.2.1. Using Semantic Web technologies in trust-based retrieval tasks (H1)

Completion Time (CT). Figure 3 shows the time required to participants to perform the task introduced in the first scenario. The CT measure is calculated for the three systems in scope, namely: three online catalogues, images.pharosresearch and mAuth. Results are grouped by user and then by system.

The average time is calculated on the basis of the CT measure, and is respectively: 04:05 minutes for searching with the three online catalogues, 03:12 minutes for searching with pharosresearch, and 01:50 minutes for searching with mAuth. It is worth to notice that some users had difficulties when using pharosimages, and some were not able to find the artwork at hand. Results show that the retrieval of the same number of information sources in mAuth requires 55% less time than a research in online catalogues, and 42% less time than a more sophisticated research in pharosresearch.

Total Pages Visited (TPV). Figure 4 shows the total number of pages visited by users to complete the first task in the first scenario in Wikidata, Wikipedia, and Zeri online catalogue.

Users were expected to open at least 7 pages in order to get all the web pages. Some users were not able to reach all the requested web pages and they answered with a lower number of pages corresponding to the number of pages visited to reach one or two web pages out of the three requested (between 2 and 7). We normalised errors to 7, so as to get significant results for the TVC measure. On average, a user visited 8.16 pages.

In images.pharosresearch users were asked to input the URL of an image previously retrieved in one of the three catalogues, and use the similarity match tool to retrieve cataloguing records. It is worth to notice that 17 participants out of 31 participants were not able to find results because no matches were found. In order to extract significant results, we estimate on average 5 pages are supposed to be visited to first retrieve the Zeri cataloguing record, which requires to visit 3 pages, and 2 pages to retrieve related results in pharosresearch.

Likewise, users were asked to input in mAuth the URL of one of web pages retrieved in one of the three catalogues and they all got results. They visited in average 5 pages to get to the final list of results. In summary, when using mAuth, a user is required to visit the same number of pages as in the images aggregator pharosresearch, i.e. 5 pages in the best case scenario, and to visit 38,7% less pages than in multiple online catalogues.

8.2.2. Users' satisfaction with regard to the ontology-based ranking model (H2)

In Figure 5 are illustrated the US, RSS, and PAS measures for each scenario.

User satisfaction wrt the information retrieval process (US). The US is high in the first and third scenario (84% of user either agree or strongly agree), since in the first scenario the artwork is unanimously ascribed to the same artist, and in the third scenario plenty of evidences supporting an attribution rather than others are available. In the second scenario the US is significantly lower (58%) since attributions are less documented, there are only two sources supported by scholars' opinions only, and there is no agreement.

Rank Satisfaction Score (RSS). When evaluating RSS, we see that in the first scenario 74% participants either agree or strongly agree; in the second scenario only 38,7% either agree or strongly agree, while 35,5% neither agree nor disagree, and 25,8% disagree; in the third scenario 81% either agree or strongly agree.

Perception of Authoritativeness Score (PAS). When evaluating PAS, in the first scenario we see that 84% either agree or strongly agree; only 42% either agree or strongly agree in the second scenario, while 51,6% neither agree nor disagree; 71% either agree or strongly agree in the third scenario.

Moreover, we collected feedbacks on users' preferences for improving the ranking model, including insights on their perception of the usefulness of scholars' indexes. Users were asked to (1) select from a list the dimensions they deem relevant for ranking attributions according to the selected scenario, and (2) to provide a feedback on how scholars' authoritativeness scores would affect the ranking - if taken into account. Precisely, in the second and third scenarios users were not told that the citation indexes do not affect the ranking, but most of them believed they were actually affecting it or that they should have affected it more. Such a social experiment provides useful insights on how to tune the current ranking model and enable future work on the analysis of scholars' indexes in the Arts and Humanities.

Finally, the Fleiss Kappa measure is calculated for the 31 raters that evaluated the three cases according to the five categories of the Likert scale: kappa is 33% when evaluating the US measure, 34% for the RSS measure, and 36% for the PAS measure, indicating a fair agreement between raters [54].

8.2.3. Fitness of HiCO Ontology

The web application serves users attributions and shows values annotated by using terms of HiCO. Specifically, HiCO terms were used by three out of six trusted providers and were further elaborated by the ranking model. HiCO terms represent features characterising argumentations around attributions. Such terms and predicates address the following features:

- The date of the statement.
- The criterion or type of source adopted to motivate the statement.
- The primary source of information, e.g. a scholar, a museum, an auction firm.
- The secondary source recording the statement, e.g. a cataloguing record.
- The agreement or disagreement with other statements.

Users' satisfaction when using mAuth and users' feedback on ranked results show that data retrieved are sufficient to assess the veracity of statements in two out of three cases (i.e. 84% in the first and third scenarios, 56% in the second scenario), and that the consequent ranking is deemed useful (respectively 72%, 40% and 84% in the three scenarios).

9. Discussion

In this section we discuss benefits and limits of using Semantic Web technologies in art historical research activities as shown by results of the user-centered evaluation, and we present our lessons learnt.

Using Semantic Web technologies in trust-based retrieval tasks. Results of CT measure allow us to validate the initial hypothesis *H1.1 Users can find relevant information faster when using mAuth than other online catalogues and web portals*. Results of TPV measure confirm the initial hypothesis *H1.2 users can find relevant information by accessing a less number of pages when they use mAuth rather than pharosresearch and online catalogues and web portals*.

Positive feedbacks collected in the first scenario reveal that Semantic Web technologies can effectively support scholars' tasks, such as gathering information, analyse internal grounds of information sources, compare sources, and efficiently support the decision-making process.

The CT measure shows benefits derived from the usage of specialised aggregators for retrieving arguments around attributions that are spread in several

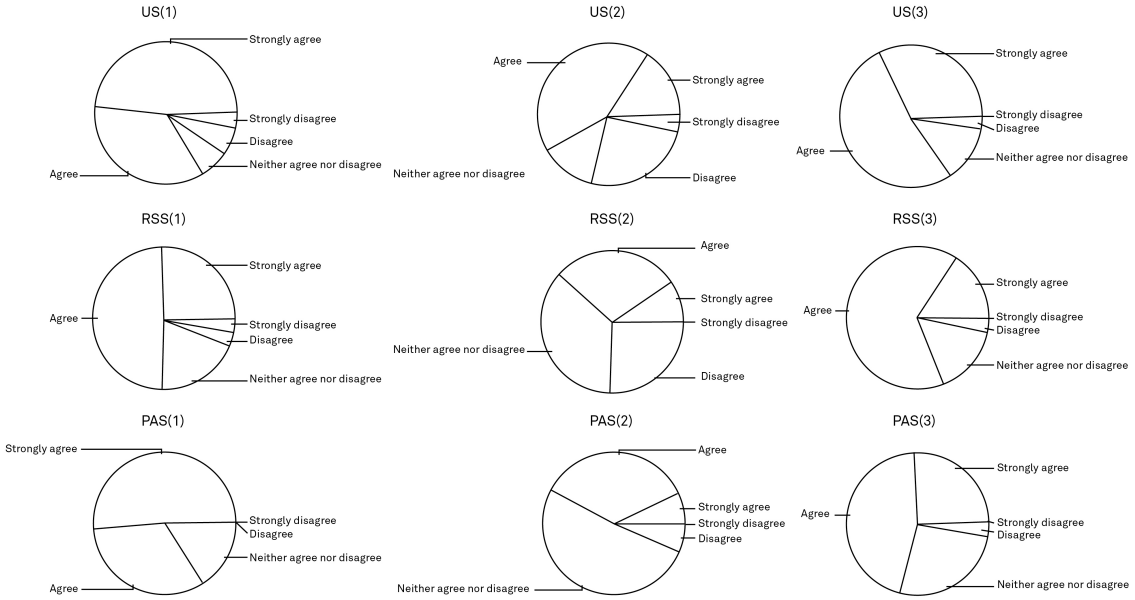

Fig. 5. US, RSS, and PAS measures in three evaluated scenarios

sources. Along with results on the user's satisfaction for the ontology-based ranking model, this measure demonstrates that Semantic Web technologies can significantly reduce time-consuming activities.

However, Semantic Web technologies still present some limitations. Querying remote triplestores may be time-consuming, since these can affected by timeouts, or other limits of third-party softwares. For this reason, when an alternative Linked Data Fragments (LDF) [55] client was available, the latter was preferred over SPARQL endpoints and content negotiation mechanisms. Specifically, VIAF[28] and DBpedia[29] offer an alternative LDF service for querying. Despite such a solution halved the query time, the total time required to query heterogeneous sources was still high. Assuming that art historical data do not change significantly over time, and fetching data is a task that can be executed on a given time interval, rather than being executed on-the-fly every time a user inputs a new URI, we preferred the storage of fetched data, i.e. the Observation graph, in a dedicated triplestore[30] to guarantee high-performance. So doing, we also provide versioning of art historical data, enabling future data comparisons for historical purposes.

The TPV measure shows that retrieving the complete list of available sources on a topic at hand is particularly error-prone, due to the lack of users' confidence when using different platforms. Hence, specialised aggregators become fundamental tools for facilitating the research.

However, on the one hand plenty of interesting data sources are currently not available as Linked Open Data, which affects the aggregation of arguments. Sec-

[28] http://data.linkeddatafragments.org/viaf
[29] http://fragments.dbpedia.org/2016-04/en
[30] https://www.blazegraph.com/

ondly, the majority of museums and cultural heritage institutions do not release machine-readable data on argumentations, because it would require a significant effort in terms of data cleaning and reconciliation. On the other hand, when detailed data on artwork attributions exist, such as the ones issued by competitors in the art market, these are not released as open data due to business interests[31]. Therefore, there is a urge of solutions based on cultural heritage open data for fostering art historical research. The realisation of linksets for reconciling artworks, artists, scholars, and organisations, currently leveraged in mAuth aims at facilitating future publication and reconciliation of art historical data.

Users' satisfaction with regard to the ontology-based ranking model. Results of US, RSS, PAS, and Fleiss Kappa measurements allow us to validate hypotheses *H2.1 users' satisfaction when validating internal grounds of attributions is higher when using mAuth rather than pharosresearch and online catalogues and web portals*, and *H2.2 users' satisfaction when validating the ranked list of attributions in mAuth is high*.

Feedback shows that content-based trust mechanisms are sufficient in the following scenarios, namely: (1) when there is an agreement between all the sources (first scenario), (2) when there is a disagreement but one source is more documented than others (third scenario), (3) when reputation-based trust mechanisms (i.e. citation indexes) confirm the rating based on contents (third scenario).

Limits of our approach are highlighted by results in the second scenario, namely: (1) when sources are not well-documented, (2) when all the sources rely on reputation-based trust mechanisms (i.e. scholars' opinions), and (3) reputation-based trust mechanisms contradict the ranking based on content-based trust mechanisms. Limits derived from not well-documented sources generally depend on data quality issues. Data quality issues affect data providers' reliability, and have a negative impact on research. Potentially, the framework of IQ measures here designed can contribute to define a number of policies for data quality improvement in art historical data sources. Specifically, data integration methods can leverage findings of this research so as to automatically enrich and update cataloguing records with cited evidences, sources, alternative attributions and, potentially, can correct attributions that have been classified with a low score. This could be especially beneficial for those organisations, such as museums or galleries, which are generally deemed reliable sources of information but that do not provide machine-readable data on argumentations around attributions.

Limits derived from sources relying on scholars' opinions depend on the lack of reliable reputation-based trust mechanisms in the Humanities. Comprehensive citation indexes for scholars in the Humanities are not available due to a number of challenges, such as the fact that not all the Humanities journals are indexed, not all scholars' publications are available in a digital format, and humanists are often acknowledged by means of different types of citations than the traditional ones (e.g. verbal communication, annotations on photographs). Therefore, existing reputation-based trust mechanisms are not sufficient to address the Arts and Humanities needs. Measures here presented for defining two different perceptions of scholars' authoritativeness (i.e. the artist-related index and the acceptance rating), despite in early stages, allow a more flexible and fit-for-purpose aid to users' judgement.

Lastly, limits derived from competing (and contradicting) reputation/content-based trust mechanisms depend on the lack of agreement in shared ranking mechanisms. However, the positive feedback provided by users with different backgrounds on the soundness of the IQ framework shows its potential application and portability to near fields. By examining arguments around other types of information (e.g. authorship attributions in literature, artwork provenance attributions) it would be possible to explore other patterns in users' judgement and define common behaviours so as to tune the ranking model accordingly.

Fitness for purpose of HiCO Ontology. Users' satisfaction is high in two out of three common scenarios. In the second scenario we recorded a lower positive feedback, due to (1) the lack of sufficient data for validating the goodness of retrieved attributions, and to (2) the need of bespoke measures to enable reputation-based trust mechanisms. However, such aspects are not in the scope of the ontology.

To this extent, the HiCO ontology can be deemed a valid means for art historical research activities since it offers the conceptual and terminological basis for (1) ontology-based data integration purposes, and for (2) defining content-based trust mechanisms. Specifically, terms of the vocabulary provide a comprehensive check list for data quality assessment, that would allow to evaluate: (1) information completeness, since

[31] See for instance Artory, https://www.artory.com/

it lists all the features that argumentations around attributions should carry, and (2) information reliability, since it provides a controlled vocabulary of terms identifying criteria and types of sources accompanied by a rating of these.

Portability of our approach to near fields In conclusion, existing Semantic web technologies provided suitable solutions for supporting the decision-making process when validating the goodness of argumentations around attributions. Moreover, the Semantic Web offered methods, models, and techniques that facilitated the research on the reproducibility of cataloguers and users' decision-making process.

Indeed, the approach here proposed for reproducing cataloguers and users' trust mechanisms, entirely relies on the application of Semantic Web technologies for analyses and classification purposes, and offers solutions that can be applied to near fields in the Humanities (e.g. philology, history). Specifically, we (1) tuned existing domain-independent metrics for evaluating reputation-based trust mechanisms to reflect peculiarities of citation mechanisms in the Humanities, (2) identified and selected existing domain-independent measures for addressing context-related variables that affect users' judgement, and (3) we proposed a domain-independent methodology for identifying, classifying, and measuring content-related variables of online secondary sources.

10. Conclusion

In this article we presented a use case of the application of Semantic Web technologies for reasoning on argumentations around attributions.

We introduced (1) an ontology for representing argumentations, (2) a framework of IQ measures derived from the analysis of a representative dataset of art historical linked data, and (3) a proof-of-concept semantic crawler for harvesting and ranking artwork attributions.

In addition, we evaluated both the ontology and the IQ measures by means of a user study with domain experts. We rely on content-based, context-based, and reputation-based trust mechanisms so as to identify likely-to-be facts and subjective opinions. We demonstrated that automatic methods for ranking argumentations can effectively support users' judgement, and can facilitate daily tasks, such as gathering relevant sources and collecting sufficient information for validating the goodness of attributions.

On the one hand, results of the user study show that content-based trust mechanisms consistently apply to a number of common scenarios in the Arts field, namely: (1) when there is an agreement between all the sources, (2) when there is a disagreement between sources but one source is more documented than others, (3) when reputation-based trust mechanisms (i.e. citation indexes) confirm the rating based on content-based trust mechanisms. On the other hand, limitations arise in more complex cases, such as (1) when sources are not well-documented, (2) when all the sources rely on reputation-based trust mechanisms (i.e. scholars' opinions), and (3) reputation-based trust mechanisms contradict the ranking based on content-based trust mechanisms. Particularly, we acknowledge the difficulty to reach representative metrics for addressing scholars' reputation due to the lack of citation indexes and bespoke measures. In future works we aim at collecting and analyzing significant amounts of cataloguing and bibliographic data in the Arts field so as to better shape and explore historical citation networks and develop more reliable metrics to be included in the ranking model.

We demonstrated the validity of Semantic Web technologies when supporting the development of a system capable of retrieving and evaluating artworks attributions. In particular, Semantic Web ontologies effectively supported data harmonisation tasks, and facilitated data analysis tasks. The ontology-based crawling allowed the presentation of relevant results to users, and the ontology-based ranking model provided results with sufficient insights for enabling users' decisions. Limits of our approach are mainly related to data providers' legacy (software limitations, restrictive licensing, data quality issues), which we aim to overcome by developing open source resources for supporting research in Digital Art History.

Lastly, feedbacks provided by participants of the user study with a heterogeneous background show that the soundness of the IQ framework is confirmed in other fields. To this extent, we believe our approach for reproducing humanists' research methodologies and judgement can be consistently replicated for in near fields (e.g. philology, history) after knowledge acquisition, and user-tailored resources and tools can be developed to support other research and cataloguing tasks.

References

[1] R. Brilliant, How an art historian connects art objects and information (1988).

[2] G. Morelli and L.M.S. Richter, *Italian masters in German galleries*, G. Bell and Sons, 1883.

[3] C. Ginzburg, Roots of a scientific paradigm, *Theory and Society* **7**(3) (1979), 273–88.

[4] D. Freedberg, Why connoisseurship matters, *Munuscula Amicorum: Contributions on Rubens and his colleagues in honour of Hans Vlieghe* **1** (2006), 29–43.

[5] I. Rahwan, Mass argumentation and the semantic web, *Web Semantics: Science, Services and Agents on the World Wide Web* **6**(1) (2008), 29–37.

[6] M. Daquino and F. Tomasi, Historical Context Ontology (HiCO): a conceptual model for describing context information of cultural heritage objects, in: *Research Conference on Metadata and Semantics Research*, Springer, 2015, pp. 424–436.

[7] M. Daquino, F. Mambelli, S. Peroni, F. Tomasi and F. Vitali, Enhancing semantic expressivity in the cultural heritage domain: exposing the Zeri Photo Archive as Linked Open Data, *Journal on Computing and Cultural Heritage (JOCCH)* **10**(4) (2017), 21.

[8] M. Daquino, A computational analysis of art historical linked data for assessing authoritativeness of attributions, *Journal of the Association for Information Science and Technology*, 1–13. doi:10.1002/asi.24301.

[9] T. Berners-Lee and M. Fischetti, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*, DIANE Publishing Company, 2001.

[10] C. Bizer and R. Oldakowski, Using context-and content-based trust policies on the semantic web, in: *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, ACM, 2004, pp. 228–229.

[11] Y. Gil and D. Artz, Towards content trust of web resources, *Web Semantics: Science, Services and Agents on the World Wide Web* **5**(4) (2007), 227–239.

[12] D. Artz and Y. Gil, A survey of trust in computer science and the semantic web, *Web Semantics: Science, Services and Agents on the World Wide Web* **5**(2) (2007), 58–71.

[13] J. Schneider, T. Groza and A. Passant, A review of argumentation for the social semantic web, *Semantic Web* **4**(2) (2013), 159–218.

[14] L. Moreau, P. Groth, J. Cheney, T. Lebo and S. Miles, The rationale of PROV, *Web Semantics: Science, Services and Agents on the World Wide Web* **35** (2015), 235–257.

[15] J.J. Carroll, C. Bizer, P. Hayes and P. Stickler, Named graphs, provenance and trust, in: *Proceedings of the 14th international conference on World Wide Web*, ACM, 2005, pp. 613–622.

[16] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann and S. Auer, Quality assessment for linked data: A survey, *Semantic Web* **7**(1) (2016), 63–93.

[17] N.A. Rakhmawati, J. Umbrich, M. Karnstedt, A. Hasnain and M. Hausenblas, Querying over Federated SPARQL Endpoints—A State of the Art Survey, *arXiv preprint arXiv:1306.1723* (2013).

[18] R. Verborgh, M. Vander Sande, O. Hartig, J. Van Herwegen, L. De Vocht, B. De Meester, G. Haesendonck and P. Colpaert, Triple Pattern Fragments: a Low-cost Knowledge Graph Interface for the Web, *Journal of Web Semantics* **37–38** (2016), 184–206. doi:doi:10.1016/j.websem.2016.03.003. http://linkeddatafragments.org/publications/jws2016.pdf.

[19] P. Jain, P. Hitzler, A.P. Sheth, K. Verma and P.Z. Yeh, Ontology alignment for linked open data, in: *International semantic web conference*, Springer, 2010, pp. 402–417.

[20] Z. Dragisic, V. Ivanova, P. Lambrix, D. Faria, E. Jiménez-Ruiz and C. Pesquita, User validation in ontology alignment, in: *International Semantic Web Conference*, Springer, 2016, pp. 200–217.

[21] A. Dimou, M. Vander Sande, P. Colpaert, R. Verborgh, E. Mannens and R. Van de Walle, RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data, in: *Proceedings of the 7th Workshop on Linked Data on the Web*, C. Bizer, T. Heath, S. Auer and T. Berners-Lee, eds, CEUR Workshop Proceedings, Vol. 1184, 2014. ISSN 1613-0073. http://ceur-ws.org/Vol-1184/ldow2014_paper_01.pdf.

[22] D. Nadeau and S. Sekine, A survey of named entity recognition and classification, *Lingvisticae Investigationes* **30**(1) (2007), 3–26.

[23] R. Verborgh and M. De Wilde, *Using OpenRefine*, Packt Publishing Ltd, 2013.

[24] R. Isele, A. Jentzsch and C. Bizer, Efficient multidimensional blocking for link discovery without losing recall., in: *WebDB*, 2011.

[25] A.-C.N. Ngomo and S. Auer, LIMESâĂŤa time-efficient approach for large-scale link discovery on the web of data, in: *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[26] J. Volz, C. Bizer, M. Gaedke and G. Kobilarov, Discovering and maintaining links on the web of data, in: *International Semantic Web Conference*, Springer, 2009, pp. 650–665.

[27] L. Moro, M. Mancinelli and A. Negri, Il ruolo dellÃćÂĂÂŹICCD nella diffusione dei modelli descrittivi del patrimonio archeologico, *established by: Mauro Cristofani and Riccardo Francovich* (2017), 35–46.

[28] M. Baca and P. Harpring, *Categories for the description of works of art*, J. Paul Getty Trust, 2006.

[29] G. McKenna and E. Patsatzi, *SPECTRUM: The UK museum documentation standard*, Museum Documentation Association, 2007.

[30] E. Coburn, R. Light, G. McKenna, R. Stein and A. Vitzthum, LIDO-lightweight information describing objects version 1.0, *ICOM International Committee of Museums* (2010).

[31] M. Baca and M. Gill, Encoding multilingual knowledge systems in the digital age: the getty vocabularies, *Knowledge Organization* **42**(4) (2015), 232–243.

[32] E. Lanzi and P. Harpring, Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images, 2006.

[33] M. Doerr, Ontologies for cultural heritage, in: *Handbook on Ontologies*, Springer, 2009, pp. 463–486.

[34] M. Doerr, The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata, *AI magazine* **24**(3) (2003), 75.

[35] M. Doerr and P. LeBoeuf, Modelling intellectual processes: the FRBR-CRM harmonization, in: *International DELOS Conference*, Springer, 2007, pp. 114–123.

[36] S.-a. Knight and J. Burn, Developing a framework for assessing information quality on the World Wide Web., *Informing Science* **8** (2005).

[37] F. Naumann and C. Rolker, *Assessment methods for information quality criteria*, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, Institut für Informatik, 2005.
[38] G.R. Mitchell, S. Church, T. Bartosh, G.D. Godana, R. Stohr, S. Jones and A. Knowlton, Measuring scholarly metrics (2011).
[39] D. Walton, *Argumentation schemes for presumptive reasoning*, Routledge, 2013.
[40] Y.W. Lee, D.M. Strong, B.K. Kahn and R.Y. Wang, AIMQ: a methodology for information quality assessment, *Information & management* **40**(2) (2002), 133–146.
[41] C. Batini, C. Cappiello, C. Francalanci and A. Maurino, Methodologies for data quality assessment and improvement, *ACM computing surveys (CSUR)* **41**(3) (2009), 16.
[42] J.-R. Park, Metadata quality in digital repositories: A survey of the current state of the art, *Cataloging & classification quarterly* **47**(3–4) (2009), 213–228.
[43] P. Wilson, Second-hand knowledge: An inquiry into cognitive authority (1983).
[44] S. Peroni and D. Shotton, The SPAR ontologies, in: *International Semantic Web Conference*, Springer, 2018, pp. 119–136.
[45] M. Daquino, mAuth - Corpus analysis, 2019, data retrieved from Figshare, https://doi.org/10.6084/m9.figshare.7411262.
[46] S. Chakrabarti, M. Van den Berg and B. Dom, Focused crawling: a new approach to topic-specific Web resource discovery, *Computer networks* **31**(11–16) (1999), 1623–1640.
[47] D. Kelly et al., Methods for evaluating interactive information retrieval systems with users, *Foundations and Trends in Information Retrieval* **3**(1–2) (2009), 1–224.
[48] L.T. Su, Evaluation measures for interactive information retrieval, *Information Processing & Management* **28**(4) (1992), 503–516.
[49] J. Cheng, X. Hu and P.B. Heidorn, New measures for the evaluation of interactive information retrieval systems: Normalized task completion time and normalized user effectiveness, in: *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem*, Vol. 47, American Society for Information Science, 2010, p. 72.
[50] F.F. Reichheld and R. Markey, *The ultimate question 2.0: How net promoter companies thrive in a customer-driven world*, Harvard Business Press, 2011.
[51] J.L. Fleiss, Measuring nominal scale agreement among many raters., *Psychological bulletin* **76**(5) (1971), 378.
[52] M. Daquino, mAuth - Results of the User Study, 2018, figshare. DOI: 10.6084/m9.figshare.7409384.v2.
[53] M. Daquino, mAuth - Results of the User Study, 2019, data retrieved from Figshare, https://doi.org/10.6084/m9.figshare.7409384.
[54] J.R. Landis and G.G. Koch, The measurement of observer agreement for categorical data, *biometrics* (1977), 159–174.
[55] R. Verborgh, M. Vander Sande, P. Colpaert, S. Coppens, E. Mannens and R. Van de Walle, Web-Scale Querying through Linked Data Fragments., in: *LDOW*, Citeseer, 2014.
[56] C.A. Knoblock and P. Szekely, Exploiting Semantics for Big Data Integration, *AI Magazine* **36**(1) (2015).
[57] S. Peroni and D. Shotton, FaBiO and CiTO: ontologies for describing bibliographic resources and citations, *Web Semantics: Science, Services and Agents on the World Wide Web* **17** (2012), 33–43.