

Introducing VISU: Vagueness, Incompleteness, Subjectivity, and Uncertainty in Art Provenance Data

Fabio Mariani

Institute of Philosophy and Art History, Leuphana University Lüneburg, Universitätsallee 1, C5.418, 21335 Lüneburg, Germany

Abstract

The acronym VISU refers to Vagueness, Incompleteness, Subjectivity, and Uncertainty found in provenance records, which document the history of ownership and socio-economic custody changes of an object. VISU information represents the intellectual effort of researchers and its limits in reconstructing historical events from archival sources. Although provenance has mainly been used in the past to assess an object's artistic and economic value, it has recently become crucial information from an ethical and legal viewpoint. In light of this, there is a growing interest in structuring provenance information in a machine-readable format and making this data openly accessible to anyone, e.g., by publishing provenance data as linked open data. However, with the impetus to publish provenance linked open data, we risk losing or simplifying VISU information. After describing VISU information and analysing current community standards, this article illustrates how to represent such information in publishing provenance linked open data.

Keywords

Provenance, Linked Open Data, CIDOC CRM, Linked Art, Nanopublication

1. Introduction

Provenance records document chains of events of ownership and socio-economic custody changes of an object. These records contain historical information that answers the question: from where did it come? This article focuses on the provenances of objects with artistic or cultural value held by a gallery, library, archive, or museum (GLAM).

In the art market, documenting provenance has been a means of establishing the value of artworks since the eighteenth century [1]. For example, if a well-known and highly respected collector owned an object, then they would contribute to its supposed authenticity and aesthetic value, determining its economic value [2]. By the late twentieth century, however, provenance's moral, ethical, and legal entanglements became a subject of scrutiny and debate. As a consequence of colonialism, totalitarian regimes and two world wars, many objects improperly changed hands due to seizures, confiscations, and looting. For this reason, documenting and establishing the life story of an object has become crucial in establishing its rightful owner. The 1998 Washington Conference on Holocaust-Era Assets foregrounded the importance of provenance research to find and return art and cultural property confiscated by the Nazi regime

COMHUM 2022: Workshop on Computational Methods in the Humanities, June 09–10, 2022, Lausanne, Switzerland

✉ fabio.mariani@leuphana.de (F. Mariani)

ORCID [0000-0002-7382-0187](https://orcid.org/0000-0002-7382-0187) (F. Mariani)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

[3]. At the conference, the 44 participating governments and 13 non-governmental organisations agreed on eleven non-binding principles (“The Washington Conference Principles on Nazi-Confiscated Art”) resolving disputes over Nazi-looted art through the study of provenance. As a result of these principles, provenance research has become more professionalised, acquiring interdisciplinary characteristics. In fact, it has become something of an academic field in its own right [4]. The increased importance of provenance from not only an economic perspective, but also an ethical and legal one, has put a spotlight on the responsibility of institutions. Indeed, the accountability and transparency to which GLAM institutions are being held also depend on researching and publishing the provenance records of the objects for which they are responsible.

However, recording provenance is a complex process and requires a considerable investment of resources. On the one hand, careful research of sources is necessary to reconstruct the history of an object. On the other hand, this effort requires consideration in curating and publishing any information obtained. Moreover, the efforts of a single institution must be coordinated with other stakeholders in the GLAM domain. Recently, digital tools and methodologies have opened up new possibilities to assist the curation, publishing, and analysis of provenance data. In particular, the publication of provenance linked open data promises unprecedented levels of standardisation, enabling researchers to analyse the context of object histories in their cross-institutional complexity [5, 6, 7].

Considering the benefits of provenance linked open data, it is crucial to identify and address its related risks and challenges. The exclusion or simplification of historical complexity could reduce the quality of information, which could, in turn, cause harm when considering the ethical and legal implications of provenance. It is no coincidence that among the principles that emerged from the 1998 Washington Conference, it is advised that “consideration should be given to unavoidable gaps or ambiguities in the provenance...” [3].

In this article, we aim to categorise such “unavoidable gaps or ambiguities in the provenance” as they are likely to be compromised in publishing provenance linked open data. Indeed, recording provenance requires considerable intellectual effort in interpreting sources and formulating hypotheses about an object’s history. Such a hermeneutic process is prone to produce Vagueness, Incompleteness, Subjectivity, and Uncertainty (VISU). In publishing provenance linked open data, it is, therefore, critical to maintain the integrity of the intellectual process, with its hypothetical statements and its dealing with gaps in knowledge. Given provenance’s complexity, this article, in addition to identifying and classifying VISU information, introduces implementation solutions to represent it as linked open data. These solutions comply with current data publishing standards in the cultural heritage domain.

2. Vague, Incomplete, Subjective, and Uncertain Information

The growing requirement for institutions to be more transparent and accountable has prompted them to publish information about the provenance of objects in their collections. Currently, the provenance of an object is recorded manually as textual metadata through collection management software. Although there is not yet a shared standard for transcribing this information, the American Alliance of Museums (AAM) has drafted guidelines for compiling provenance texts [8]. To give an example, below is the provenance text of a painting by André Derain from

1910 titled “Cagnes”, which is published on the Art Institute of Chicago website and has been compiled according to the AAM guidelines:

Galerie Kahnweiler, Paris, probably acquired directly from the artist. Louis Lion & Co., New York, by Feb. 1957 [verso inscription; this and the following according to letter from Knoedler and Co., Apr. 8, 1975, copy in curatorial file]; sold to Knoedler & Co., New York, Feb. 1957; sold to the Art Institute of Chicago, 1960.¹

According to the AAM guidelines, provenance editors should list events in chronological order, from the object’s creation to the acquisition by its current owner.² An event represents a change of ownership, or custody, of the object from one party to another. Each event consists of the acquisition method, location, date, names of the parties, and their related biographical information.

Punctuation separating events has a specific meaning: a semicolon implies that the transaction from one party to another was direct; a period indicates a gap in the reconstruction of the events. For example, the period at the end of the first recorded event listed above, when Galerie Kahnweiler received the object, indicates a gap in the provenance record of the painting “Cagnes”. This means, therefore, that it is unknown how the painting passed from Galerie Kahnweiler to Louis Lion & Co., its next recorded owner. Potentially, there could have been other owners of the object that have yet to be identified.

When there is no sufficient certainty about an event, the AAM guidelines suggest using the terms “probably” and “possibly”, depending on the level of uncertainty. In analysing the provenance text of the painting “Cagnes”, we can see that the authors were not certain about the first recorded event, and therefore used the phrase “probably acquired directly from the artist”.

Finally, notes can provide additional information regarding the provenance. In the above example, the Art Institute of Chicago uses notes in square brackets. Notes in compiling a provenance text are necessary since the chronology of events results from careful research of disparate archival sources, such as inventories, letters, and even photographs. Indeed, sometimes a provenance expert can find a source for reconstructing an event on the object itself. For example, we know that Louis Lion & Co. owned Derain’s artwork through an inscription on the back of the painting (“verso inscription”).

From what has been discussed, it is clear that reconstructing ownership histories is not a straightforward process since it requires intellectual and critical effort in analysing the available historical sources and formulating hypotheses. Moreover, sources are not always available to reconstruct events, and some information may not be immediately evident. We have classified these phenomena into four categories: Vagueness, Incompleteness, Subjectivity, and Uncertainty. We have gathered them under the acronym VISU, from the Latin *de visu*, meaning with your own eyes. Vagueness refers to information that is given with certainty but in an approximate way. An approximation can occur when describing spatial information (e.g., near Paris) or temporal information (e.g., circa 1945). In either case, the vagueness of the information does not affect the

¹<https://www.artic.edu/artworks/12402/cagnes> (accessed 2023-08-11).

²Usually, the creation event is omitted in the provenance text as it is recorded in other appropriate metadata fields, such as author, date, and place of creation.

certainty of the event. Incompleteness refers to a lack of information in the reconstruction of an object’s provenance. In this case, provenance experts may not have formulated any hypotheses yet to address the missing information. Subjectivity concerns the expert’s interpretive context when reconstructing an object’s provenance—how they formulated hypotheses through source analysis and deduction. Moreover, different assumptions may contradict each other. Finally, uncertainty refers to the level of confidence with which a provenance expert has expressed a hypothesis, using terms such as “possibly” or “probably”. Unlike vagueness, uncertainty questions the very occurrence of a given event.

The categories of what we define as VISU have already been a topic of interdisciplinary debate, from philosophy and mathematics to, more recently, computer science [9]. In Smithson’s taxonomy of ignorance, for example, the concept of uncertainty represents a generic term that, in turn, can be divided into more specific concepts such as vagueness and probability [10]. The latter is closer to our definition of uncertainty. In contrast, Smets distinguishes more sharply between uncertainty and imprecision in providing a taxonomy of imperfection [11]. Smets’ imprecision can be compared to the vagueness of VISU information. At least lexically speaking, a classification close to that of VISU is provided by Nagypál and Motik [12]. Here, the categories of uncertainty, subjectivity, and vagueness are defined in relation to expressing temporal knowledge. However, the meaning given to each term is different from that intended in VISU. In fact, according to their classifications, uncertainty (e.g., circa 1918), subjectivity (e.g., the dating of the Russian Revolution), and vagueness (e.g., in February 1918) are all ascribable to the concept of vagueness in VISU. In analysing uncertainty in the digital humanities domain, Piotrowski recognises the conflict of interpretations between scholars as an additional aspect of dealing with “uncertain, vague, incomplete, or missing information” [9]. In doing so, Piotrowski partially anticipates the classification we propose with the acronym VISU, since the conflict of interpretations is one aspect of what we define as subjectivity.

3. Provenance Linked Open Data

As previously discussed, institutions currently create and share provenance records in text format. Although provenance texts are stored and published online digitally via collection management systems, the text format limits the use of provenance as a research and study tool. Indeed, it is currently impossible to use provenance data to perform large-scale analyses across multiple institutions through the application of, for example, digital methods such as big data queries, network analysis, and spatial analysis [13]. These limitations can be attributable, on the one hand, to the fact that textual information is not machine-readable and, on the other hand, to the fact that it is not published according to FAIR principles [14]. Indeed, as it stands, provenance information, which is siloed as text in collection databases of institutions, is not findable, accessible, interoperable, and reusable. For these reasons, publishing provenance linked open data (LOD) has recently emerged as a promising possibility to address the standardization of provenance information produced by institutions in a machine-readable format compliant with FAIR principles [5]. Moreover, LOD respects the open data principles: that is, provenance LOD can be used by anyone for any purpose.³ Provenance data should be published as open data,

³<https://opendefinition.org/> (accessed 2023-08-11).

not only because it involves historical facts but also because of the significance of provenance for institutional accountability and transparency.

A significant early experiment in publishing an institution's provenance records as LOD was carried out within the Art Tracks project, an initiative of the Carnegie Museum of Art (CMAA), which took place from 2014 to 2017 [15]. In particular, Art Tracks implemented the CMAA Digital Provenance Standard for modelling provenance LOD following the CIDOC CRM schema, the international standard for exchanging digital information regarding cultural heritage (ISO 21127).⁴ The schema of CIDOC CRM is event-based since its semantic structure has temporal entities (`crm:E2_Temporal Entity`) as its core [16]. A temporal entity, such as an event (`crm:E5_Event`), can link to time (`crm:E52_Time-Span`), space (`crm:E53_Place`), or event actors (`crm:E39_Actor`). However, the centrality of the temporal entity means that an actor, such as a person (`crm:E21_Person`), cannot link directly to a time or place. For example, CIDOC CRM does not express an individual's birth date as a person's attribute, but rather as a specific event, birth (`crm:E67_Birth`), linked to a time and involving that person. In turn, the birth event can be linked to the location of the event.

In order to make CIDOC CRM modelling more accessible to institutional practitioners, Linked Art, a community of cultural heritage institutions, developed a CIDOC CRM application profile.⁵ In addition to CIDOC CRM, the Linked Art Data Model integrates the Getty's controlled vocabularies, such as the Art and Architecture Thesaurus (AAT), to identify domain-specific terms via URI.⁶ The integration of CIDOC CRM and Getty vocabularies, combined with the support of a large and active community behind the Linked Art Data Model, make this application profile an ideal candidate for the standardisation of publishing provenance LOD. Indeed, modelling provenance LOD is one of the aspects that the Linked Art Data Model covers in detail. According to Linked Art, a provenance record structured as LOD is a succession of provenance events (or activities), structured in CIDOC CRM as `crm:E7_Activity`. An activity can itself consist of multiple activities (sub-activities), expressing more complex events. Linked Art provides a pattern for defining the characteristics of events: the object(s) involved, the actors participating, the location, and the time. Examples of how to structure the data are given depending on the different types of activities. For example, an activity describing the purchase of an object may contain two sub-activities. The first activity consists of the acquisition of the object given by the seller and received by the buyer, while the second constitutes the payment made by the buyer to the seller. Similarly, exchanging two objects involves two sub-activities, each describing the respective ownership change.

⁴CIDOC CRM (version 7.2) is the Conceptual Reference Model (CRM) implemented by the International Committee for Documentation (CIDOC) of the International Council of Museums (<https://www.cidoc-crm.org/>, accessed 2023-08-11).

⁵<https://linked.art/model/> (accessed 2023-08-11).

⁶<https://www.getty.edu/research/tools/vocabularies/aat/> (accessed 2023-08-11).

```
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<https://example.com/activity/4> a crm:E7_Activity ;
  rdfs:label "Purchased by the Art Institute of Chicago from Knoedler & Co. in 1960" ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300055863> ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300417642> ;
  crm:P4_has_time-span [ a crm:E52_Time-Span ;
    crm:P82a_begin_of_the_begin "1960-01-01T00:00:00Z" ;
    crm:P82b_end_of_the_end "1960-12-31T23:59:59Z" ] ;
  crm:P9_consists_of [ a crm:E8_Acquisition ;
    crm:P22_transferred_title_to [ a crm:E74_Group ;
      rdfs:label "The Art Institute of Chicago" ] ;
    crm:P23_transferred_title_from [ a crm:E74_Group ;
      rdfs:label "Knoedler & Co." ] ;
    crm:P24_transferred_title_of [ a crm:E22_Human-Made_Object ;
      rdfs:label "Cagnes" ] ] .
```

Listing 1: RDF description, serialized in Turtle format, of the purchase of the painting “Cagnes” by the Art Institute of Chicago from Knoedler & Co. in 1960.

Listing 1 shows the RDF description of the last provenance event of André Derain’s painting “Cagnes”: the purchase of the artwork by the Art Institute of Chicago from Knoedler & Co. in 1960. RDF, Resource Description Framework, is a World Wide Web Consortium standard for information exchange as LOD. The activity (crm:E7_Activity) is classifiable according to AAT vocabulary as “provenance” (aat:300055863) and “purchase” (aat:300417642). Moreover, the activity took place in 1960, a time span expressed through its time limits: the begin of the begin “1960-01-01T00:00:00Z” (the minimum possible date) and the end of the end “1960-12-31T23:59:59Z” (the maximum possible date). The activity has a sub-activity (crm:E8_Acquisition), which describes the acquisition of the painting “Cagnes” by the Art Institute of Chicago from Knoedler & Co.

In making CIDOC CRM modelling usable to practitioners, Linked Art deliberately leaves out some aspects that would complicate the accessibility of the data model, such as uncertainty and data provenance. However, this choice compromises the integrity of VISU information when modelling provenance LOD. As discussed in the previous section, VISU information is based on the intellectual work of provenance experts, who research and record provenance. Moreover, with VISU information, historical debate and hypothesis-making become critical to achieving the most scientifically accurate reconstruction of an object’s history. Forgoing VISU information thus not only compromises the integrity of the data but also prevents debate, thereby reducing its usefulness for research. This phenomenon, also referred to as the “lure of objectivity”, is one of the major challenges in digital humanities [17]. We, therefore, intend to safeguard the complexity of VISU information by making it machine-readable according to LOD standards and compatible with the Linked Art Data Model. The following sections describe the challenges, opportunities, and solutions in dealing with VISU information as LOD.

4. Vagueness

By introducing VISU information, we have established a clear distinction between the concepts of vagueness and uncertainty that previous scholarship, as noted, has not made consistently. Vagueness indicates the approximation of a datum. Approximating a datum per se does not compromise the statement's certainty. For example, to say that an event occurred near Paris is to approximate the geographical location of a temporal entity. The fact is not called into question. Similarly, the existence of an event that occurred circa 1945 is not questioned by the temporal approximation of the date.

Since vagueness concerns spatial and temporal information, it depends on the measures and language used by historical sources. Indeed, whereas technology allows us to calculate space and time with utmost precision, human language can hardly replicate its accuracy. Compare, for example, the limitations of language in traditional art market information, such as the inventory of an art dealer, with the measures of modern digital data, such as an online auction house database. In the written inventory of an art dealer, an event date can achieve maximum accuracy by expressing the year, month, and day of the event. Usually, however, vague reference systems such as months or seasons are used. Seldom does an author of a source go into details such as the exact hour of an event. In contrast, an online auction house database can capture the moment of purchase to a thousandth of a second. Similarly, whereas human language cannot go beyond the precision of an address to indicate spatial information, technology allows us to pinpoint the geographical coordinates of a place with greater accuracy.

In addition to measuring instruments and human language, an approximation can result from a lack of information. For example, in the provenance text of the painting "Cagnes", the second provenance event states that Louis Lion & Co. owned the work "by Feb. 1957". The author of the provenance record used this expression because they had no sources to establish when Louis Lion & Co. received the object precisely. We do not even know who the previous owner was, expressed using a period that signifies a gap in the painting's provenance text. What the provenance expert can establish from the historical information available, however, is that Louis Lion & Co. had the object in February 1957 since the sources show that they sold the work to Knoedler & Co. in that month. Thus, we can assert that the acquisition of the painting by Louis Lion & Co. took place between 1910, the previous known date and thus the lower limit of the possible time interval, and 28 February 1957, the last day of the month in which Knoedler & Co. acquired the object. Experts can formulate subjective hypotheses with different degrees of uncertainty based on a vague time expression such as "by Feb. 1957". For example, according to the available information, it is possible, although very unlikely, that Louis Lion & Co. acquired the object on 28 February 1957 and sold it to Knoedler & Co. on the same day.

In CIDOC CRM and Linked Art, one can already model some vague information. This ensures that information is not falsified when publishing provenance LOD, which runs the risk of making vague information seemingly precise. It also opens up possibilities for data analysis and visualization that include this layer of complexity. Concerning the approximation of spatial data, CIDOC CRM introduces the property `crm:P189_approximates`. Using this property makes it possible to establish an approximation relation between two places. For example, in Listing 2, we see how the place "Paris", defined as a point in space, approximates the expression "near Paris". In this way, we preserve the vagueness of the information on the one hand. And on the

other hand, we model a point in space that, albeit approximate, allows us to query the geospatial datum and visualize it on a map. The Linked Art Data Model already includes this modelling solution.

```
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<https://example.com/place/1> a crm:E53_Place ;
  rdfs:label "Paris" ;
  crm:P168_place_is_defined_by "POINT(2.2769957 48.8589466)" ;
  crm:P189_approximates [ a crm:E53_Place ;
    rdfs:label "near Paris" ] .
```

Listing 2: RDF description, serialized in Turtle format, of the “near Paris” approximation.

As far as temporal information is concerned, as we have already seen when introducing CIDOC CRM and Linked Art in Listing 1, it is represented as a time span. Thanks to the properties `crm:P82a_begin_of_the_begin` and `crm:P82b_end_of_the_end`, this type of modelling makes it possible to model several vague chronological pieces of information [18]. For example, Listing 3 shows the modelling of the time span in which Louis Lion & Co. acquired the painting “Cagnes”. As previously discussed, the activity occurred sometime between 1910 (begin of the begin) and February 1957 (end of the end). In addition, this approach allows for modelling other approximate expressions in which an event occurred, such as months, seasons, years, decades, centuries, and millennia.

```
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<https://example.com/timepsan/1> a crm:E52_Time-Span ;
  rdfs:label "between 1910 and February 1957" ;
  crm:P82a_begin_of_the_begin "1910-01-01T00:00:00Z" ;
  crm:P82b_end_of_the_end "1957-02-28T23:59:59Z" .
```

Listing 3: RDF description, serialized in Turtle format, of the time span between 1910 and February 1958.

Although CIDOC CRM allows us to model temporal information as a time span, it does not allow the representation of an approximation that occurs around a date, such as the expression “circa 1945”. However, it is possible to integrate the CRMgeo module to overcome this limitation. This extension of CIDOC CRM, dedicated to a more complex representation of spatiotemporal data, introduces the property `crmgeo:Q13_approximates` [19]. Like the `crm:P189_approximates` for places, the `crmgeo:Q13_approximates` property establishes an approximation relation between two time spans [20]. As an example, Listing 4 describes how the time span “1945”—with a begin of the begin as 1 January 1945 and an end of the end as 31 December 1945—approximates the vague time span “circa 1945”. Therefore, we believe this solution, similar to the one adopted

for spatial approximation, can be integrated into the Linked Art Data Model.

```
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix crmgeo: <http://www.cidoc-crm.org/rdfs/1.2/crmgeo#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<https://example.com/timespan/1> a crm:E52_Time-Span ;
  rdfs:label "1945" ;
  crm:P82a_begin_of_the_begin "1945-01-01T00:00:00Z" ;
  crm:P82b_end_of_the_end "1945-12-31T23:59:59Z" ;
  crmgeo:Q13_approximates [ a crm:E52_Time-Span ;
    rdfs:label "circa 1945" ] .
```

Listing 4: RDF description, serialized in Turtle format, of the time expression “circa 1945” approximated by the time span 1945.

5. Incompleteness

In dealing with incompleteness, we must consider a trivial but essential fact: it is impossible to model as LOD what is unknown. Indeed, incompleteness is the only VISU information we cannot address directly in the modelling phase. However, conscious modelling of known information can help to address incompleteness through subsequent data analysis and in the hypotheses-making phase. Although we cannot model what we do not know, we can establish patterns of incompleteness against which we analyse the available information [21]. This approach first allows us to identify where and what information is missing and, secondly, to formulate new hypotheses with the help of data analysis.

The first pattern of incomplete provenance information that we can identify concerns gaps in the object’s chain of activities. The importance of considering this kind of incompleteness for the integrity of a provenance record already emerges from the AAM guidelines. As we have already described, in provenance texts, events are divided by semicolons if transactions are direct and by periods if there are gaps in the ownership history of an object. Since we cannot directly model the presence of a gap as LOD, we must define a pattern to detect this incompleteness in the data. Linked Art describes the chronological linkage of provenance activities through the properties `crm:P183_ends_before_the_start_of` and `crm:P183i_starts_after_the_end_of`. These two properties allow us to determine whether an event occurred before or after another [22]. While they may help establish a chronological order of events, these properties are insufficient for identifying gaps between them. To detect such gaps, we must formulate the incompleteness pattern of the chain of activities: there is a gap between two events, A and B, linked in chronological succession (`Activity_A crm:P183_ends_before_the_start_of Activity_B`) if the party who receives the object in `Activity_A` is not the one who parts with it in `Activity_B`.

In Listing 5, we describe the activities involving the acquisition of the painting “Cagnes” by Galerie Kahnweiler from the artist and the subsequent acquisition by Louis Lion & Co. In this case, the scenario respects the incompleteness pattern of the chain of activities insofar as

Galerie Kahnweiler was not the owner who gave the object to Louis Lion & Co. Identifying such a gap in analysis can lead to the formulation of new hypotheses since there may have been one or more intermediate owners prior to Louis Lion & Co. The gap in question is of significant interest to scholars as it conceals the events that caused the object to be moved from Paris to New York. Moreover, the gap overlaps with two world wars that affected, among other aspects, the circulation of artworks, legal or otherwise. In this scenario, the publication of provenance LOD is valuable because it allows us to analyse large amounts of provenance data from different institutions. Indeed, through network analysis, we can identify the most frequent pathways of artworks that, at some point in their lives, passed through Galerie Kahnweiler, as well as the most prominent agents from whom Louis Lion & Co. purchased artworks, thus opening up new hypotheses that try to bridge the gap.

```
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<https://example.com/activity/1> a crm:E7_Activity ;
  rdfs:label "Acquired by Galerie Kahnweiler from André Derain" ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300055863> ;
  crm:P183_ends_before_the_start_of <https://example.com/activity/2> ;
  crm:P9_consists_of [ a crm:E8_Acquisition ;
    crm:P22_transferred_title_to [ a crm:E74_Group ;
      rdfs:label "Galerie Kahnweiler" ] ;
    crm:P23_transferred_title_from [ a crm:E21_Person ;
      rdfs:label "André Derain" ] ;
    crm:P24_transferred_title_of [ a crm:E22_Human-Made_Object ;
      rdfs:label "Cagnes" ] ] .

<https://example.com/activity/2> a crm:E7_Activity ;
  rdfs:label "Acquired by Louis Lion & Co." ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300055863> ;
  crm:P9_consists_of [ a crm:E8_Acquisition ;
    crm:P22_transferred_title_to [ a crm:E74_Group ;
      rdfs:label "Louis Lion & Co." ] ;
    crm:P24_transferred_title_of [ a crm:E22_Human-Made_Object ;
      rdfs:label "Cagnes" ] ] .
```

Listing 5: RDF description, serialized in Turtle format, of the acquisition of the painting “Cagnes” by Galerie Kahnweiler from the artist, and the subsequent acquisition by Louis Lion & Co.

Different patterns of incompleteness can result from other missing constituents of an activity. As we discussed in introducing Linked Art, the data model introduces a pattern of event constituents. An activity is determined not only by its participating actors, but also by time and place and the object(s) involved. In addition, an activity can consist of several sub-activities, depending on its type. Thus, Activity_A is incomplete if the time, place, or object(s) involved are not expressed, or if one or more of the sub-activities associated with its type are missing.

When the time of an activity is unknown, incompleteness can be solved by generating vague information, that is, by defining that the event occurred in a time interval between the last previously known date before the activity and the first subsequently known date after the activity. As previously discussed in the section on vagueness, while we do not know when Louis Lion & Co. acquired the painting “Cagnes”, we can infer that the activity occurred sometime between 1910 and 28 February 1957. The incompleteness of an activity’s location proves to be a more challenging piece of information to reconstruct from the sources, except for when an event is specific, like an auction. In provenance texts, we find mainly geographical information about the actors. This can sometimes be useful in hypothesising the locations where events occurred. For example, we can infer that the purchase of “Cagnes” by Knoedler & Co. from Louis Lion & Co. occurred in New York since both companies were located there.

In contrast, the incompleteness concerning an activity and its sub-activities depends on the type of event, for which the Linked Art Data Model introduces a distinct structure. For example, as discussed, a purchase activity involves two sub-activities: 1) the acquisition of the sold object and 2) payment. In the previous section, we presented the LOD example of modelling the purchase of “Cagnes” in 1960 by the Art Institute of Chicago (Listing 1), the last event in the provenance record of that object. We can therefore assert that the activity is incomplete, since there is no sub-activity related to the payment made by the Art Institute of Chicago to the seller. Similarly, a provenance activity that concerns the exchange of one object for another will be incomplete if it consists of only one sub-activity, since one of the objects involved is not registered.

Additional types of incompleteness, which are difficult to ascribe to a fixed pattern, concern the biographical information of the actors involved. Missing biographical information of interest to the reconstruction and study of provenance may be: birth and death (or formation and dissolution, in the case of organisations), period and place of activity, and relationships to other actors. In addition to the direct intervention of historians, it is possible to use external knowledge published as LOD to fill in these gaps, such as the Getty’s Union List of Artist Names (ULAN).⁷ This controlled vocabulary can enrich our understanding of the actors of provenance activities with additional biographical information. In turn, enriching biographical information can help fill in other types of incompleteness. For example, by using the ULAN entity information of Louis Lion & Co. (ulan:500449799), we learn that the company has been in business since 1949. This new information allows us to, in turn, narrow down its purchase of the painting “Cagnes” to a time interval from 1949 to 28 February 1957.

Finally, it should be noted that provenance texts have a considerable bias in the representation of women. Many women are represented by their husbands’ names (“Mrs John Doe”) or even by the expression “the artist’s wife”. Such expressions compromise the historical representation of women and make it difficult for historians to identify female actors. For example, expressions such as “the artist’s wife” are of little help if an artist had multiple wives. Modelling provenance LOD thus becomes an opportunity for historians to remedy such bias and finally give proper representation to people.

⁷<https://www.getty.edu/research/tools/vocabularies/ulan/> (accessed 2023-08-11).

6. Subjectivity

Reconstructing the history of an art object is the result of laborious research by provenance experts, who hypothesise through the interpretation of sources what might have happened. Of course, the hypotheses of different experts may contradict each other, evolve with time, and become obsolete in light of new findings. As provenance texts stand, however, they cannot capture the hermeneutic and dialectical complexity of this intellectual process. In fact, except for notes to provide additional context for specific hypotheses, the texts are not accompanied by any publication information. For example, the author's name and publication date are critical metadata for information authority and versioning. The lack of versioning, in particular, can lead to the harmful practice of deleting a provenance text whenever an institution produces a new version. In this way, a debate concerning the provenance of an object is arbitrarily steered in a single direction, collapsing the idea that different historical interpretations can coexist.

It is possible to include publication information and versioning when publishing provenance LOD by implementing what is known as the data provenance of provenance data [23, 7, 24]. Just as we can trace an artwork's ownership history, we can trace the recording history of a given datum through data provenance. The recording history tracks when a datum was created, by whom, and when it was modified.

CIDOC CRM introduces the class `crm:E13_Attribute_Assignment`, a subclass of `crm:E7_Activity`, to describe the context in which an assertion is made regarding an entity. An attribute assignment is the entity with which CIDOC CRM represents the n-ary relationship between the asserted entity and the assertion information. In this way, in addition to defining the asserted value, we can add additional statements to describe the context of the assertion, such as the author and date. Although this solution is also adopted in Linked Art to define, for example, authorship attribution, it tends to be verbose and redundant [25]. Focusing on the case of data provenance of provenance data, we found issues related to using attribute assignments to represent this type of information. An n-ary relation enables us to describe the context of an assertion pertaining only to a single statement. However, in the case of provenance, hypothesis-making does not concern a single statement but the assertion of an entire event and, thus, multiple statements. In this scenario, should we model an attribute assignment for each statement, we would need to repeat the same information multiple times. This situation would be even more complex in case of contradictory assumptions, as this requires us to produce multiple attribute assignments to describe conflicting hypotheses, resulting in an additional increase in statements. Moreover, such a solution would result in the coexistence in the same RDF graph of different and contradictory information about the same fact, compromising the usability of the data.

Given the nature of provenance information and the issues arising using attribute assignments, we considered other approaches. Among the many methods to represent data provenance as LOD, nanopublication is one of the most suitable [26]. Nanopublication is a way of publishing an atomic unit of information as LOD, providing data provenance and publication information [27]. In this way, it is possible to trace and reference these atomic units of information independently of the entire dataset, making the knowledge expressed more authoritative and compliant with FAIR principles [28].

In presenting provenance LOD modelling according to the Linked Art Data Model, we have

Table 1

HiCO classes and properties alignment with CIDOC CRM, importing the CRMinf module.

HiCO	CIDOC CRM (CRMinf)
hico:InterpretationAct	crminf:I1_Argumentation
hico:InterpretationCriterion	crm:E55_Type
hico:hasInterpretationType	crm:P2_has_type
crm:P14_carried_out_by	crm:P14_carried_out_by
prov:startedAtTime	crm:P4_has_time-span
cito:citesAsEvidence	crm:P16_used_specific_object
prov:wasGeneratedBy	crminf:J2_concluded_that → crminf:I2_Belief → crminf:J4_that
hico:hasInterpretationCriterion	crm:P32_used_general_technique
hico:isExtractedFrom	crm:P70i_is_documented_in
prov:wasInfluencedBy	crm:P15_was_influenced_by

seen how provenance activities are the constitutive elements of an event-based model. In light of this, we consider the provenance activity as the atomic unit of a provenance record published as a nanopublication. Thus, publishing provenance LOD as a nanopublication implies publishing each provenance activity as a stand-alone, referenceable, and citable unit. In this way, two conflicting hypotheses about the same activity can coexist while older hypotheses that have become obsolete can remain accessible to scholars [29]. In addition, each nanopublication expresses metadata about the creation of the information and its publication. We can thus publish the data provenance of provenance data.

The structure of a nanopublication consists of three separate named graphs. A named graph is an RDF graph identified by a URI, which allows one to assert information about it [30]. The first graph of the nanopublication, the assertion graph, is devoted to the information on the published atomic unit. In the case of provenance data, it contains statements about a single provenance activity.

The second graph, the provenance graph, is dedicated to the data provenance related to the assertion graph. It contains statements about how the knowledge expressed in the assertion graph was produced. For example, in a nanopublication of a provenance activity, the provenance graph describes the context of the hypothesis formulated by an expert, including the author's identity, the date, the scientific method used, and the sources consulted by the author. Finally, the publication info graph, the third graph of a nanopublication, provides metadata about the entire nanopublication, such as the creator, creation date, and license.

In expressing the subjectivity of information, such as the contexts of different hypotheses and the possible conflicts between them, it is necessary to focus on modelling the interpretation context in the provenance graph. The Historical Context Ontology (HiCO) is dedicated to expressing as LOD the context of a hermeneutic activity performed by a scholar in formulating a hypothesis through the interpretation of sources [31]. HiCO is an extension of the PROV ontology, the standard model dedicated to modelling data provenance on the web [32]. Given its purpose, such an ontology is ideal for representing provenance graph information. HiCO revolves around one activity: the interpretation (hico:InterpretationAct). This activity represents the action of the scholar in formulating a hypothesis of which, among other types of informa-

tion, we can express the type of interpretation (`hico:hasInterpretationType`), the criterion of interpretation (`hico:hasInterpretationCriterion`), the time frame in which the interpretation was carried out (`prov:startedAtTime`), the resources used (`cito:citesAsEvidence`), and the influence of other hypotheses (`prov:wasInfluencedBy`). To integrate HiCO into the Linked Art Data Model, we propose aligning HiCO and CIDOC CRM, as shown in Table 1. In aligning the two ontologies, it is necessary to use CRMinf, a CIDOC CRM module dedicated to modelling inference-making activities.⁸ Specifically, CRMinf introduces the argumentation activity (`crminf:I1_Argumentation`) semantically comparable with HiCO’s interpretation act (`hico:InterpretationAct`). In addition, the module allows for more granular modelling of the argumentation result, expressed in the assertion graph. While HiCO uses the PROV ontology property `prov:wasGeneratedBy` to indicate that the assertion graph resulted from an interpretation act, CRMinf uses an n-ary relation. As a result, the argumentation generates a belief (`crminf:I2_Belief`), which is, in turn, expressed by the assertion graph. As discussed in the next section on uncertainty, an n-ary relation allows one to assert information about the relation, which is impossible in a binary relation.

```

@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix crminf: <http://www.cidoc-crm.org/cidoc-crm/CRMinf/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix np: <http://www.nanopub.org/nschema#> .
@prefix dct: <http://purl.org/dc/terms/> .

<https://example.com/nanopub/3/head> {
  <https://example.com/nanopub/3> a np:Nanopublication ;
  np:hasAssertion <https://example.com/nanopub/3/assertion_graph> ;
  np:hasProvenance <https://example.com/nanopub/3/provenance_graph> ;
  np:hasPublicationInfo <https://example.com/nanopub/3/pubinfo_graph> .
}

<https://example.com/nanopub/3/assertion_graph> {
  <https://example.com/activity/3> a crm:E7_Activity ;
  rdfs:label "Purchased by Knoedler & Co. from Louis Lion & Co. in February 1957" ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300055863> ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300417642> ;
  crm:P4_has_time-span [ a crm:E52_Time-Span ;
    crm:P82a_begin_of_the_begin "1957-02-01T00:00:00Z" ;
    crm:P82b_end_of_the_end "1957-02-28T23:59:59Z" ] ;
  crm:P9_consists_of [ a crm:E8_Acquisition ;
    crm:P22_transferred_title_to [ a crm:E74_Group ;
      rdfs:label "Knoedler & Co." ] ;
    crm:P23_transferred_title_from [ a crm:E74_Group ;
      rdfs:label "Louis Lion & Co." ] ;
    crm:P24_transferred_title_of [ a crm:E22_Human-Made_Object ;
      rdfs:label "Cagnes" ] ] .
}

```

⁸<https://www.cidoc-crm.org/crminf/> (accessed 2023-08-11).

```

<https://example.com/nanopub/3/provenance_graph> {
  <https://example.com/argumentation/3> a crminf:I1_Argumentation ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300444173> ;
  crm:P14_carried_out_by [ a crm:E74_Group ;
    rdfs:label "The Art Institute of Chicago" ] ;
  crm:P16_used_specific_object [ a crm:E33_Linguistic_Object ;
    rdfs:label "letter from Knoedler and Co., Apr. 8, 1975." ;
    crm:P2_has_type <http://vocab.getty.edu/aat/300026879> ;
    crm:P94i_was_created_by [ a crm:E65_Creation ;
      crm:P4_has_time-span [ a crm:E52_Time-Span ;
        crm:P82a_begin_of_the_begin "1975-04-08T00:00:00Z" ;
        crm:P82b_end_of_the_end "1975-04-08T23:59:59Z" ] ;
        crm:P14_carried_out_by [ a crm:E74_Group ;
          rdfs:label "Knoedler & Co." ] ] ] ;
    crminf:J2_concluded_that [ a crminf:I2_Belief ;
      crminf:J4_that <https://example.com/nanopub/3/assertion_graph> ] .
  }
}

<https://example.com/nanopub/3/pubinfo_graph> {
  <https://example.com/nanopub/3> dct:created "2023-08-11T16:31:08Z" ;
  dct:creator <https://orcid.org/0000-0002-7382-0187> ;
  dct:source <https://www.artic.edu/artworks/12402/cagnes> ;
  dct:license <https://creativecommons.org/publicdomain/zero/1.0/> .
}

```

Listing 6: Nanopublication, serialized in TriG format, of the purchase of the painting “Cagnes” by Knoedler & Co. from Louis Lion & Co. in February 1957.

Listing 6 shows the nanopublication of the provenance activity in which Knoedler & Co. purchased the painting “Cagnes” from Louis Lion & Co. in February 1957. The structure of the nanopublication is defined using the Nanopublication Ontology.⁹ According to the note in the original provenance text, the assumption made by the Art Institute of Chicago is based on a “letter from Knoedler and Co., Apr. 8, 1975.” The information is structured using HiCO’s alignment to CIDOC CRM. The Getty AAT vocabulary is used to assign the entity types, as standard practice in Linked Art. In particular, the argumentation has the entity type “provenance remark” (aat:300444173), while the linguistic object used to formulate hypotheses has the entity type “letter” (aat:300026879). The metadata of the publication info graph, such as creation date, creator, source and license, are structured using properties from the Dublin Core Metadata Initiative (DCMI) Metadata Terms, a set of standardized metadata elements to describe digital resources.¹⁰

⁹https://nanopub.net/guidelines/working_draft/ (accessed 2023-08-11).

¹⁰<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> (accessed 2023-08-11).

7. Uncertainty

According to the literature, the terms “uncertainty” and “vagueness” are related, if not conflated. For example, in documenting evidence interpretation in archaeology using CIDOC CRM, Niccolucci and Hermon merge the concepts of vagueness and uncertainty in the same concept of reliability [33]. However, as discussed in previous sections, we distinguish between vagueness and uncertainty. The reliability of vague information lies in the accuracy of the data approximation. In contrast, the reliability of uncertain information lies in the probability of the data’s factuality. In light of what was discussed in the previous section, we can therefore correlate the concept of uncertainty to subjectivity, as it expresses the degree of confidence in making a hypothesis. As we have already seen, AAM guidelines introduce the possibility of expressing uncertainty about a piece of information. Terms such as “possibly” or “probably” express levels of uncertainty depending on the provenance expert’s degree of confidence. Regarding provenance LOD modelling, Art Tracks uses a boolean value to express certainty about some information [34], and Linked Art deliberately avoids adding this degree of complexity. Examining other attempts to model uncertainty in CIDOC CRM, in the previously mentioned work by Niccolucci and Hermon, the reliability of information is expressed through fuzzy logic, with a subjective coefficient ranging from 0 (not credible) to 1 (absolutely true) [33].

When analysing provenance texts, we noticed that uncertainty coincides with the patterns we identified when dealing with incompleteness. In the presence of a gap, hypotheses become less confident. Since uncertainty is related to making hypotheses, we could have multiple contradictory hypotheses of varying degrees of certainty to fill a given gap. For this reason, the nanopublication solution is effective since it can separate various hypotheses with their degrees of certainty in different assertion graphs, allowing for the coexistence of multiple hypotheses with varying degrees of certainty.

While modelling uncertainty as information associated with the act of interpreting has already been proven possible using HiCO [35], we take a different approach. We align HiCO with CIDOC CRM, particularly with CRMInf. The use of this module to model uncertainty in provenance data has already been hypothesised by Smith in analysing the potential of provenance LOD [36]. As we have seen when dealing with subjectivity modelling, our alignment involves describing the product of the `crminf:I1_Argumentation` expressed in the assertion graph with an *n*-ary relation through the `crminf:I2_Belief` entity. The argumentation does not generate an assertion graph but instead concludes with a belief that is, in turn, expressed in the assertion graph. Thus, we can link additional information to the `crminf:I2_Belief` entity, such as the `crminf:I6_Belief_Value`. The belief value represents the truth value of a belief produced by an argumentation. The CRMInf module requires determining a belief value scale with at least three values.

Staying true to the approach of the Linked Art Data Model, we delineate a belief value scale within Getty’s AAT vocabulary. A `crminf:I2_Belief` can have as `crminf:I2_Belief_Value`: “true” (aat:300068765), “probably” (aat:300435721), “possibly” (aat:300435722), and “obsolete” (aat:300404908). The uncertainty terminology already used according to the AAM guidelines reoccurs through this new scale of values. In addition, we include the option of assuming the obsolescence of a given assumption. This option is fundamental to the data provenance of provenance data as it allows hypotheses to be discarded without eliminating them permanently, thus leaving them as evidence of the hermeneutic process concerning a given fact. What is

obsolete for one provenance expert may not be obsolete according to another.

```

@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix crminf: <http://www.cidoc-crm.org/cidoc-crm/CRMinf/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix np: <http://www.nanopub.org/nschema#> .
@prefix dct: <http://purl.org/dc/terms/> .

<https://example.com/nanopub/1/head> {
  <https://example.com/nanopub/1> a np:Nanopublication ;
  np:hasAssertion <https://example.com/nanopub/1/assertion_graph> ;
  np:hasProvenance <https://example.com/nanopub/1/provenance_graph> ;
  np:hasPublicationInfo <https://example.com/nanopub/1/pubinfo_graph> .
}

<https://example.com/nanopub/1/assertion_graph> {
  <https://example.com/activity/1> a crm:E7_Activity ;
  rdfs:label "Acquired by Galerie Kahnweiler from André Derain" ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300055863> ;
  crm:P183_ends_before_the_start_of <https://example.com/activity/2> ;
  crm:P9_consists_of [ a crm:E8_Acquisition ;
    crm:P22_transferred_title_to [ a crm:E74_Group ;
      rdfs:label "Galerie Kahnweiler" ] ;
    crm:P23_transferred_title_from [ a crm:E21_Person ;
      rdfs:label "André Derain" ] ;
    crm:P24_transferred_title_of [ a crm:E22_Human-Made_Object ;
      rdfs:label "Cagnes" ] ] ] .
}

<https://example.com/nanopub/1/provenance_graph> {
  <https://example.com/argumentation/1> a crminf:I1_Argumentation ;
  crm:P2_has_type <http://vocab.getty.edu/aat/300444173> ;
  crm:P14_carried_out_by [ a crm:E74_Group ;
    rdfs:label "The Art Institute of Chicago" ] ;
  crminf:J2_concluded_that [ a crminf:I2_Belief ;
    crminf:J5_holds_to_be <http://vocab.getty.edu/aat/300435721> ;
    crminf:J4_that <https://example.com/nanopub/1/assertion_graph> ] ] .
}

<https://example.com/nanopub/1/pubinfo_graph> {
  <https://example.com/nanopub/1> dct:created "2023-08-11T16:35:12Z" ;
  dct:creator <https://orcid.org/0000-0002-7382-0187> ;
  dct:source <https://www.artic.edu/artworks/12402/cagnes> ;
  dct:license <https://creativecommons.org/publicdomain/zero/1.0/> .
}

```

Listing 7: Nanopublication, serialized in TriG format, of the probable acquisition of the painting “Cagnes” by Galerie Kahnweiler from the artist.

Listing 7 shows the nanopublication of the provenance event in which Galerie Kahnweiler probably acquired the painting “Cagnes” directly from the artist. In this case, since the level of certainty was expressed with the term “probably” in the original text, we can describe the value held by the belief generated by the argumentation, with the entity `aat:300435721`.

8. Discussion and Conclusion

The classification of VISU information differentiates among four distinct yet correlated types of information, each pertaining to a specific intervention by the provenance expert. Vagueness, subjectivity, and uncertainty represent information categories we depend on when provenance records are incomplete. In the absence of information, the provenance expert can fill the gap by approximating data, formulating hypotheses, and expressing varying degrees of confidence in reconstructing facts.

Although these terms are often used synonymously, the VISU classification distinguishes between vagueness and uncertainty. In the classification’s context, vagueness pertains to the approximation of spatial and geographical information, thereby addressing the precision of the data. CIDOC CRM offers valuable elements for representing vague temporal information by modelling dates as time spans. Additionally, it enables the representation of vague spatial information by utilising the property `crm:P189_approximates`. Linked Art already includes such solutions. As we have discussed, to extend the modelling of temporal information approximation, we integrate the CRMgeo module. In this way, it is possible to describe a relation between a vague time span and its approximation using the property `crmgeo:Q13_approximates`.

By its nature, incompleteness is the only VISU information we cannot model in LOD. However, we can address incompleteness in analysis and hypothesis-making by carefully modelling the available information. Thanks to the event-based schema of CIDOC CRM and the application profile of Linked Art, we can formulate patterns for analysing incompleteness between and within different events. Thanks to these patterns, on the one hand, conscious modelling of the available information is possible, for example, by always including the sender, the one who parts with the object in an event. On the other hand, identifying and analysing gaps in provenance records makes it possible to gain new insights into the state of provenance research on a large scale, helping to determine which artworks, collectors, and historical periods to prioritise in research efforts.

In the classification of VISU information, subjective and uncertain information is correlated. It requires a change of approach from what CIDOC CRM proposes since modelling the assertion context for each triple related to a single provenance event proves inefficient and repetitive. For this reason, we introduced a different approach by publishing provenance LOD as a nanopublication. The nanopublication of provenance LOD involves publishing each provenance event as an atomic unit, of which we describe the data provenance information, thus implementing the data provenance of provenance data. In this way, we model the information asserted and the context of the hypothesis, such as author, date, and sources used. In addition, we can include conflicting hypotheses by modelling them in distinct RDF graphs. In the literature, there are already ontologies suitable for modelling the context in which a hypothesis is formulated, such as HiCO. We, therefore, aligned HiCO with CIDOC CRM, using the CRMinf module to describe

inference-making activities. Since uncertain information is related to the degree of confidence with which an expert makes a hypothesis, it is possible to model this uncertainty as LOD by qualifying the hypothesis-making context and implementing a belief value scale using terms from the Getty's AAT vocabulary.

Although compatible with the Linked Art Data Model, the solutions discussed for including VISU information in publishing provenance LOD should be considered an external module rather than a proposed extension. The representation and analysis of VISU information involve areas that, as we have seen, are deliberately outside the scope of Linked Art. One of the purposes of the Linked Art application profile is to make LOD information accessible and usable to institutional insiders. In this way, solutions such as nanopublications, although an established good practice in sharing scientific data compliant with FAIR principles, can be barriers to institutional practitioners. Indeed, the large volume of provenance texts from which we need to extract data would make it even more challenging to publish provenance LOD as nanopublications.

As discussed in this paper, VISU information is critical to the integrity of provenance LOD, and we cannot do without it in the name of simplicity. On the contrary, VISU information represents the complexity inherent in the effort to reconstruct historical events, as well as the contradictory assumptions that arise from the plurality of historical debates.

Balancing the effort of structuring provenance information as LOD with the qualitative care of VISU information requires a human-in-the-loop approach [37]. This means that, on the one hand, quantitative data structuring from provenance texts can be performed automatically by addressing natural language processing tasks through AI [38]. On the other hand, the qualitative curation of the data remains the responsibility of domain experts who can evaluate *de visu*, with their own eyes, the most ambiguous information.

Acknowledgments

The author would like to thank the three anonymous reviewers for their constructive feedback. I extend my gratitude to Marilena Daquino for valuable input and to Max Koss, Lynn Rother, and Liza Weber for their efforts in editing the article.

References

- [1] S. Raux, From Mariette to Joullain: Provenance and Value in Eighteenth-Century French Auction Catalogs, in: G. Feigenbaum, I. J. Reist (Eds.), *Provenance: An Alternate History of Art*, Getty Research Institute, Los Angeles, CA, 2012, pp. 85–103.
- [2] J. Gramlich, Reflections on Provenance Research: Values – Politics – Art Markets, *Journal for Art Market Studies* 1 (2017). doi:10.23690/JAMS.V1I2.15.
- [3] United States Department of State, Washington Conference Principles on Nazi-Confiscated Art, 1998. URL: <https://www.state.gov/washington-conference-principles-on-nazi-confiscated-art/>.
- [4] C. Fuhrmeister, M. Hopp, Rethinking Provenance Research, *Getty Research Journal* 11 (2019) 213–231. doi:10.1086/702755.

- [5] L. Rother, M. Koss, F. Mariani, Taking Care of History: Toward a Politics of Provenance Linked Open Data in Museums, in: E. Canning, E. Fry (Eds.), *Perspectives on Data*, The Art Institute of Chicago, Chicago, IL, 2022. doi:10.53269/9780865593152/06.
- [6] A. Luther, Digital Provenance, Open Access, and Data-Driven Art History, in: K. Brown (Ed.), *The Routledge Companion to Digital Humanities and Art History*, 1 ed., Routledge, Taylor & Francis Group, New York, NY, 2020, pp. 448–458. doi:10.4324/9780429505188-38.
- [7] D. Newbury, L. Lippincott, Provenance in 2050, in: J. Milosch, N. Pearce (Eds.), *Collecting and Provenance: A Multidisciplinary Approach*, Rowman & Littlefield Publishers, Lanham, MD, 2020, pp. 101–109.
- [8] N. H. Yeide, K. Akinsha, A. L. Walsh, *The AAM Guide to Provenance Research*, American Association of Museums, Washington, DC, 2001.
- [9] M. Piotrowski, Accepting and Modeling Uncertainty, *Zeitschrift für digitale Geisteswissenschaften* 4 (2019). doi:10.17175/SB004_006A.
- [10] M. Smithson, *Ignorance and Uncertainty: Emerging Paradigms*, Cognitive Science, Springer, New York, NY, 1989. doi:10.1007/978-1-4612-3628-3.
- [11] P. Smets, Imperfect Information: Imprecision and Uncertainty, in: A. Motro, P. Smets (Eds.), *Uncertainty Management in Information Systems*, Springer, Boston, MA, 1997, pp. 225–254. doi:10.1007/978-1-4615-6245-0_8.
- [12] G. Nagypál, B. Motik, A Fuzzy Model for Representing Uncertain, Subjective, and Vague Temporal Knowledge in Ontologies, in: R. Meersman, Z. Tari, D. C. Schmidt (Eds.), *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2003, pp. 906–923. doi:10.1007/978-3-540-39964-3_57.
- [13] P. B. Jaskot, Digital Methods and the Historiography of Art, in: K. Brown (Ed.), *The Routledge Companion to Digital Humanities and Art History*, 1 ed., Routledge, Taylor & Francis Group, New York, NY, 2020, pp. 9–17. doi:10.4324/9780429505188-3.
- [14] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, *The FAIR Guiding Principles for Scientific Data Management and Stewardship*, *Scientific Data* 3 (2016). doi:10.1038/sdata.2016.18.
- [15] D. Newbury, *Art Tracks: Using Linked Open Data for Object Provenance in Museums*, MW17: Museums and the Web 2017 (2017).
- [16] M. Doerr, The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata, *AI Magazine* 24 (2003) 75. doi:10.1609/aimag.v24i3.1720.
- [17] B. Rieder, T. Röhle, Digital Methods: Five Challenges, in: D. M. Berry (Ed.), *Understanding Digital Humanities*, Palgrave Macmillan UK, London, 2012, pp. 67–84.

- doi:10.1057/9780230371934_4.
- [18] J. Holmen, C.-E. Ore, Deducing Event Chronology in a Cultural Heritage Documentation System, in: B. Frischer, J. Webb Crawford, D. Koller (Eds.), *Making History Interactive. Computer Applications and Quantitative Methods in Archaeology (CAA)*. 37th International Conference, Williamsburg, Virginia, United States of America, March 22-26 (BAR International Series S2079), Archaeopress, Oxford, 2010, pp. 122–129.
- [19] G. Hiebel, M. Doerr, Ø. Eide, CRMgeo: A Spatiotemporal Extension of CIDOC-CRM, *International Journal on Digital Libraries* 18 (2017) 271–279. doi:10.1007/s00799-016-0192-4.
- [20] G. Hiebel, M. Doerr, K. Hanke, A. Masur, How to Put Archaeological Geometric Data into Context? Representing Mining History Research with CIDOC CRM and Extensions, *International Journal of Heritage in the Digital Era* 3 (2014) 557–577. doi:10.1260/2047-4970.3.3.557.
- [21] M. Destandau, J.-D. Fekete, The Missing Path: Analysing Incompleteness in Knowledge Graphs, *Information Visualization* 20 (2021) 66–82. doi:10.1177/1473871621991539.
- [22] M. Papadakis, M. Doerr, Temporal Primitives, an Alternative to Allen Operators, in: P. Ronzino (Ed.), *Proceedings of the Workshop on Extending, Mapping and Focusing the CRM co-located with 19th International Conference on Theory and Practice of Digital Libraries (2015)*, Poznań, Poland, September 17, 2015, CEUR Workshop Proceedings, CEUR-WS.org, 2015, pp. 69–78.
- [23] C. Huemer, The Provenance of Provenances, in: J. Milosch, N. Pearce (Eds.), *Collecting and Provenance: A Multidisciplinary Approach*, Rowman & Littlefield Publishers, Lanham, MD, 2020, pp. 2–15.
- [24] S. Al-Eryani, G. Bucher, S. Rühle, Ein Metadatenmodell für gemischte Sammlungen, *Bibliotheksdienst* 52 (2018) 548–564. doi:10.1515/bd-2018-0066.
- [25] M. Daquino, V. Pasqual, F. Tomasi, F. Vitali, Expressing Without Asserting in the Arts, in: G. M. Di Nunzio, B. Portelli, D. Redavid, G. Silvello (Eds.), *Proceedings of the 18th Italian Research Conference on Digital Libraries*, Padua, Italy, February 24-25, 2022, CEUR Workshop Proceedings, CEUR-WS.org, 2022.
- [26] L. F. Sikos, D. Philp, Provenance-Aware Knowledge Representation: A Survey of Data Models and Contextualized Knowledge Graphs, *Data Science and Engineering* 5 (2020) 293–316. doi:10.1007/s41019-020-00118-0.
- [27] P. Groth, A. Gibson, J. Velterop, The Anatomy of a Nanopublication, *Information Services & Use* 30 (2010) 51–56. doi:10.3233/ISU-2010-0613.
- [28] H. P. Sustkova, K. M. Hettne, P. Wittenburg, A. Jacobsen, T. Kuhn, R. Pergl, J. Slifka, P. McQuilton, B. Magagna, S.-A. Sansone, M. Stocker, M. Imming, L. Lannom, M. Musen, E. Schultes, FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources, *Data Intelligence* 2 (2020) 158–170. doi:10.1162/dint_a_00038.
- [29] I. Asif, I. Tiddi, A. J. G. Gray, Using Nanopublications to Detect and Explain Contradictory Research Claims, in: *2021 IEEE 17th International Conference on eScience*, IEEE, New York, NY, 2021, pp. 1–10. doi:10.1109/eScience51609.2021.00010.
- [30] J. J. Carroll, C. Bizer, P. Hayes, P. Stickler, Named Graphs, Provenance and Trust, in: *Proceedings of the 14th International Conference on World Wide Web - WWW '05*, Association for Computing Machinery, New York, NY, 2005, pp. 613–622. doi:10.1145/1060745.1060835.

- [31] M. Daquino, F. Tomasi, Historical Context Ontology (HiCO): A Conceptual Model for Describing Context Information of Cultural Heritage Objects, in: E. Garoufallou, R. J. Hartley, P. Gaitanou (Eds.), *Metadata and Semantics Research, Communications in Computer and Information Science*, Springer International Publishing, Cham, 2015, pp. 424–436. doi:10.1007/978-3-319-24129-6_37.
- [32] L. Moreau, P. Groth, *Provenance: An Introduction to PROV, Synthesis Lectures on Data, Semantics, and Knowledge*, Springer International Publishing, Cham, 2013. doi:10.1007/978-3-031-79450-6.
- [33] F. Niccolucci, S. Hermon, Expressing Reliability with CIDOC CRM, *International Journal on Digital Libraries* 18 (2017) 281–287. doi:10.1007/s00799-016-0195-1.
- [34] T. Berg-Fulton, D. Newbury, T. Snyder, *Art Tracks: Visualizing the Stories and Lifespan of an Artwork, MW2015: Museums and the Web 2015* (2015).
- [35] M. Daquino, V. Pasqual, F. Tomasi, Knowledge Representation of Digital Hermeneutics of Archival and Literary Sources, *JLIS.it* 11 (2020) 59–76. doi:10.4403/jlis.it-12642.
- [36] J. Smith, Toward “Big Data” in Museum Provenance, in: G. Schiuma, D. Carlucci (Eds.), *Big Data in the Arts and Humanities: Theory and Practice, Data Analytics Applications*, Auerbach Publishers, New York, NY, 2018, pp. 41–50.
- [37] L. Rother, F. Mariani, M. Koss, Interpreting Strings, *Weaving Threads: Structuring Provenance Data with AI*, in: *Sammlungsforschung im digitalen Zeitalter. Chancen, Herausforderungen und Grenzen*, Wallstein, Göttingen, 2023. Forthcoming.
- [38] L. Rother, F. Mariani, M. Koss, Hidden Value: Provenance as a Source for Economic and Social History, *Economic History Yearbook* 64 (2023) 111–142. doi:doi:10.1515/jbwg-2023-0005.