Provenance for Internet Art

Using the W3C PROV data model

Lozana Rossenova
London South Bank University / Rhizome
UK / USA
mehandzl@lsbu.ac.uk

Dr Karin de Wild
University of Leicester
UK
Kdw15@leicester.ac.uk

Dragan Espenschied
Rhizome
USA
dragan.espenschied@rhizome.org

Abstract – When a scholar, curator or archivist is researching an artwork, they need provenance, an essential piece of information that can help them evaluate as to whether a source can be trusted. This paper will investigate how to express the provenance of Internet art as linked data. One of the standards that has emerged to describe the provenance of digital data is the W3C PROV. It provides a model which facilitates description of the entities, agents and processes involved in producing data. This generic model has proven to be applicable in various contexts, including the cultural heritage domain [1, 2]. However, its potential to describe the provenance of Internet art is not yet fully explored [3]. This paper demonstrates how the PROV model can be used to describe the provenance of Internet art by applying it to a case study from Rhizome's ArtBase, an online archive dedicated to preserving works of Internet art. This paper is aimed at digital art conservators, digital curators, Web archivists and art historians.

Keywords – Provenance, Internet art, Rhizome, W3C PROV, linked data

Conference Topics – Collaboration: a Necessity, an Opportunity or a Luxury; The Cutting Edge: Technical Infrastructure and Implementation

I. Introduction

During the nineties, shortly after the wider introduction of the World Wide Web, artists started to create artworks within this online environment. These digital artworks were (and some of them still are) embedded within the Web. At the time of writing, there is not a definite name for these artworks. Various terms can be found in the literature, including "Internet art", "Net(work)-based art", "Web art" and "net art". Throughout this paper we will use the term "Internet art". Characteristic for Internet art is that the work evolves over time, often into various instantiations (or versions). In Rhizome's ArtBase, instantiations are referred to as "variants", a term developed by Dragan Espenschied, which we will also adopt in this paper.

Over the last 20 years, Internet artworks have increasingly entered museum collections and archives, and the study of these artworks is becoming part of contemporary art historical research, which brings new complexities. For instance, scholars are advised to proceed with caution when studying an Internet artwork as there is limited provenance information available (or made accessible), which is essential for critically evaluating the reliability of the source as evidence. Normally, researching the provenance of artworks includes an object study, going through resources about the artist (e.g. catalogues raisonnés) and other owners (e.g. auction results and exhibition catalogues). Not all of this is available, and sometimes not even applicable for Internet artworks.

In this paper we will demonstrate an approach to describing provenance for Internet art by testing the application of the PROV Data Model (PROV-DM). Instead of a history of ownership, PROV-DM describes "the people, institutions, entities, and activities, involved in producing, influencing, or delivering a piece of data or

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

a thing" [4]. It was developed as part of a family of documents published by the Provenance Working Group at W3C, which also include an OWL2 ontology (PROV-O), developed for mapping PROV-DM to RDF[1].

PROV-DM has several key characteristics which are applicable to the case of Internet Art. First, this model makes it possible to give an overview of the lifecycle of an artwork. The PROV model not only captures the creation of the artwork, but also how various actors contribute to or influence the work over time. For instance, these may include individuals or institutions, who commission, acquire, transfer or modify the work. Furthermore, PROV-DM can capture the different variants of a single artwork, even when these are preserved across various institutions. A single Internet artwork can be included in multiple (museum) collections, (Web) archives, whilst being part of the live Web. There is not yet a standard way of describing provenance adopted by all. PROV-DM is useful in this regard, as it can function as a provenance interchange model between heterogeneous systems, e.g. across (data) collections held by different memory institutions. Finally, PROV can also be used in conjunction with other ontologies in linked data repositories.[2]

The main reason we chose PROV-DM for the case study presented in this paper is that it was developed specifically for expressing provenance data on the Web. While other metadata schemas have been developed to model all data about a cultural heritage object (e.g. CIDOC-CRM), or focus on the preservation of highly abstracted digital objects (e.g. PREMIS)[3], no schema has been developed specifically to address the challenges of provenance description for Internet art. Comparing PROV to other existing schemas or ontologies is not the purpose of this paper. Rather, our objective is demonstrating how PROV-DM can be applied in a practical way towards modelling data and conceptualizing provenance for Internet art.

[1] https://www.w3.org/TR/prov-overview/

[2] PROV-O properties and concepts are already integrated in the latest OWL specification of PREMIS.

[3] Jefferies, et al, cite these reasons to explain why they chose PROV-DM, over other schemas, as a more practical approach to describe provenance in the Oxford Research Archive

We continue with a description of our methods in section II. Next, we investigate how PROV-DM can be applied to an Internet artwork, including how it can be implemented in a linked data knowledge management system. Section IV discusses key issues that came up in the modelling and implementation processes, some of which may also require future research. Finally, section V concludes the paper.

II. Methods and Methodology

A. *Application of the PROV model*

For this paper, we applied the PROV model in three steps: First, we traced the lifecycle of the artwork, based on archival research. Next, we translated this information in a PROV-DM application that illustrated the key components of provenance we considered necessary (section III.B). In the final step, we used PROV-O to map PROV-DM to RDF in Rhizome's Wikibase (section III.C).

PROV-DM consists of six components, of which we tested three for this paper (the PROV core structures) [4]. Component 1 (C1) describes the "entities" and "activities". An entity can be a variant of the artwork, and/or physical, digital or conceptual elements of the artwork. An activity is something that affects an entity within a certain time period. Component 2 (C2) describes "derivations". In our case this means how one variant of an artwork derives from or relates to another. Component 3 (C3) refers to agents and their responsibilities. An agent can be a person, as well as an organization or a piece of software. Using these components - C1, 2 and 3 - it becomes possible to pose and answer questions such as: "Who [agent] did something [activity] to this variant of this artwork [entity]?"; or: "How does this variant relate to other variant(s) [derivation]?"; etc.

B. *Interdisciplinary collaboration*

All findings presented in this paper are a result of interdisciplinary collaboration. At each stage, we made prototypes, which we reflected on and discussed, bringing our own areas of expertise. Through iterations, the outcomes were further refined. We adopted a practice-based research method, in which collaborative prototyping is a mode of enquiry [5].

C. *Case study approach*

Although this paper focuses on a single case study, we prepared models for various Internet artworks. The results were compared to further understand PROV-DM and, in particular, how to model the relationships between entities, agents and activities. Based on these initial tests, one artwork was selected and studied in more detail. All artworks tested are part of the ArtBase, one of the largest archives of Internet art that is accessible online since 1999. Maintained by Rhizome (USA), the ArtBase is also one of the few online art archives to support a functional linked data infrastructure, which can facilitate the implementation of a linked data provenance model.

The case study that is discussed in this paper is "untitled[scrollbars]" by artist Jan Robert Leegte. It was selected, because the ArtBase features several variants of the artwork, which offered the opportunity to model how the artwork evolved over time. Additionally, the record for this particular artwork features more detailed information about the accession and preservation of the work, compared to other records in the archive. Furthermore, unlike many other examples of Internet art, this artwork does not contain external media or data sources dependent on third-party services. Such media and services would require additional provenance research outside the scope of this initial study.

III. Case Study

A. *"untitled[scrollbars]"*

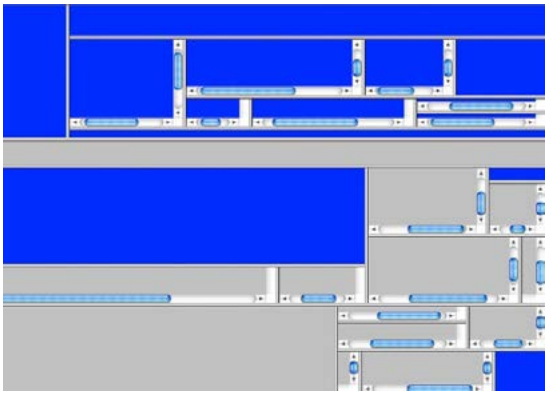


Figure 1. Jan Robert Leegte, "untitled[scrollbars]" (2000), web project, source: Rhizome ArtBase (Q2508).

The artist Jan Robert Leegte lives and works in Amsterdam, The Netherlands. He is part of a generation of artists, also known as the net art movement, who have been making art on the Web since the nineties. His early works rebuild basic, interactive elements of the (classic) Windows interface, such as buttons, window frames or scrollbars. The artwork "untitled[scrollbars]" was published online in 2000. Over time, the title of the artwork has changed from "untitled" to "untitled[scrollbars]" and "scrollbarcomposition", additionally the artwork can be encountered at various URLs. It can also be understood as part of a larger corpus of works, including a physical installation with the same title, consisting of a wooden structure and a projection of a Windows 98 or a Mac Aqua version of the scrollbar (2005, 2011). In 2001, "untitled[scrollbars]" was accessioned by the ArtBase, where an archived variant (in HTML), as well as a WARC (Web ARChived) file are preserved.

B. *Application of PROV-DM*

We developed a data model for the provenance of "untitled[scrollbars]" [Appendix A]. In this model, we applied three views on provenance that PROV supports: "data flow view", "responsibility view" and "process flow view", including the associated classes and properties from PROV-O [4].

The "data flow view" shows how one variant of an artwork derives from another. First, we identified all variants of the artwork that can be found in the ArtBase. These include two archived variants – in the custody of Rhizome, and two further URL's that pointed outside the ArtBase. In the provenance information we had available, it was not possible to find the derivation relationships for all of the variants. Building upon the data structure already in use in the ArtBase, we added one additional entity in our data model to represent the "artwork" as a general concept. PROV-DM provides support for modelling relationships between a general concept of an entity (the artwork) and its specific instantiations (the variants) with the properties *prov:alternateOf* and *prov:specializationOf* [4]. An entity that is a "specialization" of another shares all aspects of the latter. On the other hand, two "alternate" entities share some aspects of the same thing, but may also differ and may or may not overlap in time [4]. Since the way an Internet artwork evolves over time is unpredictable and variants can differ substantially

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

from one another, we found the latter, broader term more suitable in our case study.

The "responsibility view" involves assigning the agents responsible for different events in the artwork's lifecycle. We distinguished between the concepts of "attribution" and "association". In our model, we use attribution specifically in relation to the artist/creator of the work, whereas we use association to refer to any other contributors or the archivist/archival institution. In our case study, all variants of "untitled[scrollbars]" can be attributed to the same artist – Jan Robert Leegte. Therefore, we assigned the attribution property to the general concept of the artwork. In contrast, we assigned association agents to each specific variant. We found this to be a useful way of documenting custodial care, i.e. whether the variant is in the care of an artist (*:variant* > *prov:wasAssociatedWith* > *:artist*) or an archive (*:variant* > *prov:wasAssociatedWith* > *:archivist* > *prov:actedOnBehalfOf* > *:archivalOrganisation*). Association can be related to entities, or activities. In the latter case, it becomes part of the "process flow view".

In the "process flow view", we refined our data model by including "generation" activities (i.e. creation), and appended dates and locations.[1] The generation activities in our case study include "Cloning" and "Webrecorder capture". "Cloning" refers to creating a file directory copy of the artwork from the artist's server to Rhizome's server. "Webrecorder capture" refers to creating a WARC (Web ARChive) of the artwork, using Rhizome's tool Webrecorder. This activity is subject to the decisions of an archivist performing the capture. In order to document this agent's influence, we assign the association directly to the activity, rather than the variant. Furthermore, this association can be qualified (modelled as a *prov:qualifiedAssociation* in PROV-O) by additional properties, e.g. adding a "plan" to the activity, for example the archival instructions used during capture. A generation activity can also be qualified (modelled as a *prov:qualifiedGeneration)* by dates and times with the *prov:AtTime* property. When no particular activity of generation is assigned to a variant, the variant can still be dated using the *prov:generatedAtTime* property. Lastly, in addition to multiple times and activities of generation, the variants in our case study had different URL addresses. We used *prov:atLocation* to assign URL locations to each variant.

C. *Implementation in Rhizome's Wikibase*

1. *Wikibase and Wikidata*

Rhizome is one of the first cultural heritage organisations to use Wikibase as a collection management system for its archive, the ArtBase [6]. Wikibase is the open source software environment built to run Wikidata – a knowledge base of public domain structured data maintained by the Wikimedia foundation (WMF). Originally, the software infrastructure was not designed as a linked data system. Linked data capabilities were added later to serve the community needs for interoperability with existing linked data sets [7].

Wikidata's knowledge base follows RDF principles, and is organised in subject-predicate-object triples. These translate to item-property-value statements in terms of Wikidata syntax (e.g. artwork [item] > attributed to [property] > artist name [value]) [7]. Statements can have bibliographic references, too. This is how PROV-O is currently used in RDF data that can be exported from Wikidata. The *prov:wasDerivedFrom* property is used to link a bibliographic source to a particular statement. This application of PROV is insufficient in the case of Internet artworks and additional concepts and properties are needed.

Crucially for our use case, Wikibase can be deployed as a stand-alone instance, independent from Wikidata. While it still follows the RDF data modelling conventions of Wikidata, a Wikibase installation requires a custom configuration of concepts and properties. [6]. This is how Rhizome's Wikibase can adopt some PROV-DM concepts and PROV-O properties. Derivation and attribution, for instance, can easily be modelled as item-property-value triples. Owing to its legacy ties to Wikimedia

[1] Please note that the level of abstraction in this model is different compared to other digital preservation standards, such as PREMIS. This level of abstraction is concerned with artistic and historic integrity, not just technical integrity. We are modelling activities and actors involved in the creation, acquisition, or modification of an artwork variant, rather than tracking file system activities related to individual files such as checksum creation, etc.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

software, Wikibase is very good at representing data related to things (e.g. Wikipedia pages), which become "items" in its RDF statements. The concept of the item can be mapped to the PROV-DM concept of an entity. The concept of the qualified activity (or process) from PROV-O, however, is more challenging to represent in Wikibase.

2. Mapping concepts and properties

The possibility to map concepts across different concept schemas is an integral part of the design of linked data. The SKOS (Simple Knowledge Organisation System) data model was developed specifically to facilitate such linking across knowledge organization systems on the Web.[1] The SKOS mapping property (i.e. *skos:exactMatch*) is already being used in Wikidata as a way to match a concept from one standard schema to another[2]. Using *skos:exactMatch*, it is also possible to link properties and concepts from Rhizome's Wikibase to corresponding PROV-O properties and concepts [Table I].

The "data flow" and "responsibility views" can be fully represented in Wikibase via such mapping. The "process flow" view presents challenges with regards to the qualified relations, which in order to be represented in Wikibase may have to be broken down and simplified [Appendix B][3]. In some cases, it is possible to model PROV qualified processes with Wikibase "qualifiers" – these are sub-properties which can be added to statements, providing additional detail such as time periods, locations, etc. Qualifiers add flexibility to data modelling in Wikibase, however, this flexibility can make querying more difficult, because a user would need to know the exact structure of the data model in order to make a meaningful query [7].

While this may be considered a limitation of the system, it also provides an opportunity to model provenance data in statements that are both easier to present to end-users accessing the data via a graphical user interface (GUI), as well as easier to query by users who want to find the provenance of artworks without being experts in the particular data model used in the ArtBase.

TABLE I
Mapping PROV-O properties to properties in the ArtBase

Art Base Property	Property ID	skos:exact Match mapping	PROV view
variantOf	P56	prov:alternateOf	Data flow
derivedFrom	P102	prov:wasDerivedFrom	Data flow
artist	P29	prov:wasAttributedTo	Responsibility
collaborator	P120	prov:wasAssociatedWith	Responsibility
associated With	P118	prov:wasAssociatedWith	Responsibility
onBehalfOf	P119	prov:actedOnBehalfOf	Responsibility
generatedBy	P117	prov:wasGeneratedBy	Process flow
inception	P26	prov:generatedAtTime	Process flow
accessURL	P46	prov:atLocation	Process flow
startTime	P11	prov:startedAtTime	Process flow
endTime	P13	prov:endedAtTime	Process flow
archivalPlan	P121	prov:hadPlan	Process flow

IV. Discussion

A. *Linking variants to a general concept*

To gain insights into the provenance of an artwork, it is important that a query can retrieve all variants of the work. While a general concept is not required in PROV-DM, we used *prov:alternateOf* to connect all variants of the artwork to a general concept. This strategy is compatible with other cultural heritage and bibliographic semantic models, such as CIDOC-CRM and FRBRoo, where our concept of the "artwork" is equivalent to E28 Conceptual object (CIDOC) or F1Work (FRBR), and "variant" is equivalent to E73 Information Object (CIDOC) or F2 Expression (FRBR).[4][5] This compatibility offers the potential for information exchange between different cultural heritage collections.

[1] https://www.w3.org/TR/skos-reference/

[2] https://www.wikidata.org/wiki/Property_talk:P2888

[3] See the current record for "untitled[scrollbars]" in the ArtBase with partial PROV implementation: https://staging.catalog.rhizome.org/wiki/Item:Q2508

[4] http://www.cidoc-crm.org/

[5] https://www.ifla.org/publications/node/11240

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

B. *Modelling historical gaps*

Although closing knowledge gaps within the provenance of an artwork remains the goal, a 'complete' story can rarely be established. Considering the complexity of historical data, it is important to find ways to model gaps, inconsistencies and/or errors. PROV-DM offers the opportunity to state partial or incomplete provenance about an entity. For example, in the provenance for "untitled[scrollbars]" we added a date towards the general concept of the artwork (2000), but not to the variant that entered the ArtBase. This indicates that it is known that the artwork was created in 2000, but it is not precisely known when the particular variant accessioned by the ArtBase was created. Not only is it possible to leave out unknown information, PROV-DM also allows modelling in increasing levels of detail. For example, the provenance for "untitled[scrollbars]" indicates that the agent who accessioned the artwork is an archivist (their role). When provenance research reveals additional information, it is possible to further refine the provenance by adding, for example, a person's name (taking into consideration that their identity can be revealed) or more details about their actions (e.g. archival plans, etc).

C. *Accessing provenance data*

The considerations for implementing PROV-DM in a linked data art archive extend to how it will be accessed by end-users. Despite its limitations with regards to expressing qualified processes, the Wikibase system does provide a GUI, where the complexity of a graph database is made intelligible to end users [6]. While full integration of the PROV-O into Wikidata RDF expressions is not yet possible, if equivalent properties and concepts are accurately mapped, users will have the ability to query the ArtBase for PROV statements using a query service, such as the Wikidata Query Service. The formal terms for running federated queries across knowledge bases remain a matter of debate within the Wikidata community. A universal adoption of a standard mapping notation such as *skos:exactMatch* would improve the usability of the query service, particularly for users who would like to use it via a GUI. Until this adoption is implemented, users will need to first express the desired mapping in SPARQL and then formulate the particular query of interest.

V. Conclusion

In this paper we have presented a practical approach to expressing provenance for Internet art using PROV-DM. This approach facilitates the description of the lifecycle of the artwork, including any changes that were made over time and who was responsible for them. Additionally, it offers the opportunity to provide an overview of all the variants of an artwork, even when they are included in different collections and archives.

In future work, we will test PROV-DM against further case studies to include other types of provenance entities, agents and relationships. Further research is also needed to test the application of our proposal on a larger scale – e.g. entire collections. Another area for future research is how to fully integrate PROV-DM in Wikidata RDF expressions. For the time being, we have demonstrated alternative ways of working with PROV-O and PROV-DM in the ArtBase. We see future collaborations between digital preservation professionals, historians and the Wikimedia community as a key route to the wider adoption of PROV as a standard practice for preserving and presenting provenance of Internet art as linked data on the Web.

Acknowledgment

The authors would like to thank the following colleagues for their support and contributions to this research: Dr Annet Dekker, Anisa Hawes, Dr Katherine Thornton, Rhizome's Digital Preservation team, and Wikidata's development team at Wikimedia Deutschland. Lozana Rossenova would like to acknowledge that her PhD research is funded by AHRC (UK) with additional support from Rhizome.

References

[1] C. Li and S. Sugimoto, "Metadata-driven approach for keeping interpretability of digital objects through formal provenance description," in *Proceedings of the 14th International Conference on Digital Preservation iPRES 2017, Japan*, 2017.

[2] N. Jefferies, et al., "Contextual and provenance metadata in the Oxford Research Archive (ORA)', in *Proceedings of 9th Research Conference, MTSR 2015,* Manchester: Springer International Publishing, 2015.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

[3] K. de Wild, *Internet art and agency*, PhD diss. Dundee: University of Dundee, 2018, pp. 183-184.

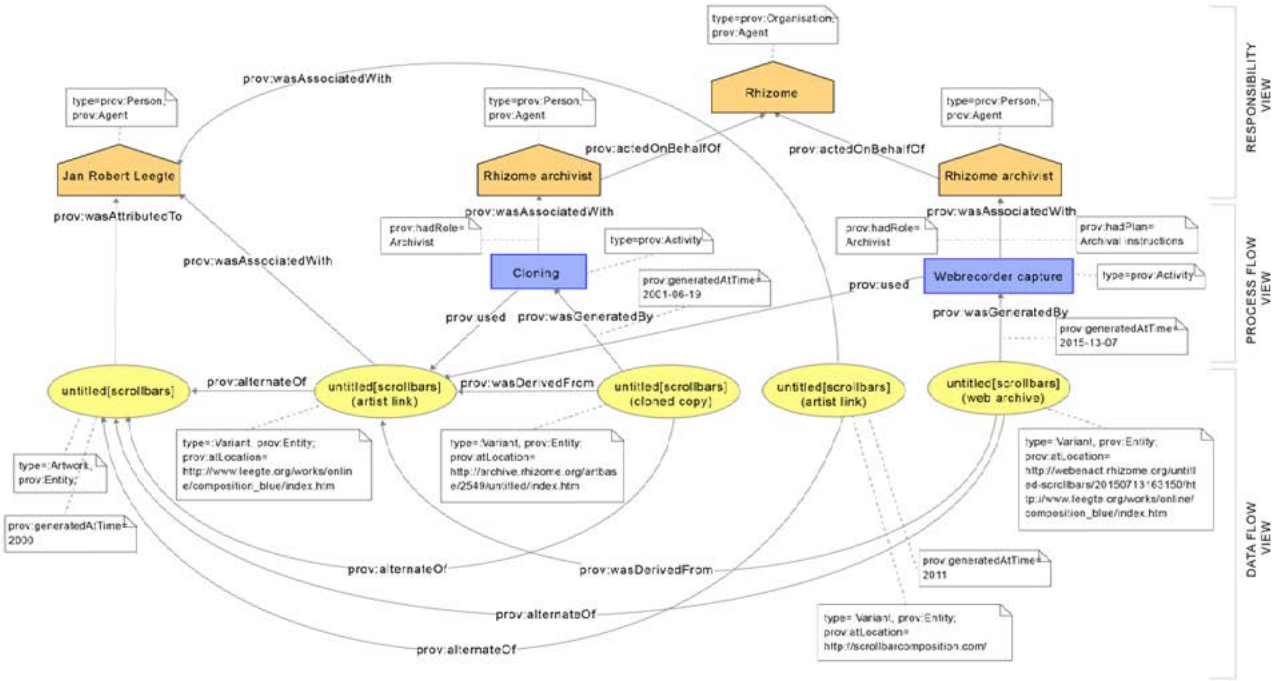
[4] L. Moreau and P. Groth, *Provenance: An introduction to PROV.* CA: Morgan & Claypool Publishers, 2013, pp. 3, 22, 33.

[5] F. Kräutli and S. Boyd Davis, "Digital humanities research through design," in *Proceedings of DH Early Career Conference,* King's College London, 2016, pp. 4.

[6] S. Fauconnier, "Many Faces of Wikibase: Rhizome's Archive of Born-Digital Art and Digital Preservation: Interview with Dragan Espenschied, Lyndsey Moulds and Lozana Rossenova", Wikimedia blog, 2018. https://wikimediafoundation.org/2018/09/06/rhizome-wikibase/

[7] K. Thornton, et al., "Modeling the Domain of Digital Preservation in Wikidata", in *Proceedings of the 14th International Conference on Digital Preservation iPRES 2017*, Kyoto, Japan, 2017, pp. 2, 5-6.

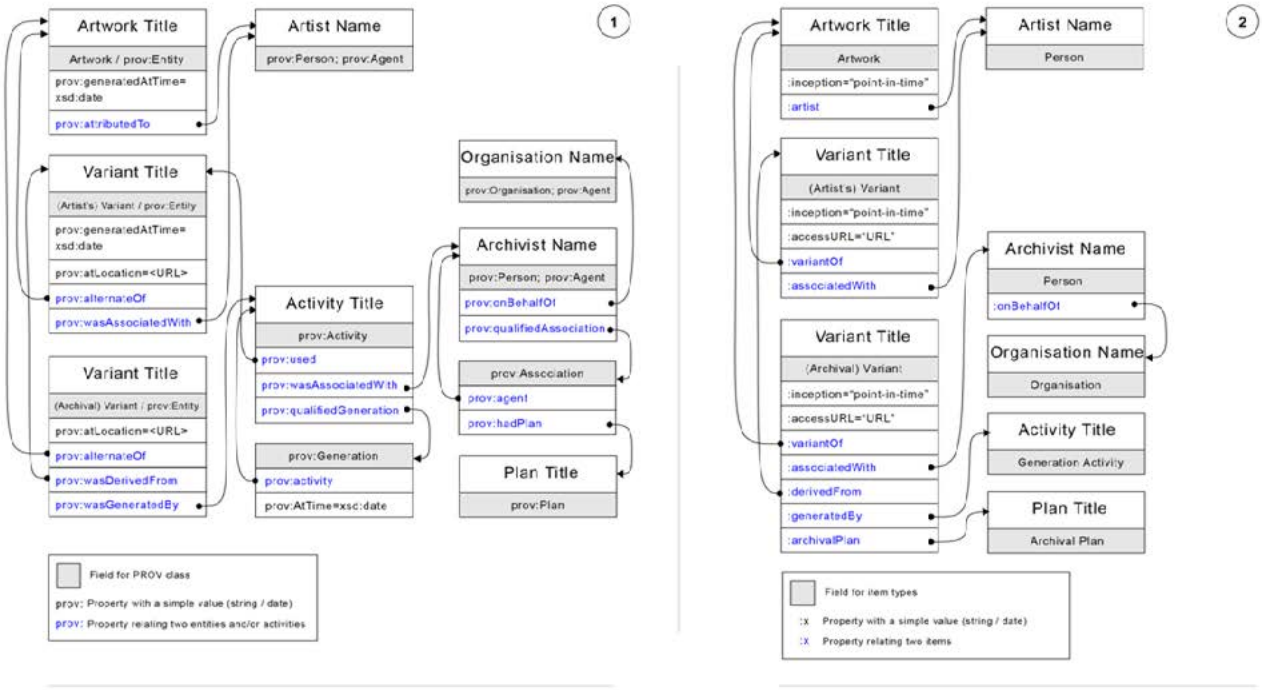
APPENDIX A



PROV-DM application for "untitled[scrollbars]". Visualisation follows the PROV Graph Layout Conventions, specified by the W3C recommendation: https://www.w3.org/2011/prov/wiki/Diagrams. Image can also be accessed in the Open Science Framework repository.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

APPENDIX B



A comparison between 1) the generalised PROV-DM application for the ArtBase developed in the case study, vs. 2) the proposed practical implementation of PROV concepts in Rhizome's Wikibase.
Image can also be accessed in the Open Science Framework repository.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.