From self-documentation to federated querying using Wikibase: a new topology to Media Art Archiving?

Raphael Tsz Kin Chau
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
tszkin.chau@epfl.ch

Abstract

Traditionally, domain datasets such as media art exhibition records and collection records are managed and published by archiving or research institutes. However, it is rare for an institute to compile or possess an entire exhibition history of an artist before his archive enters its custody. This paper provides an alternative paradigm in which artists may self-document and publish their creative history using the easy-to-use Linked Open Data suite Wikibase. This paper also argues that self-documentation is a two-for-one endeavour where artists can consolidate their CVs while the archiving and research communities will have access to a distributed archive.

Keywords

Artist as Archivist, Linked Open Data, Big Data, Graph Database, Backend for Media Art, Self-documentation, Community-driven Archiving, Distributed Knowledge Making

Introduction

This paper is built upon two previous works from the ISEA New Media Art Archiving Submit: *Preservation Begins at Creation: Integrating an Embedded Digital Archivist Within an Academic Media Art Program* (2020) and *Interconnecting Archives: Paving a Path Forward* (2022).

In the first work, the author argued that media artworks are unlikely to be preserved and documented by archiving professional before it enters the custody of an art museum. As a result, their associated digital artefacts might be kept in unarchival, obsoleting formats, challenging their future conservation and restoration. To counter this, the author promoted the concept of *preservation begins at creation*, in which media artists should be coached with archiving knowledge to perform preliminary preservation. [1] This paper extends this "Artist as Archivist" idea by advocating *self-documentation by artists*. Through the process of curating the Artist's CV in *Linked Open Data* (LOD), artists can benefit from the extensibility and reusability of the data created in their websites or art projects, while the archiving and research communities may be benefited by improving intergraph technologies, such as SPARQL federated query.

In the second work, the programmers from ISEA, SIGGRAPH, and FILE developed a system to connect people and events that are common in the respective online databases, enabling users to discover further information through hyperlinks. [2] This paper suggests that, with artists' data available as LOD, "Connecting Archives" can stretch deeper into artists' websites/endpoints.

This paper first discusses why self-documentation by the artist is desirable and necessary. Next, we will illustrate why we choose LOD technologies as our data backend and how LOD technologies benefit an artist and the archiving and research communities.

Motivation: Typical Use Cases

Data Management for the Artist Portfolio

The origin of this paper stemmed from the data management challenge of the Jeffrey Shaw Compendium, the portfolio, the retrospective project, and the future SPARQL endpoint of the Australian media artist. [3] Beginning in 2015, the Compendium was first built using a conventional content management system (CMS) as a website and the supporting installation during *Prix Ars Electronica 2015*. [4] Initially, it consists of webpages, images, videos, and credit information of about 100 selected works.

A major edit was made to the Compendium website right after *Prix Ars Electronica*. Exhibition records of the 100 works were reconstructed from archive documents, ephemera, and catalogues and are ready to be added to their respective page. Instead of adding the exhibition records as text, we determined to store the records in a relational database and generate the records via a function call. In such a way we can also generate the artist's CV using the same data source with another function, avoiding the problem of inconsistency and redundant work to synchronize information.

Expanding Application and Data Model

In 2019, the Compendium was presented at Enghien-les-Bains, France, with a different layout customized for the exhibition. [5] Instead of data export and import, we wanted the on-site instance to access the live data of the Compendium, in such a way that any update on the main Compendium is reflected on the installation instance. We migrated our data from the relational database to Google Sheets for two reasons. 1) Easier to add new columns / adjust schema to reflect new application needs, 2) Google Sheets is web-native and thus, when configure correctly, is more secure and reliable for retrieving data from remote locations using its API.

Figure 1. Jeffrey Shaw Compendium, *Passé Augmenté x Présent Augmenté*, Arts Center of Enghien-les-Bains, Enghien-les-Bains, France, 2019

Meanwhile, mass digitization and archival research of the Jeffrey Shaw Archive resulted in an immense growth of data and metadata. Apart from exhibition records, we now have records about lectures, conferences, museum collections, interviews, publications, and descriptive metadata for various kinds of media, such as video tapes, floppy disks, and computer tapes. Along with descriptive data, we also saw a growth in relational data. For example, a publication can be supporting evidence for an exhibition event or a citation for an artwork. Our existing Google Sheets data source is becoming more and more clumsy to adjust its schema to fulfil future data curation needs.

Supporting Media Art Project

As Lev Manovich pointed out, "the database becomes the centre of the creative process in the computer age." [6] The emergence of mass digitization, advancement of retrieval and interactive technology has brought further a new genre of interactive art that centers on the performance, navigation, and interpretation of archives. [7] I describe this trend as a movement from data(base)-driven media art to archive-

Figure 2. *The Infinite Line* illustratively demonstrates a strategy to "let the archive speaks". The installation gives visitors the opportunity to recombine the poetic ensemble of Edwin Thumboo, using the actual recital recording from the poet. [8]

driven media art. Developed from the archival principles of respecting provenance and evidence, this genre of art strives to "let the archive speaks". Consequently, new aesthetics and strategies emerged to "interface" the content and context of an archive collection.

Originally conceived to support the artist's retrospective exhibition, the database of the Compendium pre-version was first employed as an exhibition backend in 2013 as *Touch-table*. In this installation, the audience could browse the dual archive of Jeffrey Shaw and Hu Jieming, exploring image and video documentation of a chosen theme or keyword. [9] The limitation of the 2013 iteration is that the content is organized with a few handcrafted keywords and themes by the respective artist. The next iteration of the artist's retrospective project may include a richer dataset, stretching from temporal, spatial, thematic, people or sensory features (e.g. colour, contrast). In this respect, our existing database backend is not powerful enough to cater to the description and query requirements.

Figure 3. *Touch-table*, Jeffrey Shaw and Hu Jieming Twofold Exhibition, Chronus Art Center, Shanghai, China, 2014

Self-documentation in Linked Open Data

Requirement Specifications of the New Database

- NoSQL, graph, non-tabular database.
- Open Standard.
- Enrichable.
- Queryable.
- Web-native

NoSQL ("Not only SQL") refers to a collection of next-generation database systems developed from the big data challenge. Among various types of NoSQL databases, we are particularly interested in graph databases because of their flexibility and superiority in describing relationships. In a typical relational/tabular database (often also referred to as RDBMS, SQL database), the database engineer needs to define what attributes ("the columns") to be included to describe a collection ("the table"). Designed to handle transactional data, relational databases assume items/records ("the rows") to be homogeneous. This works well in an institutional context if the collection departments are divided according to medium. In our case, to describe the artworks of an artist who works in a diverse medium, attributes may not be consistent across these artworks. For example, "duration" is irrelevant in non-time-based works and "software engineer" is irrelevant in performance or architectural projects.

In a graph database, data is represented as interconnecting nodes and edges. Essentially, there is no structural imposition that an attribute must exist for a node, nor how many times such attribute is attached to a node. Moreover, when a new attribute is added to a node as an edge, no structural change to the database is needed, in contrast to an SQL database in which the table structure must be changed.

Another advantage of a graph database is spontaneous relationship creation. To create a relationship in a relational database, such as linking an artwork and its collaborators, a relationship table containing artwork and actor ID should be created. In a graph database, adding a relationship is as simple as linking an artwork node to an actor node via some "collaborator" edge. As a result, there is no structural change needed to add a new relationship. As a result, we can then implement new relationships without changing the database structure. Selecting a database with open standard is desirable. Disclosure, documentation, community adoption levels are the keys to the sustainability of the chosen database.

Enrichability refers particularly to the ability to connect, align and reconcile with public datasets to download new attributes. Data enrichment will enable us to step closer to curating an operable, programmable high-quality dataset. For example, reconciling a text-based location name to a unique ID on a geolocation database allows us to retrieve GPS coordinates for map visualization.

Queryablility refers to the capability to return a result set from a structured, general-purpose query like SQL. Web-native refers to a database package with an abstraction layer to provide an interface for building web applications, including websites, web apps, and custom exhibition applications. These two features are important to us as we aim at building websites and applications on the reworked database.

Linked Open Data

Many of the above requirements are satisfied by Linked Open Data (LOD) (also known as the Semantic Web), a concept conceived by Tim Berners-Lee. [10] First, LOD employs an RDF graph as its core data structure which is NoSQL and non-tabular. Created by the World Wide Web Consortium (W3C), RDF (Resource Description Framework) is a framework for representing information in the Web which has gained popularity in Library Information Science and Digital Humanities. [11] In RDF, data is represented by triples where each triple is a statement of three parts — the subject, the predicate, and the object — essentially replicating the node-edge schema of a graph database.

RDF requires triples to be represented by HTTP URIs (Hereinafter referred to as the 'URIs', for brevity).

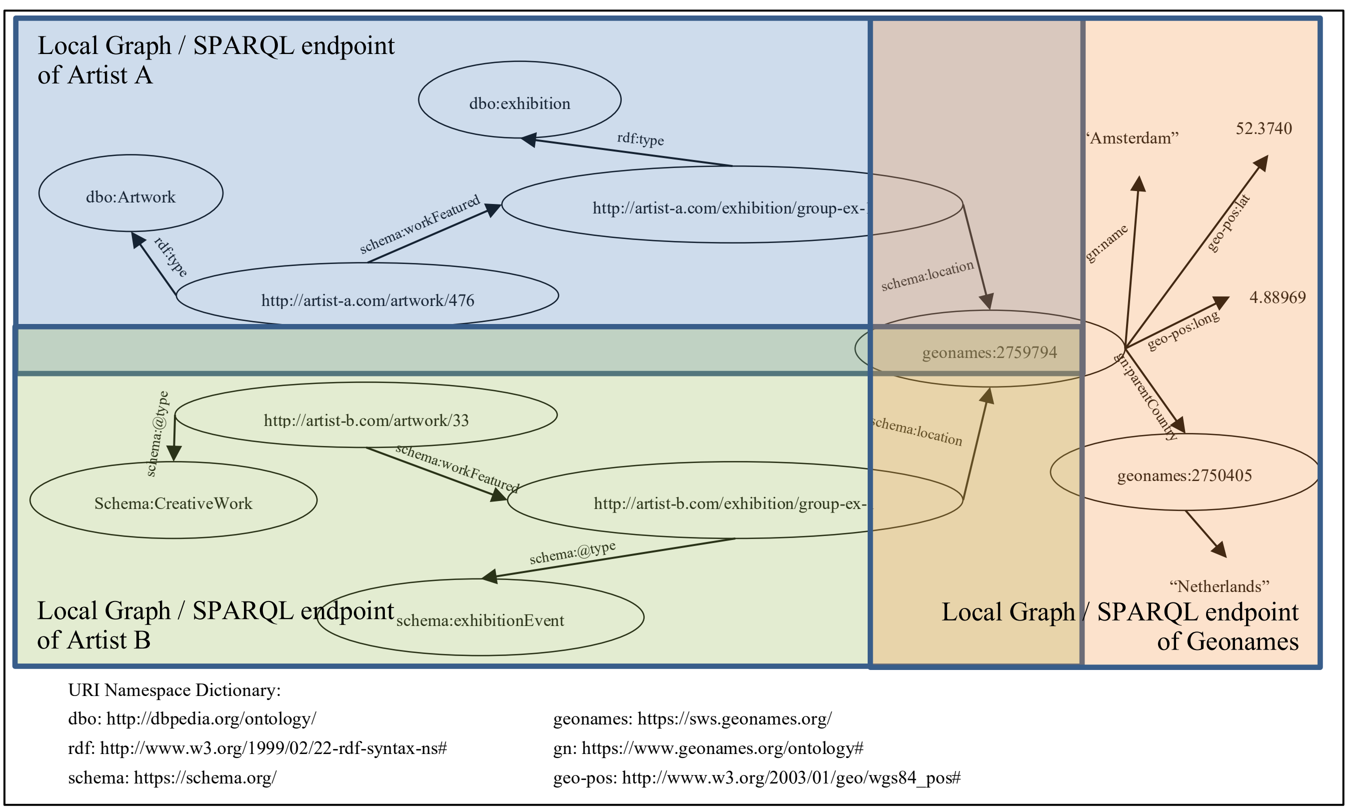

Figure 4. In the topology of federated querying, each data owner (Artist A, Artist B, Geonames) publishes and maintain their graph data.

Structurally, a URI comprises a domain and a path. For example, in Figure 4, “http://artist-a.com/artwork/476” contains the domain “artist-a.com” and the path “/artwork/476”. In the RDF schema, the domain represents namespaces while the path represents a resource in a namespace. Fundamentally, a resource can be any concept, or any “thing”, such as “artwork”, “exhibition”, “person”, or “event”, followed by a unique identifier, while namespace/domain partitions the web-scale database into separate graphs. Effectively, “artwork/476” of “artist-a” is different from “artwork/476” of “artist-b”. Furthermore, “artist-a” and “artist-b” may describe their data according to their own needs by choosing relevant attributes/predicates.

Predicates are also represented by URIs. There are many community projects (some examples are listed in Figure 4) providing terms to name predicates. One example is “https://schema.org/workFeatured” (shorthand schema.org:workFeatured) from the Schema.org project/namespace. Accessing this URI with an internet browser one can find its definition “A work featured in some event”. [12]

There are also many publicly available LOD datasets/projects published in RDF. For example, Geonames is a publicly available graph with hierarchical geographical information, such as regions, countries, GPS coordinates, and names in various languages. When two resources are referring to the same Geonames resource, we know that there is a connection. (Figure 4)

As one might notice, the second advantage of employing LOD is the possibility to offload certain data curation tasks to domain experts/communities, such as the example in Figure 4. Through linking the graph from Geonames, one’s data point is extended the latter’s upward branches.

Third, LOD comes with a powerful, SQL-like query language called SPARQL to interact with the RDF graph. For example, one may query“?wk” “schema:workFeatured” “http://artist-a.com/exhibition/group-ex-1”, which essentially means which artwork(s) is/are shown in group exhibition 1. Like SQL, SPARQL can be used to formulate complex queries, aggregate results, order results, and alter values. With SPARQL, we can migrate some features of the Compendium to a graph-based backend.

Forth, SPARQL results can be formatted as JSON and JSON-LD which is natively supported by web applications. And fifth, LOD is an open standard with interoperable schemas, this means if we transform our data into LOD we can also publish it as data apprehensible by a third-party *human* and *machine*.

SPARQL Endpoint, Federated querying and the new Typology of Local/Global Graph

One of the core innovations of LOD technologies is federated querying. Any service/domain exposed to provide a SPARQL query interface to its stored RDF graph is a SPARQL endpoint. Federated querying allows SPARQL endpoints to extend a query to other remote SPARQL endpoints. For example, in Figure 4, both artists A and B curated their exhibition record data in their local graph. When they expose their data through their SPARQL endpoints, people may combine their data to retrieve complex information such as common exhibition location patterns of the two artists over a period. People may also connect the data to the Geonames endpoint to obtain information about those locations, such as latitude and longitude for spatial visualization.

Federated querying thus opens up a new typology of local/global graph. Local graphs refer to the individual SPARQL endpoints published by artists, collectives, institutions, and research communities. Global graph refers to the joint representation of media art history by connecting local graphs. (Figure 7) There are three characteristics of this new typology. First, data autonomy. Within each local graph, data curators are free to create their data model, choose attributes (predicates) relevant to their needs, and enter data useful to them. Second, data interoperability. When curating data in local graphs, it is preferable to choose attributes and resources from existing, trustworthy projects than create a new one, if possible, to maximize the connection to other graphs. Third, distributed representation and provenance. Maintained by different individuals/communities, Each local graph may have its specific attention. For example, institution endpoints are more interested in their

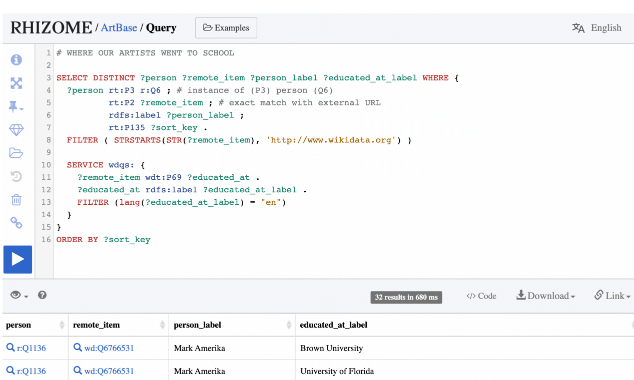

Figure 5. SPARQL endpoint of Rhizome ArtBase. The above example federated the graphs of Rhizome ArtBase and Wikidata. Here, the person’s name and the alma mater are derived from ArtBase and Wikidata respectively [13]

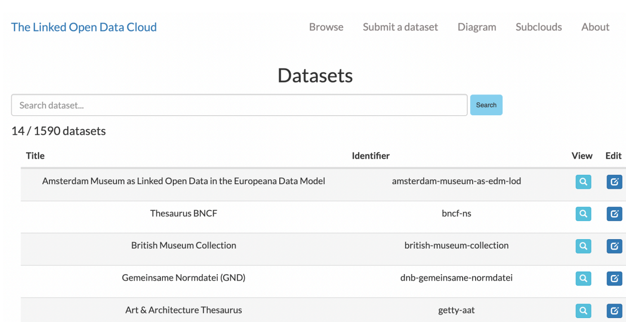

Figure 6. LOD dataset/endpoints published by museums. [14]

collection and exhibition, while an artist/collective endpoint is committed to managing data of their creative record.

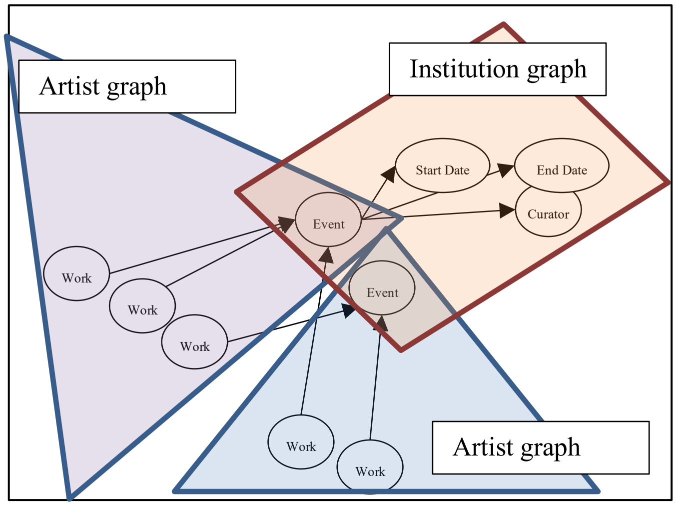

Figure 7. Concept of the Typology of Local/Global Graph. Assuming that the institute is publishing authoritative data of their exhibitions. The artist may offload the effort of documenting event data by connecting to/retrieving data from the institution graph. The global graph of media art history is the UNION of these graphs.

Our LOD Implementation

There are many open and proprietary implementations of LOD technologies. For a full stack implementation of LOD technologies, one has to select a triplestore (e.g. Apache Jena) engine for storing the RDF graph, a query engine (e.g. Fuseki) to implement SPARQL query, and an endpoint service (e.g. YASGUI) for structuring SPARQL result in human (tabular, visualization) and machine (RDF, XML, JSON) uses. Optionally, it is common to provide human-readable information in HTML for a URI. (HTML content negotiation) (e.g. LodView).

Wikibase: (almost) a Turnkey LOD Bundle

Implementing the above full stack requires experience in web server administration and backend development which needs many years to acquire. This paper proposes a ready-to-use alternative which requires a small amount of configuration. Developed by Wikimedia Deutschland, Wikibase is a full stack LOD bundle which is almost a turnkey LOD application. [15] It is also the data-storage backend for Wikidata. Relevant features and specifications of Wikibase are,

- Wikibase can represent data in an RDF graph on its Blazegraph backend.
- Wikibase includes a feature-rich SQARQL endpoint called Wikibase Query Service.
- Wikibase Query Service is accessible in mainstream programming languages such as Python, JavaScript, PHP, and Ruby.
- Wikibase format data URI in HTML to provide human-readable information.
- Wikibase provides a richer data model than RDF which is useful to provide context to archival records.
- Wikibase can be deployed as a self-contained docker service.
- Wikibase provides an intuitive WebGUI to enter statements.
- Wikibase edit can be done in batch on Openrefine.
- A vast general-purpose graph from Wikidata is ready to be used as an enrichment source.

Wikibase as LOD backend

The Wikibase LOD implementation is slightly different from the canonical LOD. However, it is fully compatible with the RDF triples and SPARQL queries. The notable differences are:

- Wikibase, an RDF triple is known as a “statement”.
- In canonical LOD, URIs are normally human-recognizable. (e.g. http://example.com/person/bob) In Wikibase, URIs are represented by numeric identifiers. (e.g. http://www.wikidata.org/entity/Q325)
- Resources are called “entities” / “items” in Wikibase and are represented with Q-ids.
- Predicates are called “properties” in Wikibase and are represented with P-ids.
- Canonical LOD consists of an Ontology, which is the data model for describing data of a knowledge domain for achieving interoperability. Wikibase lacks this domain ontology.

Data Model Related Work

Today, there are 10,724 properties on Wikidata. It is impossible to achieve the effect of federated queries if each data owner “speak their own” language. Therefore, to harness the full potential of Wikibase’s complete stack, researches are conducted to experiment and adapt domain ontology usage of Wikibase properties and entities. [16] In media art, Lozana Rossenova piloted using Wikibase as a data platform for collection management and archival description. The research led to the publication of the Wikibase-powered online archive of Internet Art (Rhizome ArtBase) and the publication of the ArtBase data model. [17]

Our Data Model

At the present stage, we planned to transform artworks, exhibition records and publications data from the Compendium as a pilot test. We experiment to see to what extent the ArtBase data model can be used in artists’ self-documentations.

Artwork

An experiment was done on transforming the textual information of the artwork Legible City to LOD. [18]

Property label	Wikidata ID	Type	Target Item	Remarks
instance of	P31	Item	Artwork	
inception	P571	Point in time		
author	P170	Item	Person/ Collective/ Institution	ArtBase uses “artist” as the property label.
collaborator	P767	Item	Person/ Collective/ Institution	We extended the uses of collaborator with the qualifier: role.
*role	P3831	Item	Occupation/ Profession	
description	X	X	X	
imageURL	X	URL		ArtBase uses “images” (P18) to store the thumbnail image
videoURL	X	URL		ArtBase uses “video” (P10) to store related video
exhibition history	P608	Item	Exhibition	
*Jeffrey Shaw Compendium ID				external ID Our legacy artwork id
*described at URL	P973	URL		Link to the text version of the artwork
subject of	P1343	Item		publication

Table 1. Artwork data model, Jeffrey Shaw Compendium Wikibase

The ArtBase artwork model is suitable for describing artworks. We can easily add/adjust properties for a better representation of our data without changing the schema from ArtBase. We extended the collaborator property from ArtBase using Wikibase’s qualifier feature.

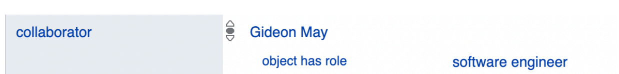

Figure 8. Describing collaborator role in Wikibase.

Exhibition

An experiment was done on transforming the exhibition records of Legible City to LOD. [19]

Property label	Wikidata ID	Type	Target Item	Remarks
instance of	P31	Item	Exhibition	
inception	P571	Point in time		Precision to Year
*start time	P580	Point in time		Precision to Datetime
*end time	P582	Point in time		Precision to Datetime
*location	P198	Item	Venue/ Museum/ Gallery	
*located in the administrative territorial entity	P202	Item	City/ administrative territorial entity	
*coordinate location	P625	Geographic coordinates		Data enriched from Wikidata
*country	Q183	Item	Country/ Region	
*Jeffrey Shaw Compendium Event ID			external ID	Our legacy event id
Same as	P2888			Link to the same Wikidata entity

Table 2. Exhibition data model, Jeffrey Shaw Compendium Wikibase

Due to the nature of Internet Art, the ArtBase exhibition model is not very comprehensive. Thanks to the graph data structure, we can easily add the missing spatiotemporal properties to the data model. Considering that our research has closely documented the source of the exhibition records, Wikibase’s statement reference model enables us to also publish these supporting materials.

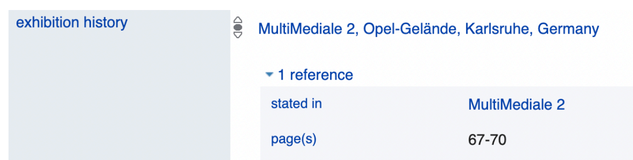

Figure 9. Describing reference source for an exhibition record.

Publication

An experiment was done on transforming one of the publication records of Legible City to LOD.

Property label	Wikidata ID	Type	Target Item	Remarks
instance of	P31	Item	Publication	
publication date	P14	Point in time		
described at URL	P973	URL		Link to the publication on OCLC WorldCat
OCLC control number	P596	external ID		For data reconciling and enrichment
ISBN-10	P597	external ID		For data reconciling and enrichment
editor	P316	Item	Person/ Collective/ Institution	
publisher	P348	Item	Collective/ Institution	
Jeffrey Shaw Compendium Publication ID		external ID		

Table 3. Publication data model, Jeffrey Shaw Compendium Wikibase

The primary objective of the publication model is to support the generation of the "writing" pages of the Compendium website and the Artist's CV. With Wikibase, we can further associate a publication as a citation for an artwork or as a reference source for an exhibition record.

Wikibase: Values to Artist

Returning to the origin of the problem. How does self-documentation using Wikibase benefit the artist and the archiving communities?

An Extensible DB for Non-expert User

Wikibase comes with a friendly user interface for adding data manually. It is particularly suitable for artists in their early careers where they can start documenting their career entirely with the WebGUI without too much data wrangling.

Figure 10. Adding statement using Wikibase WebGUI

Language	Label	Description	Also known as
English	Legible City	interactive art installation where the visitor rides a stationary bicycle through a simulated representation of a city	
German	Die lesbare Stadt	No description defined	
French	The Legible City	No description defined	The Legible city (…
Japanese	読みやすい都市	No description defined	
Dutch	De leesbare stad	No description defined	

Figure 11. Wikibase natively support multilingual labels

Data Enrichment

One of the foremost use cases of LOD is the reconciliation and enrichment of one's data against large public datasets. For example, our exhibition data contain lots of city and country names and we wish to link these names to the vast geographic databases on Wikidata. Through reconciliation with Wikidata, we can retrieve additional data such as coordinate location useful for further visualization.

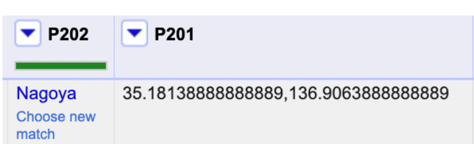

Figure 12. Enriching data using Openrefine. Cell on the left: our data reconciled with WikiData (entity Q11751). Cell on the right: coordinate data retrieved from WikiData based on Q11751.

Query Engine for Application/Artwork Uses

A SPARQL endpoint is a general-purpose query platform that can be used to retrieve data for application uses. There are libraries in mainstream programming languages such as Python, JavaScript, Ruby, and PHP, making it suitable to function as the backend for artists' websites and data sources for artists' retrospective installations.

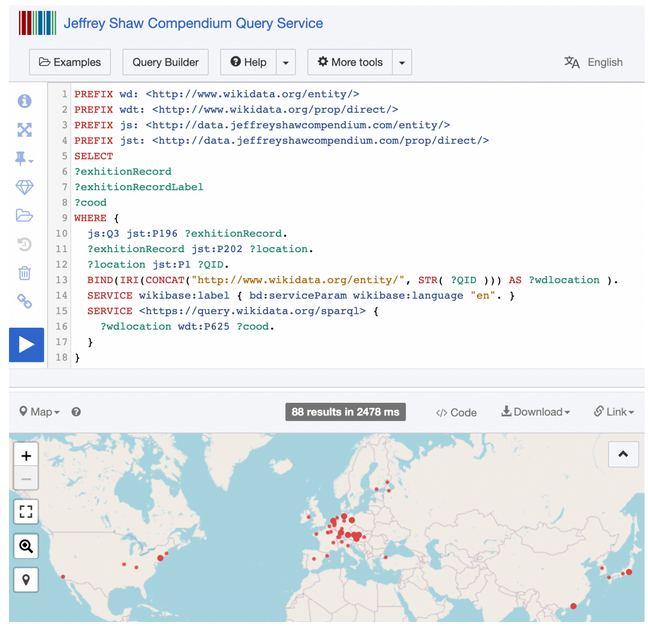

Figure 13. Federated query (Jeffrey Shaw Compendium + Wikidata) visualizing the exhibition location of Legible City. (mark size corresponds to the number of exhibitions)

Final Words

At the beginning of this paper, we envision a new paradigm of media art archiving in that the compilation of an artist's chronology is accomplished spontaneously by the artist's *self-documentation* in LOD. The experiment of this paper has identified Wikibase as a user-friendly and extensive database for *self-documentation*. This paper also described typical use cases and values *self-documentation* can bring to an artist. However, up to now, there is no artist's self-documentation endpoint listed on LOD-Cloud. [20] Therefore, we are yet to see how we will use federated querying to analyze the "Big Data" / global graph of media art history.

References

[1] Devon Mordell, "Preservation Begins at Creation: Integrating an Embedded Digital Archivist Within an Academic Media Art Program," (paper based on a talk presented online at Montreal, Canada, October, 2020). *ISEA Summit On New Media Archiving.*

[2] Alexa Mahajan, Luis Wilson and Dalton Lopes Martins, "Interconnecting Archives: Paving a Path Forward," (paper based on

a talk presented at MACBA - Convent dels Àngels, Barcelona, Catalonia, Spain, June, 2022). *ISEA Summit On New Media Archiving*.
[3] "Jeffrey Shaw Compendium," https://www.jeffreyshawcompendium.com/
[4] *Prix Ars Electronica 2015*, OK im OÖ Kulturquartier, Linz, Austria, 2015.
[5] *Passé Augmenté x Présent Augmenté*, Arts Center of Enghien-les-Bains, Enghien-les-B ains, France, 2019.
[6] Lev Manovich, *The Language of New Media* (Cambridge, Mass.: MIT Press), 227.
[7] An example project employing an artist's archive as the primary material, see Tiago Martins, Christa Sommerer, Laurent Mignonneau, "AR[t]chive – Augmented Reality Experience for a Digital Art Archive," (paper based on a talk presented at MACBA - Convent dels Àngels, Barcelona, Catalonia, Spain, June, 2022). *ISEA Summit On New Media Archiving*.
[8] *Jeffrey Shaw*, Sarah Kenderdine, Edwin Nadason Thumboo, The Infinite Line, 2014.
[9] *Touch-table*, Jeffrey Shaw and Hu Jieming Twofold Exhibition, Chronus Art Center, Shanghai, China, 2014.
[10] Tim Berners-Lee, "Linked Data (2014)", Design Issues, accessed January 15, 2023, https://www.w3.org/DesignIssues/LinkedData
[11] W3C, "RDF Schema 1.1," accessed January 15, 2023, https://www.w3.org/TR/rdf-schema/
[12] Schema.org, "workFeatured," accessed March 9, 2023, https://schema.org/workFeatured/
[13] This federated query pulled the "educated at" property from Wikidata to extend the biographical data of the artists featured in Rhizome ArtBase. "Rhizome ArtBase Query: WHERE OUR ARTISTS WENT TO SCHOOL", Rhizome ArtBase, accessed January 15, 2023, https://tinyurl.com/yzrsehs8
[14] John P. McCrae, "Museum Datasets", The Linked Open Data Cloud, accessed January 15, 2023, https://lod-cloud.net/datasets?search=museum
[15] Wikibase, Wikimedia Deutschland, 2012.
[16] Joonas Kesäniemi, Mikko Koho, and Eero Hyvönen, "Using Wikibase for Managing Cultural Heritage Linked Open Data Based on CIDOC CRM," in *New Trends in Database and Information Systems*, eds. Silvia Chiusano, Tania Cerquitelli, Robert Wrembel, Kjetil Nørvåg, Barbara Catania, Genoveva Vargas-Solar, and Ester Zumpano (Cham: Springer International Publishing, 2022), 542–49.
[17] Lozana Rossenova Mehandzhiyska, "ArtBase Redesign Data Models (2020)," ArtBase Redesign Documentation, accessed January 15, 2023, https://sites.rhizome.org/artbase-re-design/datamodels.html
[18][19] "Legible City," Jeffrey Shaw Compendium, accessed January 15, 2023, https://www.jeffreyshawcompendium.com/portfolio/legible-city/
[20] John P. McCrae, "The Linked Open Data Cloud," https://lod-cloud.net/

Bibliography

Berners-Lee, Tim, "Linked Data (2014)", Design Issues, accessed January 15, 2023, https://www.w3.org/DesignIssues/LinkedData

DBpedia, "The DBpedia Ontology," accessed January 15, 2023, http://dbpedia.org/ontology/

Kesäniemi, Joonas, Mikko Koho, and Eero Hyvönen, "Using Wikibase for Managing Cultural Heritage Linked Open Data Based on CIDOC CRM," in *New Trends in Database and Information Systems*, eds. Silvia Chiusano, Tania Cerquitelli, Robert Wrembel, Kjetil Nørvåg, Barbara Catania, Genoveva Vargas-Solar, and Ester Zumpano (Cham: Springer International Publishing, 2022), 542–49.

Mahajan, Alexa, Luis Wilson and Dalton Lopes Martins, "Interconnecting Archives: Paving a Path Forward," (paper based on a talk presented at MACBA - Convent dels Àngels, Barcelona, Catalonia, Spain, June, 2022). *ISEA Summit On New Media Archiving*.

Manovich, Lev, *The Language of New Media* (Cambridge, Mass.: MIT Press)

Martins, Tiago, Christa Sommerer, Laurent Mignonneau, "AR[t]chive – Augmented Reality Experience for a Digital Art Archive," (paper based on a talk presented at MACBA - Convent dels Àngels, Barcelona, Catalonia, Spain, June, 2022). *ISEA Summit On New Media Archiving*.

Mordell, Devon, "Preservation Begins at Creation: Integrating an Embedded Digital Archivist Within an Academic Media Art Program," (paper based on a talk presented online at Montreal, Canada, October, 2020). *ISEA Summit On New Media Archiving*.

Mehandzhiyska, Lozana Rossenova, "ArtBase Redesign Data Models (2020)," ArtBase Redesign Documentation, accessed January 15, 2023, https://sites.rhizome.org/artbase-re-design/datamodels.html

Mehandzhiyska, Lozana Rossenova, "Model–Database–Interface: A study of the redesign of the ArtBase, and the role of user agency in born-digital archives," (Ph.D. diss., London South Bank University School of Art and Creative Industries, 2021.)

QA Company, the, "WikibaseSync", 2022, https://github.com/the-qa-company/WikibaseSync/

Schema.org, "Schema.org Project," accessed January 15, 2023, http://schema.org/

W3C, "RDF Schema 1.1," accessed January 15, 2023, https://www.w3.org/TR/rdf-schema/

W3C, "SPARQL 1.1 Overview," accessed January 15, 2023, https://www.w3.org/TR/sparql11-overview/

W3C Semantic Web Interest Group, "Basic Geo (WGS84 lat/long) Vocabulary," accessed January 15, 2023, https://www.w3.org/2003/01/geo/

Wick Marc, “GeoNames,” accessed January 15, 2023, https://www.geonames.org/

Wikimedia Deutschland, “Wikibase/DataModel,” accessed January 15, 2023, https://www.mediawiki.org/wiki/Wikibase/DataModel

Author Biography

Raphael Tsz Kin Chau is a PhD student in Digital Humanities from École polytechnique fédérale de Lausanne, Switzerland. Chau is interested in the exhibition, experience and retrieval of Big Cultural data. Chau obtained his master’s degree in Digital Humanities from KU Leuven, Belgium. His master’s thesis reflected on the data publication strategies of memory institutions in the context of the Europeana aggregator project. He has developed the physical archive and the digital infrastructure of the Jeffrey Shaw Archive and Jeffrey Shaw Compendium which is exhibited in Austria, France, Hong Kong and Switzerland. Chau is currently working on the Digital Murten Panorama project (DIAGRAM) collaborated with Stiftung für das Panorama der Schlacht bei Murten.