alj

50/1 2025

Smithsonian Libraries & Archives: BIBFRAME linked data experiment with the Share Family technology

Jackie Shieh

The semantic web promises interlinks of cultural heritage objects, both original and digital, of which the Smithsonian Institution (SI) holds plenty in artifacts, books, audio/video, etc. At the core, structured data in RDF is likely the one member of data infrastructure that carries the major lifting for interlinking. Parsing millions of bibliographic data mainly locked in MARC21 into the BIBFRAME/RDF syntax is not an insignificant endeavour for the library community. The Smithsonian Libraries & Archives, a network of specialized libraries (including art and design), is not an exception. It became part of the Share Family to utilize its technology so as to increase user resource discovery and the library staff's ability to resource curations. Working collaboratively with scores of research, academic and national libraries, and the Share Family team, the Smithsonian Libraries & Archives experimented with BIBFRAME, a Library of Congress RDF-based ontology, which has benefited greatly from the Share-VDE semantic technology and its system design. From such efforts, the Smithsonian Libraries & Archives looks to move away from the web of documents into a web of data. Fulfilling SI's strategic vision of availing collections through digital solutions for the increase and diffusion of human knowledge.

Structured data and the Semantic Web

Sir Tim Berners-Lee unveiled his Web 3 thesis in 1998.[1] Since then, system designers, software engineers and information service professionals have worked tirelessly designing software, and formulating policies and best practices concerning data models, structures and system infrastructure.

Structured data in RDF (Resource Description Framework)[2] syntax is the base of the Linked Open Data (LOD) ecosystem. This is an entity-relationship data model in a triple form of descriptive statements that give the context and meaning to entity data facilitating machine negotiation, manipulation, and making connections. RDF establishes meaningful relationships between entities that can be set apart from the original framework, forming new connections, encouraging remix and reuse of data. For library data, this structured data syntax liberates its long-locked infrastructure to freely up- or down-link to nodes that share the same HTTP identifiers.

URIs (Uniform Resource Identifiers), the axioms of Web architecture[3] carries the global uniqueness that unambiguously identifies the resource it points to. It shows clarity about the resource the URI represents, conveying and meeting the surrounding entity, and is easier for user agents (e.g., browsers) to understand when processing and manipulating information. On the open web, URIs bridge RDF data and the Web. Connecting the abstract world with concrete resources across the web, URIs glue the Semantic Web.

1. Tim Berners-Lee, *[Semantic Web (Web 3.0)] Design Issues,* accessed March 4, 2024, https://www.w3.org/DesignIssues/Overview.html; "*The Next Web,*" filmed February 2009, video, 16:03, https://www.ted.com/talks/tim_berners_lee_on_the_next_web.
2. Berners-Lee, "Relational Databases on the Semantic Web," *Design Issues.*
3. Berners-Lee, "Axioms of Web Architecture," *Design Issues.*

 © Smithsonian Institution, 2025. This is a work of the US Government and is not subject to copyright protection within the United States. Published by Cambridge University Press on behalf of ARLIS doi:10.1017/alj.2025.5

https://doi.org/10.1017/alj.2025.5 Published online by Cambridge University Press

Library bibliographic description

Library bibliographic descriptions are document-based syntax catalogue records that have a long tradition rooted in over century-old principles and practices.[4] Description coded in MARC format was developed in the mid-1960s. In response to Web technology, and with the ability to quickly engage and align with open and new Web standards, the library community (under the leadership of Library of Congress and many academic and national libraries) in 2012 embarked on the journey of migrating library MARC21 data to bibliographic framework (BIBFRAME)[5] to describe library collections. BIBFRAME, an RDF-based ontology, aims to be the replacement of the current bibliographic encoding structure, MARC21. For over a decade, BIBFRAME has had a relatively small number of adoptions throughout the world. The momentum of implementations picked up when the Library of Congress announced that its cataloguing staff would incorporate BIBFRAME in its production workflow by the year-end of 2024, along with its implementation of a new library service platform, FOLIO. In September 2023, during the BIBFRAME Workshop in Europe, several library service providers shared their roadmap of incorporating BIBFRAME in their respective library service platforms for cataloguing workflow.[6]

BIBFRAME is a three-tier abstraction with core categories (Work, Instance, and Item) and three classes for Agent, Subject and Event that relate to the core categories.[7] This data model leverages the RDF modelling practice of classes and properties and identifies entities, attributes and relationships, between and among them.

Libraries, archives, and museums hold vast collections of resources and data that have remained mostly hidden. In some cases, many are yet to be discovered even in individual institutional catalogues. With a linked open data framework, the BIBFRAME data model is likely a better tool to facilitate data exchange and reuse with other widely used Web ontologies revealing local holdings to a wider audience. Hence this encourages greater engagement between the collections of cultural heritage institutions and their users. BIBFRAME, though born in the library, is designed to be agnostic to data content types

4. Cataloguing principles and practices trace its roots for almost 150 years–from Charles Cutter's *Object of Cataloguing (1876)* to the Paris Principles (1961) and its revision, the IFLA International Cataloguing Principles in 2009 and the various revisions of Anglo-American Cataloguing Rules (1967, 1978, 1988-2002).

5. Bibliographic Framework Initiative, last modified January 31, 2024, https://www.loc.gov/bibframe/.

6. LSP providers (EBSCO FOLIO, Index Data FOLIO, ExLibris, OCLC) shared BIBFRAME roadmap in the BIBFRAME Workshop in Europe, 2023, https://www.bfwe.eu/brussels_2023.

7. Overview of the BIBFRAME 2.0 Model, accessed March 11, 2024, https://www.loc.gov/bibframe/docs/bibframe2-model.html.

The Smithsonian Libraries & Archives

Smithsonian Libraries & Archives (SLA), a unit of the Smithsonian Institution, is a network of 21 specialized research libraries (including art/design-centric collections) and an archives unit that preserves, organizes, and manages documents and records of the Smithsonian since its inception in 1846. SLA must be equipped to describe the data types from 21 museums, specialized research centres, institutes, and laboratories. In addition, SLA serves as the central agency that creates and manages library metadata and oversees authority records for subjects and names of persons, organizations, concepts, events, and places.

SLA's library catalogue holds close to 2 million items and over 900 thousand bibliographic records. MARC21 has been the library's default catalogue data format. Recognizing the inevitable linked data momentum within the research libraries community, SLA began preparing its MARC21 data for the future. Its linked data journey commenced in 2019 with two different ecosystems addressing the needs for authority control as well as bibliographic description. Expanding an already familiar wiki infrastructure and staff knowledge of Wikipedia to Wikidata, SLA staff experimented with authority data in RDF, the W3C syntax for linked data utilizing the wiki platform, known as Wikidata. SLA staff also moved forward with BIBFRAME experimentation, the natural candidate for renovating bibliographic description. The hope was to energize SLA's next data generation journey and expand the library collection data horizon.

The first step, on the bibliographic description front, was to upgrade descriptive standards to RDA (Resource Description & Access), adding FAST headings and embedding HTTP URIs to headings from authority and entity identity services. Followed by a gradual move of authority data to a collaborative platform – seeking to modernize library operations and services by utilizing Web-based solutions. Experimentation began to manage name identity entities on a wiki platform (Wikidata) and a local installation, Wikibase. In the process, staff knowledge and

https://doi.org/10.1017/alj.2025.5 Published online by Cambridge University Press

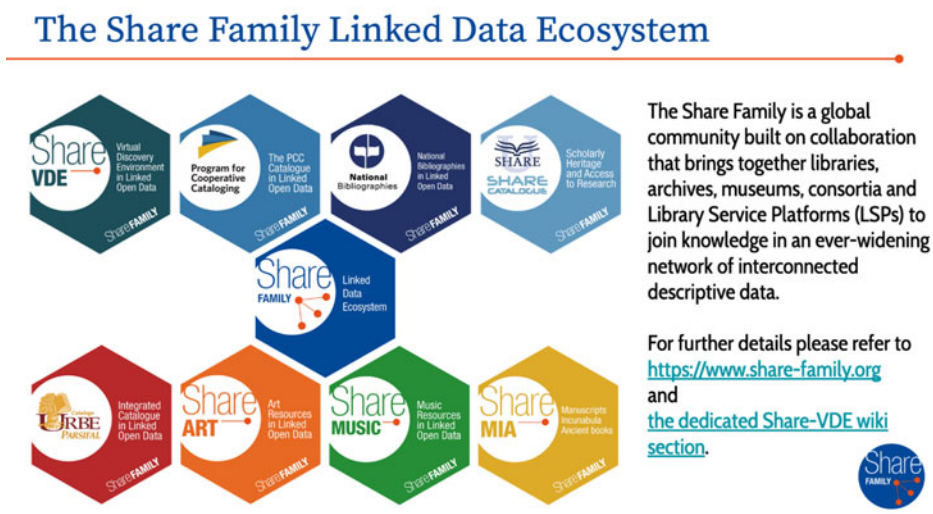


Fig. 1. The Share Family clusters.

skills with linked data was broadened by moving authority control to an identity entity management service.[8]

The Share Family technology offered a real possibility for an institution such as SLA and its bibliographic linked data experiments. The Share Family's sustainable model fits so very well with this Chinese idiom, '*sending charcoal in the snow*' — 'building a community with a shared future for mankind'. For a federal agency like SLA that has long-suffered insufficient personnel and resources but whose systems must perform well for production and services to its constituents, this model seemed ideal. As the diverse member communities in the Share Family share their respective initiatives and expertise, all members benefit from one another. More members also help to lower the cost to participate. Such engagement allows SLA to quickly embrace a more modern and dynamic approach to represent bibliographic description and to seamlessly interconnect resources across domains and disciplines in the Smithsonian museums and other resources on the Web.

8. A series of library blogs by staff who shared their insights on Wikidata as a viable alternative to provide authority work on a non-library traditional platform. https://blog.library.si.edu/blog/tag/wikidata-series/.

The Share Family Initiative

The Share Virtual Discovery Environment (Share-VDE) originated as a research project in October 2016, developing a partnership with the LD4L (Linked Data for Libraries) Mellon funded project, spearheaded by the Stanford Library and major academic libraries from the Program for Cooperative Cataloguing (PCC) community.[9] With an overall goal of enabling and making connections for cultural heritage collections beyond the library community by enriching MARC records with URIs, optimizing the linked data principles that are based on the Library of Congress BIBFRAME ontology, and then publishing to the open Web. Over the years, Share-VDE, a member-driven initiative, expanded to be the Share Family Initiative that includes the following projects—the Share Catalogue, the National Bibliographies, the Catalogue Urbe Parsifal, Share-Art, Share-MIA (manuscripts, incunabula, ancient books), the PCC, etc. (Fig. 1) The Share Family initiative adds cultural heritage institutions and domains to work together for the dissemination and discovery of human knowledge—history, art, and science. The core principles of the Share Family Initiative are: cooperative, sustainable, interoperable, open, flexible, and authoritative.[10]

9. Linked Data for Libraries (LD4L) Gateway, last modified October 6, 2020, https://wiki.lyrasis.org/display/ld4lgw.

The Share-VDE ontology

The Share-VDE ontology (Fig. 2) is a BIBFRAME ontology (Fig. 3) with an extension for a higher abstraction entity known as Opus which eventually form the three core classes: svde:Opus (Fig. 4), svde:OpusType and svde:Work to support the discovery functionality and the Share Family search systems.

The svde:OpusType (Fig. 5) individuals are related to the svde:Opus using the object property svde:hasOpusType.

10. Share Family: Linked Data Ecosystem, accessed March 13, 2024, https://www.share-family.org/ The Share Family: Cooperation and Interconnection for a Sustainable Linked Open Data Ecosystem, accessed June 1, 2024, https://www.cni.org/topics/metadata/share-family-cooperation-and-interconnection-for-a-sustainable-linked-open-data-ecosystem.

https://doi.org/10.1017/alj.2025.5 Published online by Cambridge University Press

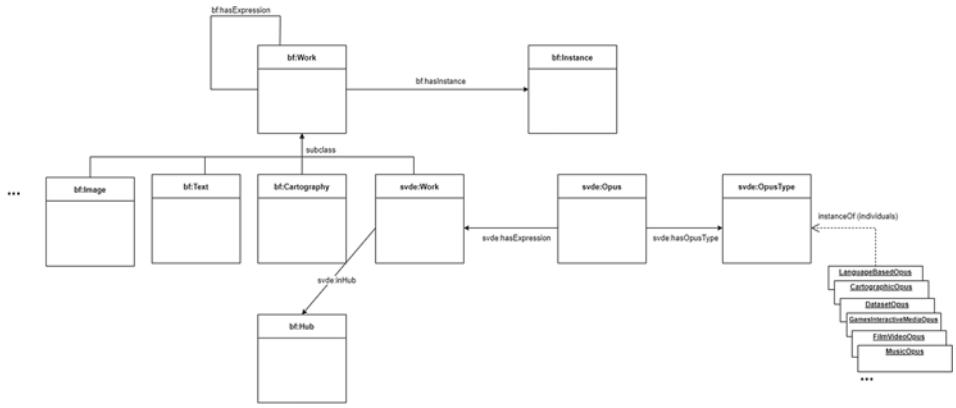


Fig. 2. SVDE Ontology diagram.

```
<!-- https://svde.org/ontology -->

<Ontology rdf:about="https://svde.org/ontology/">
        <versionIRI rdf:resource="https://svde.org/ontology/svde-1-0-0"/>
        <import rdf:resource="https://id.loc.gov/ontologies/bibframe.rdf"/>
        <cc:license
rdf:resource="https://creativecommons.org/publicdomain/zero/1.0/"/>
        <terms:publisher
rdf:resource="http://id.loc.gov/vocabulary/organizations/itfic"/>
        <skos:comment rdf:datatype="http://www.w3.org/2000/01/rdf-schema#Literal">
Share-VDE (SVDE) is a library-driven initiative which brings together the
bibliographic catalogues and authority files of a community of libraries in a shared
discovery environment based on linked data.</skos:comment>
        <skos:definition rdf:datatype="http://www.w3.org/2000/01/rdf-
schema#Literal">The Share-VDE ontology is an extension of the Library of Congress
BIBFRAME vocabulary.</skos:definition>
</Ontology>
```

Fig. 3. SVDE Ontology.

```
<!-- https://svde.org/ontology/Opus -->

<Class rdf:about="https://svde.org/ontology/Opus">
<disjointWith rdf:resource="https://svde.org/ontology/Work"/>
<terms:relation rdf:resource=" http://id.loc.gov/ontologies/bibframe/Hub"/>
    <rdfs:label>Opus</rdfs:label>
    <skos:definition>The svde:Opus is a distinct conceptual outcome of
artistic or intellectual activity. The highest level of abstraction in Share-VDE, an
Opus is an entity that permits the grouping of works that are considered functional
or near equivalents. The Opus is defined by a constellation of elements that form
the shared content of works and provides a grouping for svde:Work
entities.</skos:definition>
    <skos:note>The Opus is not the same as the bf:Hub.</skos:note>
    <skos:scopeNote>The Opus may be a piece of art, literature, music,
a scientific result, or a creation within some other artistic or intellectual
domain.</skos:scopeNote>
    <svde:closeMatch rdf:resource="http://iflastandards.info/ns/lrm/lrmer/E2"/>
    <svde:closeMatch rdf:resource="http://rdaregistry.info/Elements/c/C10001"/>
</Class>
```

Fig. 4. SVDE Ontology: Opus.

```
<!-- https://svde.org/ontology/OpusType -->

<Class rdf:about="https://svde.org/ontology/OpusType">
    <rdfs:label>OpusType</rdfs:label>
    <skos:definition>Individuals of the OpusType class support
identification of Opus categories.</skos:definition>
</Class>
```

Fig. 5. SVDE Ontology: OpusType.

https://doi.org/10.1017/alj.2025.5 Published online by Cambridge University Press

```
<!-- https://svde.org/ontology/Work -->

<Class rdf:about="https://svde.org/ontology/Work">
    <rdfs:subClassOf
rdf:resource="http://id.loc.gov/ontologies/bibframe/Work"/>
    <rdfs:label>Work</rdfs:label>
    <skos:definition>The svde:Work is defined by a constellation of
elements representing the specific intellectual or artistic form that
an Opus takes each time it is "realised." Individuals of the class
svde:Work hold an Opus entity identity.</skos:definition>
    <svde:closeMatch
rdf:resource="http://iflastandards.info/ns/lrm/lrmer/E3"/>
    <svde:closeMatch
rdf:resource="http://rdaregistry.info/Elements/c/C10006"/>
</Class>
```

Fig. 6. SVDE Ontology: Work.

```
<rdf:RDF xmlns="http://www.w3.org/2002/07/owl#" xmlns:rdf="http://www.w3.
org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/
rdf-schema#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:terms
="http://purl.org/dc/terms/">
<Ontology rdf:about="https://svde.org/ontology">
<imports rdf:resource="http://www.rdaregistry.info/xml/Elements/c.xml"/>
<imports rdf:resource="http://www.rdaregistry.info/xml/Elements/e.xml"/>
<imports rdf:resource="http://www.rdaregistry.info/xml/Elements/u.xml"/>
<imports rdf:resource="http://www.rdaregistry.info/xml/Elements/w.xml"/>
<imports rdf:resource="http://iflastandards.info/ns/lrm/lrmer.xml"/>
<imports rdf:resource="https://id.loc.gov/ontologies/bibframe.rdf"/>
```

Fig. 7. SVDE Ontology: imports function and other ontologies.

```
<AnnotationProperty rdf:about="https://svde.org/ontology/closeMatch">
<rdfs:label>close match to</rdfs:label>
<skos:definition>Refers to a semantically similar entity (
typically class or property) in another ontology or scheme.
</skos:definition>
<skos:scopeNote>The svde:closeMatch is different than the skos:
closeMatch in the sense that the annotation property is
designed for annotation of classes whereas the skos:closeMatch
provides a mechanism for relating class individuals. The svde
:closeMatch annotation does not directly connect any of the
classes that are the focus of annotation.</skos:scopeNote>
<rdfs:subPropertyOf rdf:resource="http://purl.org/dc/terms/
relation"/>
</AnnotationProperty>
```

Fig. 8. SVDE Ontology: annotation property: svde:closeMatch.

The svde:Work (Fig. 6) is defined by a constellation of elements representing the specific intellectual or artistic form that an Opus takes each time it is 'realized.' Individuals of the class svde:Work hold an Opus entity identity.

The SVDE ontology adheres to the design of data reuse principle reducing redundancy and ambiguity. With the BIBFRAME vocabulary as a base, relationships are added as imports among the SVDE classes and the existing linked data vocabularies, RDA Work and Expression classes, RDA unconstrained elements, and the IFLA LRM conceptual model. (Fig. 7)

This design achieves interoperability among bibliographic models, such as IFLA LRM and other widely used ontologies, including Getty, Wikidata, etc. enabling the discovery portal capability to adapt ontologies from, but not limited to, GLAM domains. In addition, an annotation property, svde:closeMatch (Fig. 8), gives definition and scope relating to other ontologies.[11]

The Share Family technology

As a member of the Share Family linked data ecosystem, the SLA's MARC21 data, like its peers', went through a suite of Linked Open Data (LOD) service platform processes—data mapping, URI enhancement, heading reconciliation to create

11. SVDE ontology model examples reformatted from Jim Hahn and Tiziana Possemato's article, "The Share-VDE ontology: a BIBFRAME extension for linked data discovery" September 10, 2023. https://zenodo.org/doi/10.5281/zenodo.8332350.

https://doi.org/10.1017/alj.2025.5 Published online by Cambridge University Press

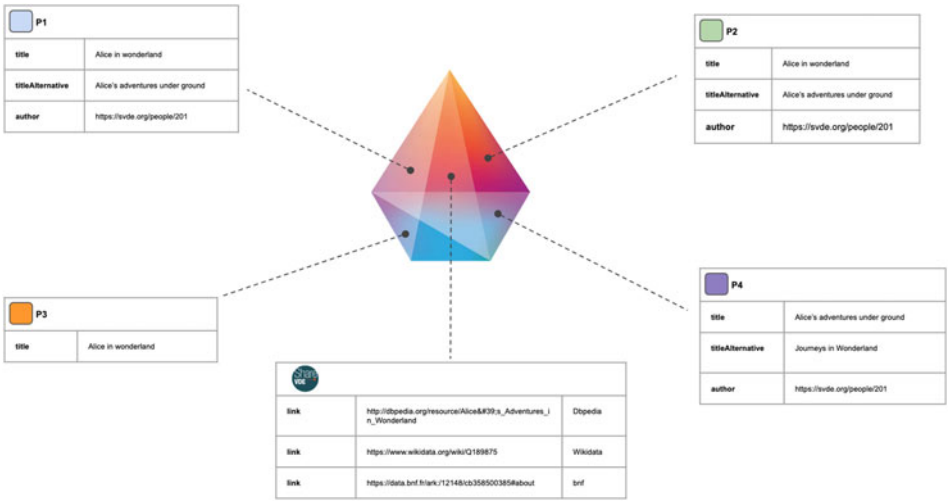


Fig. 9. SVDE technology: PRISM.

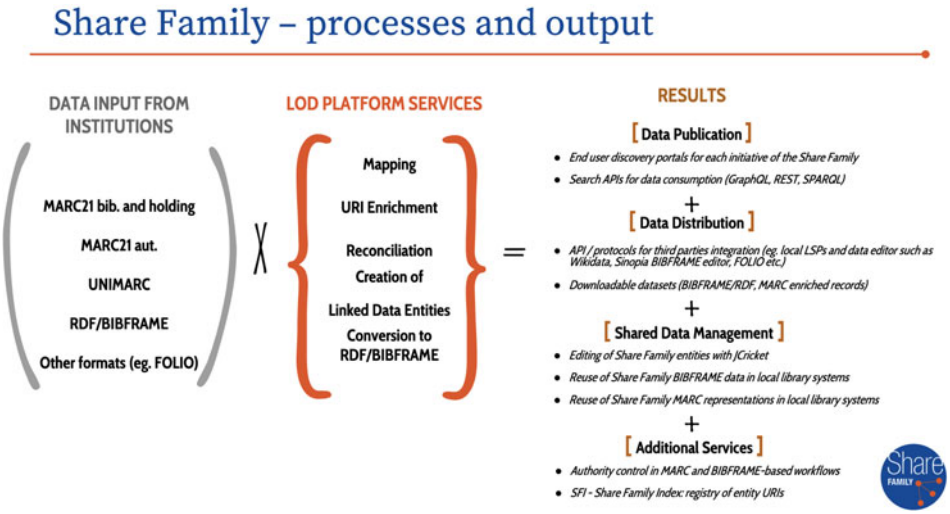


Fig. 10. Share Family data flow.

linked data entities and transformation to BIBFRAME RDF data format. Clusterization takes place simultaneously as member libraries' data is being ingested into the LOD platform. The data from each contributing institution contributes to the building and enriching of one or more clusters in the SVDE database. A cluster is a PRISM where each face represents a given provenance and maintains a link to the entities from which it originated (Fig. 9).

After their data is published, member institutions are able to search in the discovery portals that are supported by a wide range of search APIs (based on GraphQL (SPARQL, StructQL, SVDEQL, TermsQL, REST) and tools (PostgreSQL, SolrCloud, Hadoop) for data consumption and management. Data can also be integrated into third parties such as local library service platforms (LSPs), other data repositories, like Wikidata, FOLIO, etc. The Share Family entity management editor, JCricket, provides collaborative workflow opportunities for members to create and maintain entities that can be reused in BIBFRAME or MARC formats in members' local environments. The Share Family Index (SFI) orchestrates the indexing of entities across the various projects of the Share Family (Fig. 10).

Together facing a brighter future for information services

Standing on the Share Family technology, the SLA's descriptive journey toward a linked data ecosystem has laid a solid foundation for connecting library data to the

https://doi.org/10.1017/alj.2025.5 Published online by Cambridge University Press

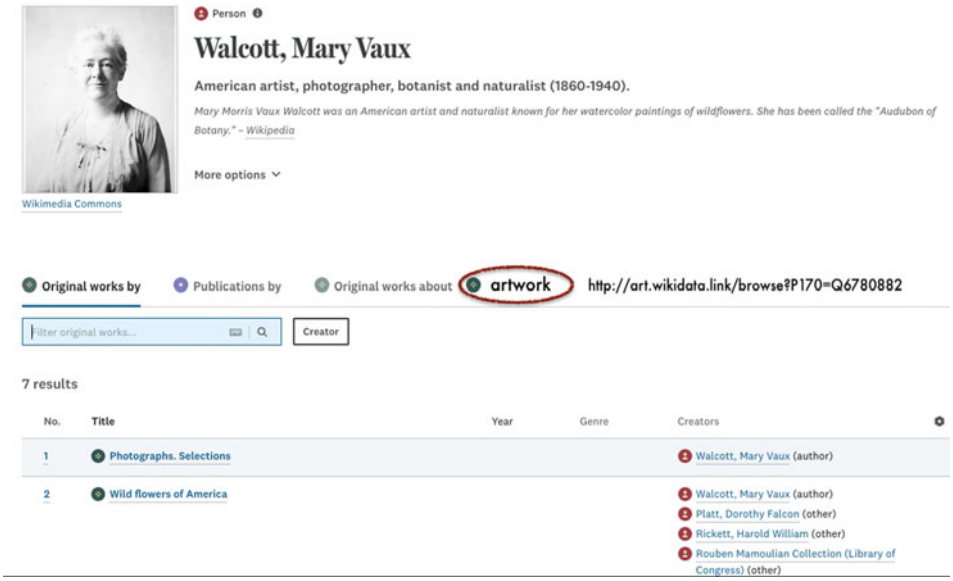


Fig. 11. Smithsonian SVDE Discovery portal: showing Mary Vaux Walcott, including links to her works and artworks on WikiCommons.

Institution's 18 million resources within and beyond the Smithsonian.[12] Imagine the full advantage availed from the semantic technology, coupled with the widely deployed internationally accepted structured data syntax and the technology platform – institutional data can be easily remixed to represent various needs from different disciplines and visualization preferences (e.g., data graphs, charts, map, timelines, etc.). It's a look into a brighter future of interconnected resources! Users may effortlessly journey from an artist to artworks, to archival records, to specimens depicted in an artwork, to the drawings of field reports, and digital surrogates of them. For example, Mary Vaux Walcott, a famed botanist and gifted illustrator for her scientific pursuits, was also the wife of the fourth Smithsonian Secretary, Charles Doolittle Walcott. For some records Mary was known as Mrs. C.D. Walcott, in her research publications and accompanying illustrative works she was known as M.V. Walcott. With HTTP URIs via interconnections to/from a variety of linked data endpoints, it straightforwardly shows her accomplishments, and the network of her family and profession. Shown here is the Smithsonian SVDE discovery portal, http://smithsonian.svde.org with potential connections to freely available artworks by Mary Vaux Walcott (Fig. 11).

On the horizon is data visualized in graphs from Wikidata, in the language of a user's preference, such as this one in Chinese scripts, https://w.wiki/A77p. It is an exciting prospect for the Smithsonian resources! And many others as well! As for this team, fulfilling the institutional vision—availing SI collections through digital solutions for the increase and diffusion of human knowledge is drawing ever closer!

12. Jackie Shieh. "Library Descriptive Data Unleashed in SVDE: Making Data More Meaningful," *Unbound (blog),* September 14, 2022, https://blog.library.si.edu/blog/2022/09/14/library-descriptive-data-unleashed-in-svde-making-data-more-meaningful/.

Jackie Shieh
Descriptive Data Management Librarian
Smithsonian Libraries and Archives
10th St. & Constitution Ave. NW
Washington, D.C.
United States
Email: shiehj@si.edu

https://doi.org/10.1017/alj.2025.5 Published online by Cambridge University Press