A Shared Approach to BIBFRAME Implementation

The SHARE-Virtual Discovery Environment

With many thanks to the SVDE Transformation Council, Working Groups, Tiziana Possemato and Michele Casalini for sharing diagrams and slides from previous presentations.

Hosted by ALCTS, Association for Library Collections and Technical Services
What We’ll Talk About Today

- Overview of SHARE-Virtual Discovery Environment (SVDE) project
- History and development of SVDE
- Provide an overview of processes, workflow, and models
- Contextualize SVDE in relation to other linked data initiatives
- Provide an overview of the SVDE Transformation Council
- Outline the work of SVDE working groups
  - Work Identifier Working Group
  - Cluster Knowledge Base Editor Working Group
  - Authority and Identifier Management Working Group
  - Discovery UX & UI Working Group
- Next steps and further analysis
What is SHARE-VDE?

Introduction and Project Overview
SHARE-VDE is a community-driven initiative to implement linked data. While the aim is a more general focus on transitioning traditional GLAM institution data thus far the project focus has been on moving from MARC to BIBFRAME.

The process enriches library data with additional information and relationships, previously unexpressed with MARC, and converts bibliographic and authority data in linked data.

A virtual discovery platform with a four-layered adaptation of the BIBFRAME data model was developed to provide a linked data discovery option.
The main areas of the SHARE-VDE project:

- Enrichment of MARC records with URIs
- Conversion from MARC to RDF using the BIBFRAME vocabulary (and other additional ontologies as needed)
- Data publication according to the BIBFRAME data model
- Batch/automated data updating procedures
- Batch/automated data dissemination to libraries
- Progressive implementation of further use cases in the priority order defined by the community
SHARE-VDE is a collaborative endeavour, based on the requirements and perceptions of libraries, developed by:

- Casalini Libri, provider of bibliographic and authority data as member of the Program for Cooperative Cataloguing

- @Cult, provider of ILS, Discovery tools and Semantic web solutions for the cultural heritage sector

- with input and active participation from an international group of 22 research libraries and influenced by the vision of the LD4P initiative

The collaborative initiative is steered by the library community
Participating Institutions

SVDE Full Members
Duke University
New York University
Stanford University
University of Alberta – NEOS consortium
University of Chicago
University of Michigan at Ann Arbor
University of Pennsylvania
Yale University
The National Library of Norway

LD4P Cohort
Cornell University
Frick Art Reference Library
Harry Ransom Center
Harvard University
National Library of Medicine
Northwestern University
Princeton University
UC Davis
UC San Diego
University Colorado at Boulder
University of Minnesota
University of Texas A&M
University of Washington

And with the cooperation of
The Library of Congress

SHARE-VDE Phases

**Phase 1**
October 2016 – January 2017
1985 and 2015 imprint titles; 2,249,397 bib-records and 3,601,327 auth-records.

**Phase 2**
March 2017 – May 2018
The entire catalogues for all resource types; 94,378,728 bib-records and 24,150,238 auth-records.

**Phase 3**
Production environment: January 2019-
In progress.

Linked

D1. The catalogue of each library converted into BIBFRAME 2.0 format

D2. The Cluster Knowledge Base in RDF format

D3. The dataset converted in BIBFRAME 2.0 with external URIs included

D4. The MARC21 version of D3

Entities are reconciled in the dataset and linked to SHARE-VDE project URIs of D2 for identification.

Common for all institutions as it includes data from all of the participants. Entities in D2 are enriched with URIs from external sources. All variant forms are included.

This dataset includes a certain number of relationships already present in the knowledge base. Works autonomously from D2.

It includes all of the institution's records enriched with URIs.
Overview of SVDE Workflow

OliSuite: manual process → Marc enriched/URIs → Entity detection

Entity detection:
- Enrichment
- Reconciliation/Cluster

Database of relationships → Knowledge base of clusters

Knowledge base of clusters → RDF/Bibframe dataset

OliSuite: manual process:
- MARC
- XML

Authify:
- Similarity’s score
- Dump db
- APIs

External sources

Possemato, T. (2018). The mapping and conversion workflow from MARC to BIBFRAME. European BIBFRAME Workshop, Firenze, Italy.
The current SHARE-VDE entity model
The future of the data model

Future model version 1

Future model version 2

Project Overlap: The role of SVDE in library communities
“It is time to move beyond knowledge and skills related to linked data at a theoretical level and into implementation. Building on the PCC’s strong tradition of providing training for metadata creators, active experimentation and piloting of linked data practices will help inform policy decisions, training, and operationalizing such practices. As we move to a culture of greater data sharing, it is crucial to extend our community, both by engaging a more diverse range of members in the work of the PCC and by collaborating with vendors, open source communities, and others.” (Program for Cooperative Cataloguing, 2018)
BIBFRAME

- Initiative started by the Library of Congress and community partners and collaborators in 2011
- Provides a foundation for the future of bibliographic description on and of the web
- Based on linked data principles and standards
- Goes beyond “replacing” MARC
  - different model for expressing and connecting bibliographic data
- https://www.loc.gov/bibframe
- http://bibframe.org/
BIBFRAME 2.0

The SVDE Community will be making proposals to the Library of Congress for the evolution of BIBFRAME, providing a mechanism to structure feedback from multiple institutions.

- Three core levels of abstraction
  - Work
  - Instance
  - Item

- Additional key concepts
  - Agents
  - Subjects
  - Events

- Consists of RDF classes and properties
  - members of a class share certain characteristics and may have subclasses
  - properties describe characteristics of resources as well as relationships among resources
Linked Data for Production has been focusing on:

- Developing standards, guidelines, and infrastructure to communally produce metadata as linked open data
- Developing end-to-end workflows to create linked open data in a technical services production environment
- Extending the BIBFRAME ontology to describe library resources in specialized domains and formats
- Engaging the broader library community to ensure a sustainable and extensible environment
Linked Data for Production: Pathway to Implementation (LD4P Phase 2)

A collaborative project among four institutions (Cornell, Harvard, Stanford, and the University of Iowa School of Library and Information Science) and the Library of Congress and the Program for Cooperative Cataloging (PCC), this phase of LD4P has seven broad goals:

1. The creation of a continuously fed pool of linked data expressed in BIBFRAME-based application profiles.
2. The development of an expanded cohort of libraries (the LD4P Cohort) capable of the creation and reuse of linked data through the creation of a cloud-based sandbox editing environment.
3. The development of policies, techniques and workflows for the automated enhancement of MARC data with identifiers to make its conversion to linked data as clean as possible.
4. The development of policies, techniques, and workflows for the creation and reuse of linked data and its supporting identifiers as libraries’ core metadata.
5. Better integration of library metadata and identifiers with the Web through collaboration with Wikidata.
6. The enhancement of a widely-adopted library discovery environment (Blacklight) with linked-data based discovery techniques.
7. The orchestration of continued community collaboration through the development of an organizational framework called LD4.
Sinopia -- “The underdrawing for the new world of linked data in libraries”
SVDE Transformation Council

«The SHARE-VDE Transformation Council's role is to provide insight and analysis of the MARC to BIBFRAME transformation to make recommendations for improvements based on member library data analysis, and project documentation. Initial recommendations are based on Phase 2 deliverables, but the work of the team will be ongoing into the foreseeable future.»

There are 4 sub-committees focusing on specific areas:

• Work Identification Working Group
• Authority/Identifier Management Services Working Group
• Cluster Knowledge Base Interaction/Editor Working Group
• User experience/User Interface Working Group
Conversion Review Process

1. Reviewed project partner reports from Yale, University of Chicago, University of Alberta, and Duke University to gauge the scope and commonalities of issues discovered
2. Analyzed conversion documentation provided by Casalini Libri, @Cult, and the LC Conversion Specifications
3. Analyzed data through the LC BIBFRAME Comparison Tool and
4. Viewed MARC and BIBFRAME/RDF data side-by-side using a special, project specific comparison tool created by Tim Thompson for this use

The SVDE-TC attempted to be as systematic as possible, but our time frame did not allow for a comprehensive review of every MARC element through conversion to BIBFRAME. Additionally, the subgroup that was to undertake review of MARC authorities did not form in time to give input. The SVDE-TC began meeting on August 15, 2018; the initial recommendations were finalized on November 30, 2018.
SVDE Transformation Council Recommendations to Improve the MARC to BIBFRAME Conversion

General Recommendations

1. BIBFRAME properties should be used instead of other ontologies whenever possible in order to promote consistency and sharing of data.

1. SVDE conversion specifications, based on LC Conversion Specifications, need to be published to support community development and ongoing analysis. This documentation will be accompanied by a white paper.

1. The library community should be prepared for the need to iterate the conversion process in 1-2 years to support implementation strategies. SVDE-TC should continue communicate with library community members.

1. SVDE developers should coordinate with LD4P developers to ensure that supported vocabularies for Sinopia are utilized by SVDE.
The SVDE Work ID Working Group

At the LD4 Workshop at Stanford University in May of 2018 conversations identified the need for a group to work on refining the creation of work identifiers for the SHARE Virtual Discovery Environment project (SVDE). In the months following members of the Work ID Working Group (WIDWG) were identified from within stakeholder institutions and based on relevant expertise. Representation included Casalini Libri, @Cult, George Washington University, Library of Congress, NLM, Ohio State University, PCC, Stanford University, UCDavis, University of Alberta Libraries, and University of Chicago.

Charge:

1. Review the SHARE work clustering outline and submit feedback on potential improvements or optimizations
2. Review the use of primary resource identifiers in the SHARE VDE data set and provide feedback as appropriate
3. Engage with the PCC to identify and/or develop best practices for use of these identifiers in BIBFRAME and MARC data
Grouping under a single work title of the many publication titles in the catalogue for *Cimento dell’armonia e dell’inventione* brings together different publications/resources present in different catalogues.

http://share-vde.org/sharevde/searchTitles?t_cluster_id=11287

Possemato, T. (2018). The mapping and conversion workflow from MARC to BIBFRAME. European BIBFRAME Workshop, Firenze, Italy.
The Challenge - Lots of Work in a Short Time

LC bf: example slide

```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmns:bf="http://id.loc.gov/ontologies/bibframe/"
xmns:bflc="http://id.loc.gov/ontologies/bflc/"
xmns:madsrdf="http://www.loc.gov/mads.rdf/v1#">
  <bf:Work rdf:about="http://example.org/6020396#Work">
    <bf:adminMetadata>
      <bf:AdminMetadata>
        <bf:generationProcess>
          <bf:GenerationProcess>
            <rdfs:label>DLC marc2bibframe2 v1.3.0-SNAPSHOT</rdfs:label>
          </bf:GenerationProcess>
        </bf:generationProcess>
        <bf:status>
          <bf:Status>
            <bf:code>p</bf:code>
          </bf:Status>
        </bf:status>
      </bf:AdminMetadata>
    </bf:adminMetadata>
  </bf:Work>
</rdf:RDF>
```
Existing Vocabularies?

OCLC Work ID
- Challenges with technical methods of inclusion in SVDE
- schema:exampleOfWork, but no open outline of how the identifiers are created or defined
- At the time OCLC works were still considered experimental
  - Could be a next step for enrichment of SVDE data
  - It would be interesting to compare approaches and come to a standard for creation of primary resource identifiers through algorithmic processes

Library of Congress Work ID
- Based on LCCN
  - Example: [http://id.loc.gov] /resources/works/c010579972
- Or making use of nametitle authorities
  - Example: [http://id.loc.gov] /resources/works/no98044787
- These are not of universal applicability, and at the time was not ready for use in SVDE
  - Could be a next step for enrichment in SVDE depending on approach
Defining Works

FRBR
“… a distinct intellectual or artistic creation. A work is an abstract entity; there is no single material object one can point to as the work. We recognize the work through individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work.”¹

LRM
“A work is an abstract entity that permits the grouping of expressions that are considered functional equivalents or near equivalents. A work is a conceptual object, no single material object can be identified as the work. The essence of the work is the constellation of concepts and ideas that form the shared content of what we define to be expressions of the same work. A work is perceived through the identification of the commonality of content between and among various expressions.”²

BIBFRAME
“The highest level of abstraction, a Work, in the BIBFRAME context, reflects the conceptual essence of the cataloged resource: authors, languages, and what it is about (subjects).”³

There are differences in definition, and this has been a subject of discussion for many years, but the WIDWG needed a practical solution in a short period of time. We wanted to keep an open mind about the definition of “Work”, and how work identifiers are created in SVDE.

Converting from MARC creates a lot of Work!
Are Work to Work Relationships Sufficient?

1. While Work -> Expression relationships can currently be expressed in BIBFRAME, these are ultimately Work-Work relationships, and determining the initial or primary work, or hierarchical relationships between works may prove difficult with this structure.

2. Through conversion from MARC to BIBFRAME, or automatic work ID generation based on BIBFRAME elements, unless we can define a difference (a fingerprint for each cluster or constellation) between Work and SuperWork elements then these relationships (work-expression) cannot be captured through conversion or automated processing. With the scale of data conversion underway, not doing this would seem like a missed opportunity. Once a separate fingerprint is defined for this primary work, it needs a name, thus the creation of SuperWork.
Defining SVDE Work and SuperWork

SVDE Work

- Is equivalent to a BIBFRAME Work, but is no longer the highest level of abstraction
- Identifiers for SVDE Work are created algorithmically based on unique constellations of elements for BIBFRAME Works (including RDA work and expression level elements)
- The types of SVDE Work and the definitions for which elements are used in its creation are outlined in the Work ID Cluster Mapping

SVDE SuperWork

- The highest level of abstraction in SVDE, the new SuperWork class (subclass of bf:Work) is meant to aggregate or group functional or near equivalent bf:Work clusters
- Identifiers for SVDE SuperWork are created algorithmically based on unique constellations of elements for BIBFRAME Works, minus RDA expression level elements
### Work ID Cluster Mapping

<table>
<thead>
<tr>
<th>Field</th>
<th>Indicators</th>
<th>Subfields</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>W1</strong></td>
<td>130</td>
<td>Sa$6StSg$6kSm$nSnSr$7StSp</td>
<td>Sn may or may not indicate a part; Sp is a part (portion of a work)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>W2</strong></td>
<td>240</td>
<td>Sa$6StSg$6kSm$nSnSr$7StSp</td>
<td>Only if no W1 or W2 identified*</td>
</tr>
<tr>
<td>1XX and 7XX</td>
<td></td>
<td></td>
<td>Used for clustering - Outline here</td>
</tr>
<tr>
<td><strong>W3</strong></td>
<td>245</td>
<td>Sa$6n$7Sp</td>
<td>Only if no W1 or W2 identified*</td>
</tr>
<tr>
<td>1XX and 7XX</td>
<td></td>
<td></td>
<td>Used for clustering - Outline here</td>
</tr>
<tr>
<td><strong>W4</strong></td>
<td>700/710/711</td>
<td>Sa$6StSg$6kSm$nSnSr$7StSp</td>
<td>Note differentiation in relationships based on 2nd indicator 2 (analytic) or &quot;blank&quot; (related work)</td>
</tr>
<tr>
<td><strong>W5</strong></td>
<td>730</td>
<td>Sa$6StSg$6kSm$nSnSr$7StSp</td>
<td>Note differentiation in relationships based on 2nd indicator 2 (analytic) or &quot;blank&quot; (related work)</td>
</tr>
<tr>
<td><strong>W6</strong></td>
<td>758</td>
<td>Sa$45$0$1</td>
<td>As a separate work or clustered in conjunction with other work types. Use identifiers and relationships in SVDE data</td>
</tr>
<tr>
<td><strong>W7</strong></td>
<td>780</td>
<td>Sa$6g$3Ss$7St$w</td>
<td>Note 2nd indicator values for relationship assignment (0-7); also only use $w if the identifier is for a work ID (LCCN, Work URI, ...)</td>
</tr>
<tr>
<td><strong>W8</strong></td>
<td>785</td>
<td>Sa$6g$3Ss$7St$w</td>
<td>Note 2nd indicator values for relationship assignment (0-8); also only use $w if the identifier is for a work ID (LCCN, Work URI, ...)</td>
</tr>
</tbody>
</table>
Work & SuperWork Example

The way the Work and SuperWork are defined in the data:

1. Some Works identified as type SuperWork
2. Works have designation -1 to indicate the difference between the two types
3. hasExpression used from SuperWork to Work
4. expressionOf used from Work to SuperWork

# =========== SUPER WORK TUPLES ===========
Further Analysis, Refinement, and Iteration
Cluster Knowledge Base Editor WG

Terms of Reference
SVDE-TC Cluster Knowledge Base Editor Working Group

Mandate
The Cluster Knowledge Base Editor working group will:

1. Review the CKB Interaction/Editor document and submit feedback on the suggested use cases and processes
2. Propose additional use cases identified as essential for effective knowledge base management
Cluster Knowledge Base Editor WG

Terms of Reference
SVDE-TC Cluster Knowledge Base Editor Working Group

1. Mandate
The Cluster Knowledge Base Editor working group will:
   1. Review the CKB Interaction/Editor document and submit feedback on the suggested use cases and processes
   2. Propose additional use cases identified as essential for effective knowledge base management
Authority and Identifier Management Working Group

Authority work in BIBFRAME and MARC
## URIs table for external sources

<table>
<thead>
<tr>
<th>Code</th>
<th>Source</th>
<th>Institution</th>
<th>Link to Project Page</th>
<th>URI</th>
<th>URI Example</th>
<th>MARC21 Tag</th>
<th>API/WS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NAF (NameTitle + Titles)</td>
<td>Library of Congress</td>
<td><a href="http://id.loc.gov/authorities/names.html">http://id.loc.gov/authorities/names.html</a></td>
<td><a href="http://id.loc.gov/authorities/names">http://id.loc.gov/authorities/names</a></td>
<td><a href="http://id.loc.gov/authorities/names/r00019197">http://id.loc.gov/authorities/names/r00019197</a></td>
<td>100/110/111(Si) $5</td>
<td>S/M</td>
</tr>
</tbody>
</table>

Note: The URIs provided are examples and may vary depending on the specific resource or service.
Every x months, the Service provider checks all authority records to see if any of them has changed.
Discovery UX & UI Working Group
How the Grinch stole Christmas

ID: 1215121

Creators:

Subject:
- PRO Gotham Book Mart (former owner) (Gotham Book Mart Collection copy)
- Christmas stories, -- Juvenile literature
- Fiction
- Fantasy
- Grinch (Fictitious character)
- Christmas stories
- CHR 1985
- Stories in rhyme
- Grinch (Fictitious character) -- Juvenile fiction
- Grinch (Fictitious character) -- Fiction
- Christmas -- Fiction
- Juvenile works

(390367) Seuss, Dr
Search for people, original works and publications
e.g. “Jules Verne” or “Around the world in 80 days”
William Shakespeare

1564-1616. English writer.

William Shakespeare (bapt. 26 April 1564 – 23 April 1616) was an English poet, playwright and actor, widely regarded as the greatest writer in the English language and the world's greatest dramatist.[2][3][4] He is often called England's national poet and the "Bard of Avon".[5][b] His extant works, including collaborations, consist of approximately 39 plays,[c] 154 sonnets, two L... — Wikipedia

Original Works by Shakespeare

<table>
<thead>
<tr>
<th>Original Work title</th>
<th>Format</th>
<th>Year of publication</th>
<th>External links</th>
</tr>
</thead>
<tbody>
<tr>
<td>All's well that ends well</td>
<td>Book</td>
<td>1601</td>
<td></td>
</tr>
<tr>
<td>Antony and Cleopatra</td>
<td>Book</td>
<td>1601</td>
<td></td>
</tr>
<tr>
<td>As you like it</td>
<td>Book</td>
<td>1599</td>
<td></td>
</tr>
<tr>
<td>The Comedy of Errors</td>
<td>Book</td>
<td>1592</td>
<td></td>
</tr>
<tr>
<td>Coriolanus</td>
<td>Book</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hamlet

Written by William Shakespeare in English

The Tragedy of Hamlet, Prince of Denmark, often shortened to Hamlet (ˈhæməlt), is a tragedy written by William Shakespeare sometime between 1599 and 1602. Set in Denmark, the play depicts Prince Hamlet and his revenge against his uncle, Claudius, who has murdered Hamlet’s father in order to seize his throne and marry Hamlet’s mother. Hamlet is Shakespeare’s longest... — Wikipedia
The SHARE-VDE portal (www.share-vde.org)
Where will we go from here: Areas for further analysis

- Further analysis and review of conversion specifications
- Analysis of primary resource clustering
- Development of methods to “clean” up clusters further (aggregate or disambiguate)
- Establish clear authority/identifier profiles
- Create tools and techniques for data creation and maintenance while maintaining the integrity and permanence of identifiers
- Establish workflows for use of the CKB by the wider library community for maximum impact
- Work on sharing experiences with PCC towards development of standards
- Application to workflows for the LD4P Cohort
- In general, to facilitate the implementation of linked data in libraries, determine appropriate processes for data flow
  - MARC - BIBFRAME
  - BIBFRAME - MARC
  - Source of original data
  - ETL
- The list goes on…but so does the work, and improvements come with iteration
Thank You!

“Never let the future disturb you. You will meet it, if you have to, with the same weapons of reason which today arm you against the present.”