



Enhancing an OAI-PMH Service Using Linked Data

The case of the Sheet Music Consortium



Stephen Davison, University of California, Los Angeles

Dynamite Rag.

A NEGRO DRAG.

J. RUSSEL ROBINSON,
Composer of "Bepko Rag," etc.

The musical score for 'Dynamite Rag' is presented in five systems of piano accompaniment. The first system is marked 'Solo March' and includes dynamic markings of *mf*, *f*, *ff*, *sfz*, and *ffz*. The subsequent systems continue the piece with various dynamics such as *f*, *mf*, and *sfz*. The score is written in 2/4 time and features a complex, rhythmic melody with many beamed notes.

Copyrighted, 1910, by Southern California Music Co., Los Angeles, Cal.
International Copyright Secured

Hold Me Parson Hold Me I Feel Religion Coming On

I FEEL RELIGION COMIN' ON.



A First Night Hit with Gov. Bowen of
Al. G. Field's Minstrels
A Coon Song that is Original

The musical score for 'I Feel Religion Coming On' is written in a single system. It includes a vocal line and a piano accompaniment. The score is in 2/4 time and features a simple, catchy melody. The lyrics are written below the vocal line.

Our Big Hit

As Sung by
Lee Lloyd
and
Jay Roberts



Our Big Hit

And 100 Other
Headliners on
the Vaudeville
Stage

For Sale where You Bought this
Published by
Southern California Music Co.,
LOS ANGELES, CAL.

MADE IN U.S.A.

Los Angeles: Southern California Music Co., 1910

New York: Howley, Haviland, Dresser, 1903

Race relations

Performance and performers

Graphic art

Musical composition

“When it’s moonlight on the
Levee, Caroline”

“When I hear the banjos
ringing”

Has: composer, lyricist, graphic
artist, publisher, performers



men

women

Society and Culture--Sentimental song

Songs with piano

Songs

Landscapes

Legacies of Racism and Discrimination--Afro-Americans

Entertainment

Legacies of Racism and Discrimination--Stereotypes--Afro-Americans

Singers

Couples

Afro-Americans

rivers

Society and Culture--Couples

Performers--Men--Kenny

Kenny

Subject headings assigned by Duke University

The Nature of Sheet Music

- Cultural documents
- Multidimensional (variety of purposes)
- Various communities of interest
- Ephemeral in nature
- Printed components mixed, remixed upon reissue
- Variety of descriptive methods and levels
 - Special collections: Finding aids
 - Libraries: Library catalogs
 - Collectors: often interested in graphical components
- All this results in a challenge for a data aggregation service

The Sheet Music Consortium: history and background

- First version launched in 2002
 - 4 members
 - 7 contributing institutions
- “Next Generation” launched in 2011
 - 2 supporting institutions (UCLA, Indiana U)
 - 31 institutions, 29 collections, 228,000+ records
 - metadata mapped to MODS
 - user-contributed metadata services
- Going forward...
 - leveraging “next generation” infrastructure to support publication of linked data

Keep normalized and user-supplied data separate ...

- ... from the harvested metadata
- New data is not easily written back to contributing institution
- Association of harvested and contributed metadata could be lost upon reharvesting
 - Harvested data maintained in XML format and indexed using Solr
 - User contributed data is stored in a separate database

SEARCH

Limit to: Titles Names Subjects Place Publisher Digitized sheetmusic

Community Activity

[Advanced Search Options](#)

Welcome

The Sheet Music Consortium provides tools and services that promote access to and use of online sheet music collections by scholars, students, and the general public.

Not sure what to search for? Try [Browsing](#)

News

June 6, 2013

Linked Open Data project

The Sheet Music Consortium is exposing music publisher information extracted from the Consortium's data as linked open data (LOD). [Read more...](#)

February 15, 2013

Collections and services update

The Sheet Music Consortium now provides cross collection searching to 22 sheet music collections and a total of 226,914 items. If you manage a collection that you would like to be searchable through this site read the [Participation](#) page and [contact us!](#)

New services to be added soon include the ability to download records, assign additional metadata and save back to the Consortium, search user contributed metadata, and access our RSS feed.

June 7, 2011

New Collections, New Services

<http://digital.library.ucla.edu/sheetmusic/>
<http://www.dlib.indiana.edu/smc/migrator/>

Starting today this site becomes the official site of the Sheet Music Consortium. The previous site is still temporarily available for reference purposes at <http://digital.library.ucla.edu/smc/>. The new site offers improved searching and browsing, access to collections at 16 institutions (up from seven), and guidelines for the creation of descriptive metadata.

In addition there is a [Metadata Mapping Tool](#), developed at Indiana University, making it possible to upload, crosswalk and validate metadata from a variety of sources, including plain text files, Microsoft Excel spreadsheets (.xls or .xlsx), or dBase files (.dbf). If you have data in one of these

Activity Stream

Thu, 12 Sep 2013 19:02:16 -0700



emcaulay added a Title to At Memorie's Gate [Read More »](#)

Mon, 1 Jul 2013 23:55:30 -0700



emcaulay added a Title to California [Read More »](#)

Mon, 1 Jul 2013 23:53:01 -0700

emcaulay added an Extent to California and you [Read More »](#)

Tue, 4 Jun 2013 13:19:04 -0700



emcaulay added a Name to Remember Me [Read More »](#)

Mon, 20 May 2013 11:02:10 -0700

BROWSE BY

Featured Repositories

[Title](#) [Subject](#) [Name](#) [Date Published](#)



American University
Library Historical
Sheet Music
Collection
2744 records



Auburn University,
Piano Bench Digital
Collection
323 records



**BALL STATE
UNIVERSITY.**
Ball State University,
Sheet Music from the
Hague Collection
1268 records



Baylor University,
Frances G. Spencer
Collection of
American Sheet Music
4960 records



Butler University,
Sheet Music
Collection
2222 records



Duke University,
Digital Collections
20157 records



Historic Sheet Music
Collection
59 records



Indiana University,
Sam DeVincent
Collection of
American Sheet Music
17431 records



Indiana University,
Starr Sheet Music
Collection
509 records



Johns Hopkins
University, Lester S.
Levy Collection of
Sheet Music
28943 records



Library Of Congress
62976 records



Maine Music Box
18475 records



Mississippi State
University, Charles
Templeton Sheet
Music Collection
12121 records



National Library of
Australia
12385 records



Southern Illinois
University
Edwardsville, KMOX
Popular Sheet Music
Collection
118 records



Temple University,
Sheet Music
Collections
844 records



Tulane University,
Guisepppe Ferrata
Score Collection
78 records



Tulane University,
Louisiana Sheet Music
Collection
630 records

PRIVATE COLLECTIONS

Login to View and Create Your Own Virtual Collections

SHARED COLLECTIONS

FFLsheetmusic

Melbourne

Texas Cowboy Songs

PUBLIC COLLECTIONS

Sacred Christian

Sean

Sean

Songs with ukulele

The Titanic (digitized sheet music)

Wild Blue Yonder

Yiddish Songs

classical piano

dddddd

dogdays

flight of bumble bee

getty test

Yiddish Songs

Share

Mein Weib Shlept Sich Mit. [English, Hebrew, and Yiddish]

Publisher J. & J. Kammen
Repository Johns Hopkins University, Lester S. Levy Collection of Sheet Music



VIEW ONLINE

More

Ich Benk a Heim. Das Voegele.

Publisher Theodore Lohr
Repository Johns Hopkins University, Lester S. Levy Collection of Sheet Music



VIEW ONLINE

More

Levine Levine Der Yid Mit Zein Machine. [In Yiddish]

Publisher Gus Goldstein
Repository Johns Hopkins University, Lester S. Levy Collection of Sheet Music

VIEW ONLINE

More

The Wedding Dress. [German dialect]

Publisher Volk's Theatre Co., 235-37 Bowery
Repository Johns Hopkins University, Lester S. Levy Collection of Sheet Music

VIEW ONLINE

More

Der Naser Keiver Oder Churban Titanic

Publisher New York : Hebrew Publishing Co.
Date 1912
Repository Mississippi State University, Charles Templeton Sheet Music Collection



VIEW ONLINE



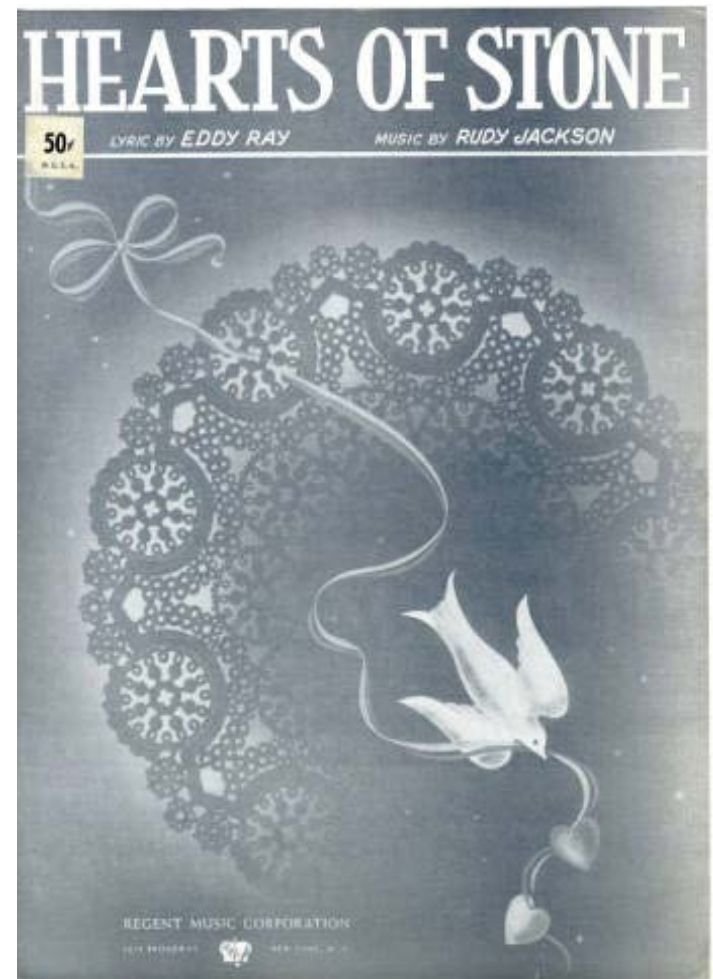
Schemas and Workflows used to harvest records for the Sheet Music Consortium.

SCHEMA	institutions	records
Dublin Core	14	98,317
Qualified Dublin Core	9	26,236
MODS	4	103,504
WORKFLOW		
Direct harvesting using the OAI protocol	25	205,914
Harvesting the metadata via the Static Repository Gateway	1	2,222
Manual extract of MARC records from an integrated library system and mapping to MODS and ingest	1	19,921

SMC and Name Authority

SMC metadata is harvested from diverse institutions, with varying practices

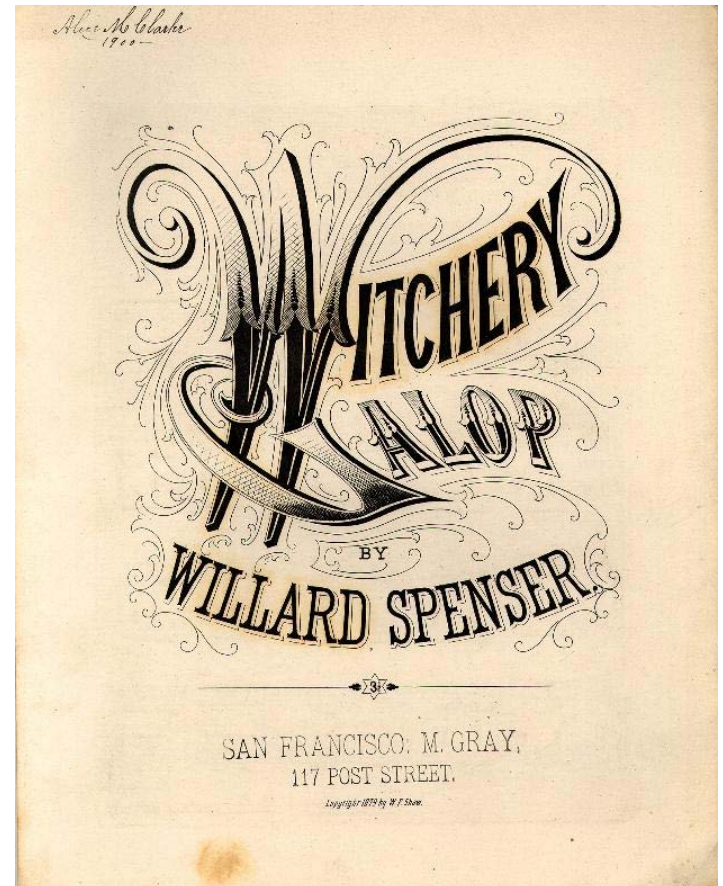
- inventories & finding aids
- spreadsheets
- bibliographic “records”
- focus on music vs. focus on illustrations



California: Granite Music, 1954

SMC and Name Authority

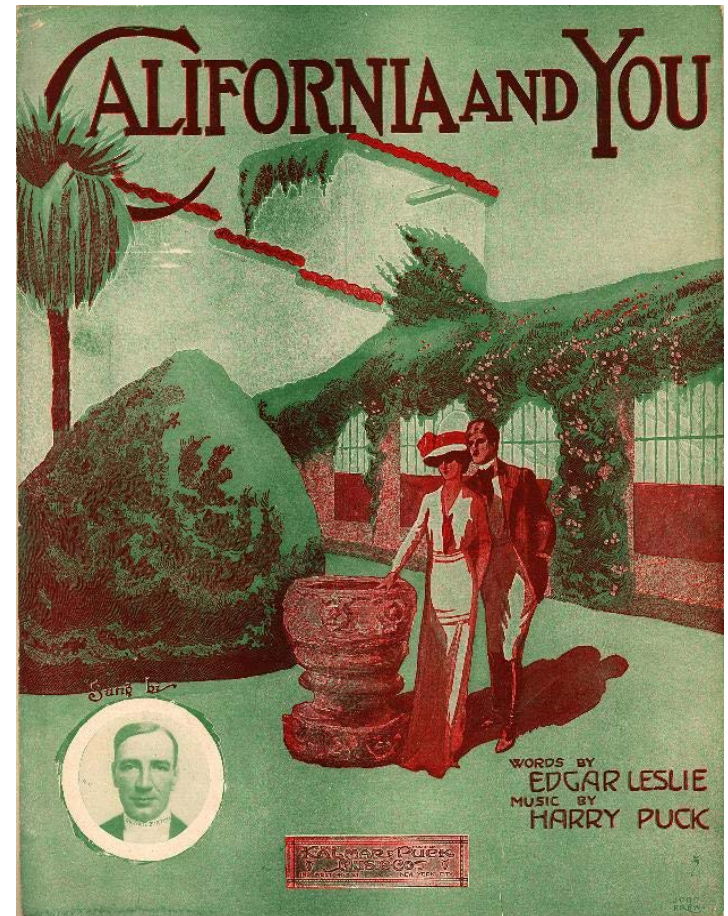
- Resources not always available for authority work at the point of description *or* aggregation
- Some important elements (e.g. Publisher) not traditionally subject to authority control



San Francisco: M Gray, 1879

Challenges of Aggregated Metadata

- Aggregating sheet music records by “work” (as identified by composer & title)
- Variations in practices by contributing institutions
- Example:
 - Harry Puck (composer)
 - Puck, Harry, 1890-1964
 - Puck, Harry [composer]



New York: Bert Kalmar & Harry Puck, 1914

Challenges of Aggregated Metadata

- Sheet music “titles” difficult to define
 - First line of text
 - First line of the chorus
 - The same song may be published under multiple titles
 - California and you
 - California (and You)
 - Oh! you old Pacific Coast
 - A variety of distinct songs may have the same title

Options for publishing linked data

- Works
 - identified by title, composer, lyricist
 - Hard to identify reliably
- Creators
 - authority files exist, e.g. LCNAF
- Subjects
 - authority files exist, e.g. LCSH, TGM
- Publishers
 - generally not represented in existing authority files... some are represented in LCNAF, but usually because they have “authored” works (e.g. catalogs)

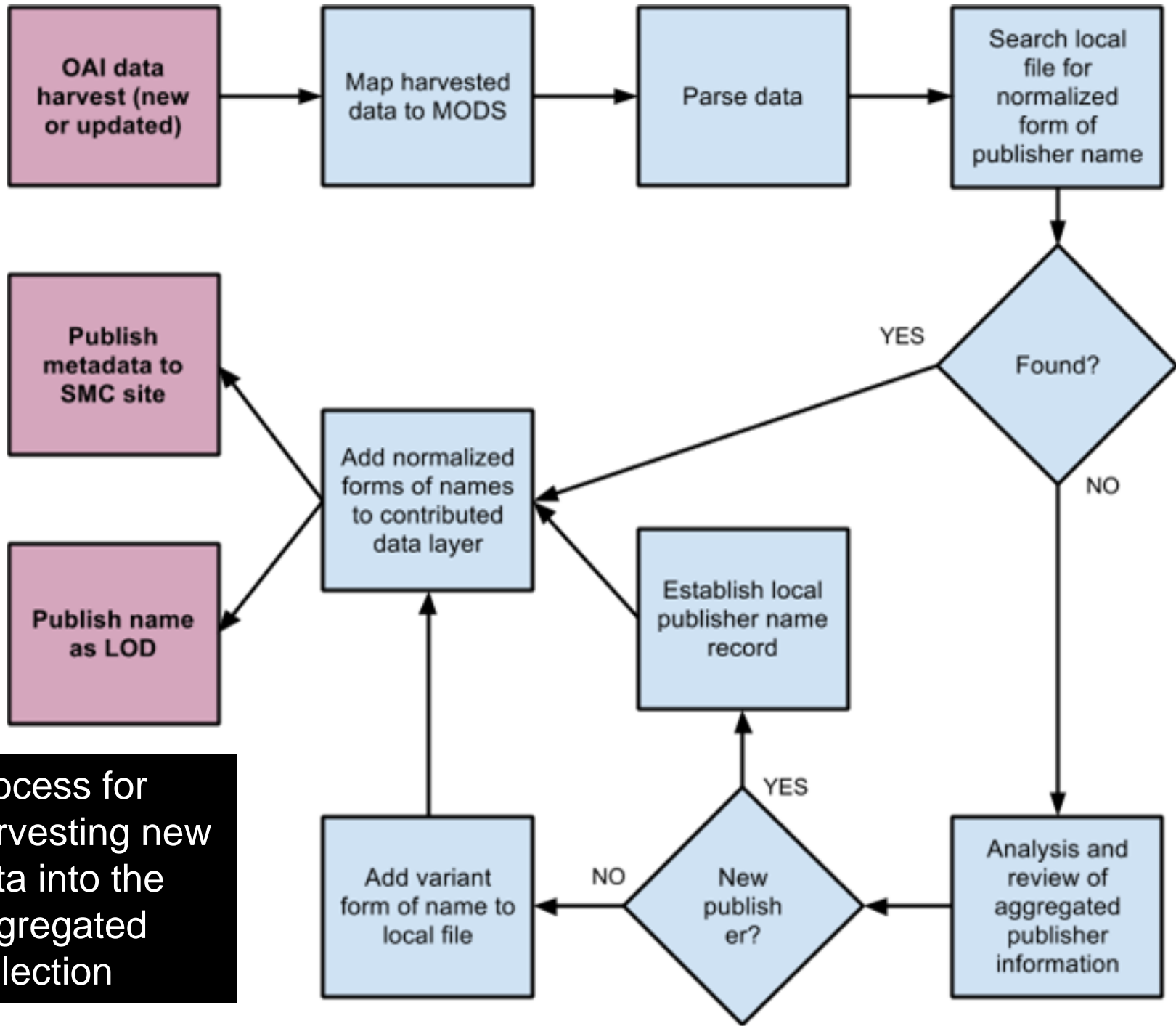
Publishing Aggregated Data as Linked Data: a Pilot Project



- Roles of composers, lyricists, publishers & performers more interrelated than in many other forms of publication
- On published items publisher names and locations change frequently
- LOD provides us with a means of enriching bibliographic information and creating actionable metadata

Strategy for normalizing data

1. Extracted data (**names, titles, publishers**) from MODS records
2. Rank ordered word frequency using **Voyeur/Voyant** tools
3. Chose to work on group of dozen most important publishers
4. Used word frequency data to establish name and title groups
5. Used both internal and external information to establish when publishers really changed identity or ownership
6. Used **Google Refine** to normalize forms of name. Based choice of “preferred form of name” on frequency
7. Wrote these preferred forms back into the repository as “user supplied metadata” (i.e. separate from the harvested data)
8. Published publisher information on the web as HTML and LOD (RDF/XML) (plan also to publish RDFa)
9. Established unique ID's, permanent URLs and link resolution for each publisher



Process for harvesting new data into the aggregated collection

Summary of publisher information generated from SMC data

PUBLISHER NAME	PUBLISHER ADDRESS	DATES OF PUBLICATIONS
Kalmar & Puck		1905
Kalmar & Puck	152 West 45th Street, New York	1913-1915
Kalmar & Puck	New York	1913-1916
Bert Kalmar & Harry Puck	New York	1914-1915
Maurice Abrahams Music Co.	New York	1913-1915
Maurice Abrahams Music Co.	1570 Broadway, New York	1913-1916
Kalmar Puck & Abrahams	New York	1915-1918
Kalmar Puck & Abrahams	1570 Broadway	1917
Kalmar Puck & Abrahams	Strand Theatre Building at 47th St	1917-1918
Maurice Abrahams, Inc.	1591 Broadway, New York	1923
Maurice Abrahams, Inc.		1923-1926
Kalmar & Ruby Music Corp.	6301 Sunset Boulevard, Hollywood	1937-1939



Timeline for Oliver Ditson, Music Publisher

DATE	PUBLISHER	EVENT
1835	Oliver Ditson, Boston	firm founded by Oliver Ditson
1867	Oliver Ditson, Boston	acquired Firth, Son & Co., New York
1867	Charles H. Ditson, New York	firm founded by Oliver's son
1873	Oliver Ditson, Boston	acquired Miller & Beacham, Baltimore
1875	Oliver Ditson, Boston	acquired Wm. Hall & Son, New York acquired Lee & Walker, Philadelphia
1875	James E. Ditson, Philadelphia	firm founded by Oliver's son
1877	Oliver Ditson, Boston	acquired G. D. Russell & Co., Boston acquired J.L. Peters, New York
1879	Oliver Ditson, Boston	acquired G. André, Philadelphia
1883	Theodore Presser, Philadelphia	firm founded by Theodore Presser
1890	Oliver Ditson, Boston	acquired F.A. North & Co., Philadelphia
1931	Theodore Presser, Philadelphia	acquired Oliver Ditson



Publisher LOD project objectives

- Add a layer of information to the aggregation that leverages existing information through a mixture of machine and human analysis
 - Map relationships between names
 - Additional derived information
 - Addresses and dates
- Publish publisher info in a variety of ways:
 - HTML
 - Visualization tools, mapping, timelines
 - RDF
 - RDFa

Archival Resource Keys (ARK) for publishers

PUBLISHER	IDENTIFIER
Kalmar & Puck	ark:/21198/r23x84k8
Maurice Abrahams Music Co.	ark:/21198/r27p8w9m
Kalmar Puck & Abrahams	ark:/21198/r2cc0xm5
Kalmar & Ruby Music Corp	ark:/21198/r2057cvv

The Name-to-Thing (N2T) Resolver:

- Permanent URLs
e.g. <http://n2t.net/ark:/21198/r2cc0xm5>
- Institutional commitment: 21198 = UCLA
- Maintained by the UC Curation Center

<skos:prefLabel>Kalmar Puck & Abrahams</skos:prefLabel>

<skos:altLabel>Kalmar, Puck & Abrahams</skos:altLabel>

<skos:altLabel>Kalmar, Puck & Abrahams Consolidated Inc.

</skos:altLabel>

<skos:altLabel>Kalmar, Puck & Abrahams Consol't'd, Inc.

</skos:altLabel>

<rdfs:seeAlso rdf:resource="http://n2t.net/ark:/21198/r27p8w9m"/>

<!--Maurice Abrahams Music Co.-->

<rdfs:seeAlso rdf:resource="http://n2t.net/ark:/21198/r23x84k8"/>

<!--Kalmar & Puck-->

<rdfs:seeAlso rdf:resource="http://n2t.net/ark:/21198/r2057cvv"/>

<!--Kalmar & Ruby Music Corp-->

MADS/RDF (Metadata Authority Description Schema in RDF) vocabulary

- a data model for authority and vocabulary data
- MADS/RDF is a knowledge organization system (KOS) designed for use with controlled values for names (personal, corporate, geographic, etc.), thesauri, taxonomies, subject heading systems, and other controlled value lists
- fully mapped to SKOS vocabulary
- designed specifically to support authority data as used by and needed in the library community
- designed to support the description of cultural and bibliographic resources

```
<madsrdf:Address>
  <rdf:Description>
    <madsrdf:streetAddress>Strand Theatre Building at
      47th Street</madsrdf:streetAddress>
    <madsrdf:city
      rdf:resource="http://sws.geonames.org/5128581/" />
    <time:year>1917</time:year>
    <time:year>1918</time:year>
  </rdf:Description>
</madsrdf:Address>
```

Conservatoire François Mitterand, Mauritius – SMC's newest member

- Small collection of sheet music
- Looking for advice
- Wants to publish digital surrogates on the web
- Our strategy:
 - Create descriptive metadata in a local DB
 - Map to MODS using SMC's online tool
 - Upload metadata to SMC's Static Repository
 - Ingest to MSC using Static Repository Gateway
 - Metadata added to the Web of Data through SMC



Have demonstrated a strategy for mitigating some of the problems in aggregated metadata and publishing normalized data on the web as linked data.

Over time normalized linked data may take on the role that authority records do in OPACs, and may its way into formal authority vocabularies.

Publishers are just a start... now we need to republish other normalized elements to the "web of data."

OAI is still a useful tool for harvesting data.

With mapping tools and static repositories even the smallest of players can contribute.

A possible model for other bibliographic projects.

New York: Howley, Haviland, Dresser, 1903

*With special thanks to my
collaborators and co-authors:*

Yukari Sugiyama

East Asia Library, Yale University

Elizabeth McAulay

UCLA Digital Library Program

Claudia Horning

UCLA Cataloging & Metadata Center

Stephen Davison

sdavison@library.ucla.edu

