

Increasing Access to Archives Through Linked Open Data

Karen F. Gracy
Kent State University
SAA 2012 Research Forum

A Research Report on ...

The Metadata Vocabulary Junction Project:

Connecting Library Data to the Unfamiliar Data and Metadata Resources in the LOD Universe

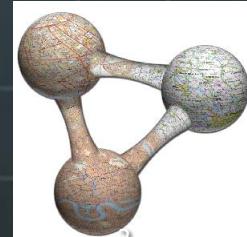
PI: Marcia Lei Zeng, Ph.D.

<http://lod-lam.slis.kent.edu/>

Co-PI: Karen Gracy, Ph.D.

Research Assistants:

Laurence Skirvin, MLIS; Sammy Davidson, MLIS; Shadi Shakeri, MLIS;
Riley Stoermer, MLIS





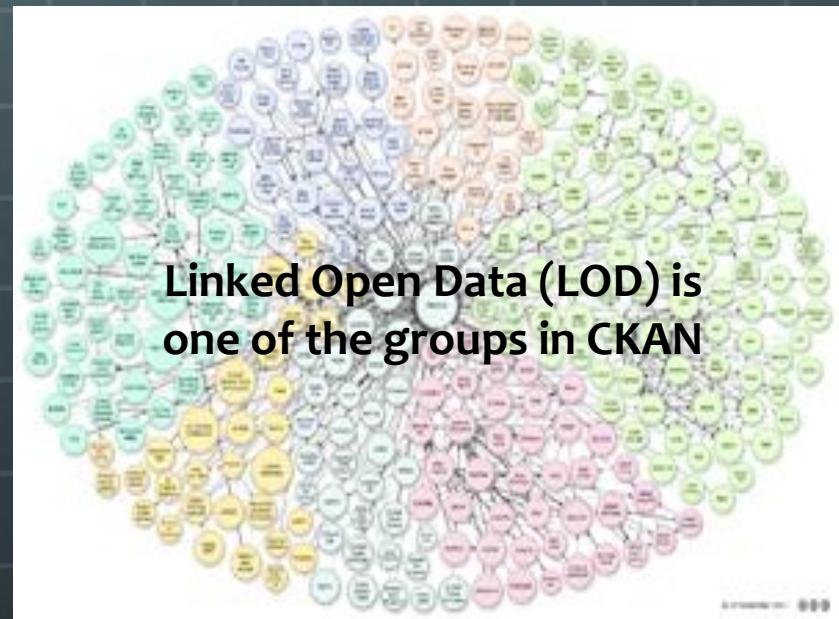
the Data Hub — The easy way to get, use and share data

[Search](#) [Groups](#) [About](#)

Project purpose:

To connect library,
archive, and museum
(LAM) data to the
unfamiliar datasets
available in the Linked
Data (LD) community's
CKAN Data Hub.

5,951 datasets
(as of August 4, 2012)



Preliminary Focus Areas

- Authority information
- Music and music-related information
- Archives and archival finding aids
- Museum object documentation

Aligning Library/Archive Data with LOD Data Constructs

LAM data:

- Library bibliographic records
 - MARC records
 - schema.org markup
- Digital collection metadata
 - Dublin Core-based
 - Other locally defined
- Archival descriptions
 - EAD, MARC, other
- Museum and Visual Resources object descriptions
 - VRA Core, other



LOD Datasets:

- Ontologies
- Metadata Application Profiles
- Structured data in XML/RDF samples
- Documentation



Data Collection and Analysis (1): Archives Information Systems

1. Analyze archival descriptive and authority standards to determine what information may be useful as linked data:
 - Encoded Archival Description
 - Encoded Archival Context
2. Review finding aid exemplars from several sources to discover how standards are applied:
 - OhioLink EAD Repository
 - Online Archive of California
 - Library of Congress Finding Aids
3. Identify and review other potential sources of information as needed

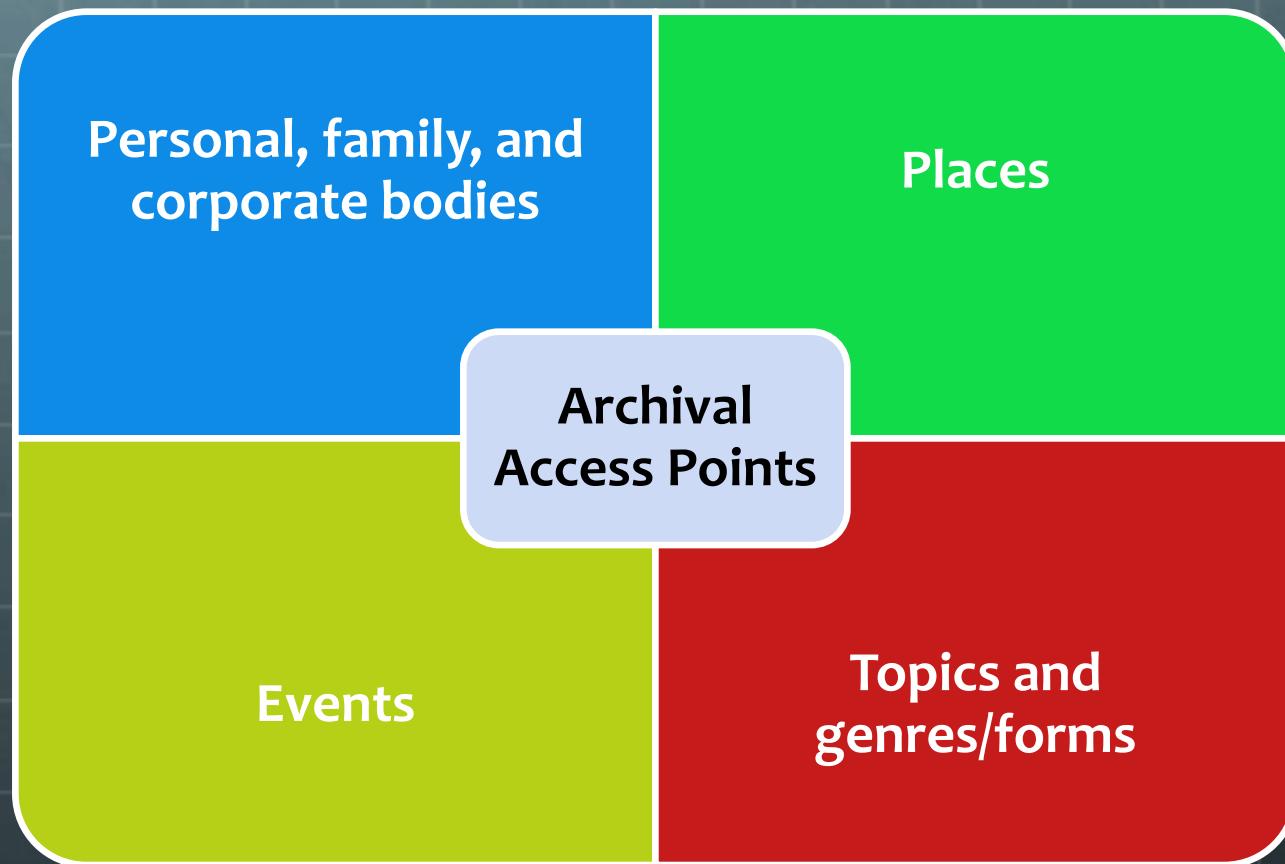


4. Then, break down the “records” and find:
 - Common elements
 - Possibilities for linking archival data elements to LOD classes and properties
 - Major access points
 - Hidden access points

Data Collection and Analysis (2): LOD Datasets

- 1. Identify datasets potentially relevant to archival collections**
 - a. Records creator/agent information
 - b. Geographic information
 - c. Event and activity information
- 2. Analyze the source of the data structures**
 - a. Ontologies used
(e.g., Dublin Core Terms, FOAF, Event Ontology, LODE Ontology, GeoNames Ontology, etc.)
 - b. Metadata schemas and application profiles
(e.g., schema.org, LOCAH, Europeana EDM Schema, Archival Schema from U.Mass-Amherst) and derivatives)
 - c. Structured data in XML/RDF samples
 - d. Documentation
- 3. Crosswalk their properties, indicating the matching level**
(e.g., broadMatch, equivalent, narrowMatch, etc.)
- 4. Identify major classes and properties useful to archives data**
(e.g., foaf:name, dbpedia:Place, dc:subject schema:Event)

Potential Intersection Points in the Archival Descriptions



Alignment Example: EAD to FOAF (partial)

Root: EAD container tag(s)	Root: EAD Data Element	Target: FOAF Class	Class Mapping	Target: FOAF Property	Property Mapping
<archdesc> <controlaccess>	<persname>	foaf:Person	CM	foaf:name, foaf:familyName, foaf:givenName, foaf:nick	N, B, B, B
<archdesc> <controlaccess>	<famname>	foaf:Group (?)	N	foaf:name	N
<archdesc> <controlaccess>	<corpname>	foaf:Organization	N	foaf:name	N
<archdesc>	<scopecontent>	foaf:Person, foaf:Agent, foaf:Organization	B, B	foaf:name, foaf:familyName, foaf:givenName, foaf:nick	B, B, B, B
<archdesc>	<bioghist>	foaf:Person, foaf:Agent, foaf:Organization	B, B	foaf:name, foaf:familyName, foaf:givenName, foaf:nick	B, B, B, B

