A Minimum Viable Strategy for Archives and Linked Data using Schema.org

Mark A. Matienzo, Stanford University / @anarchivist

Society of American Archivists Session 303 **#s303**
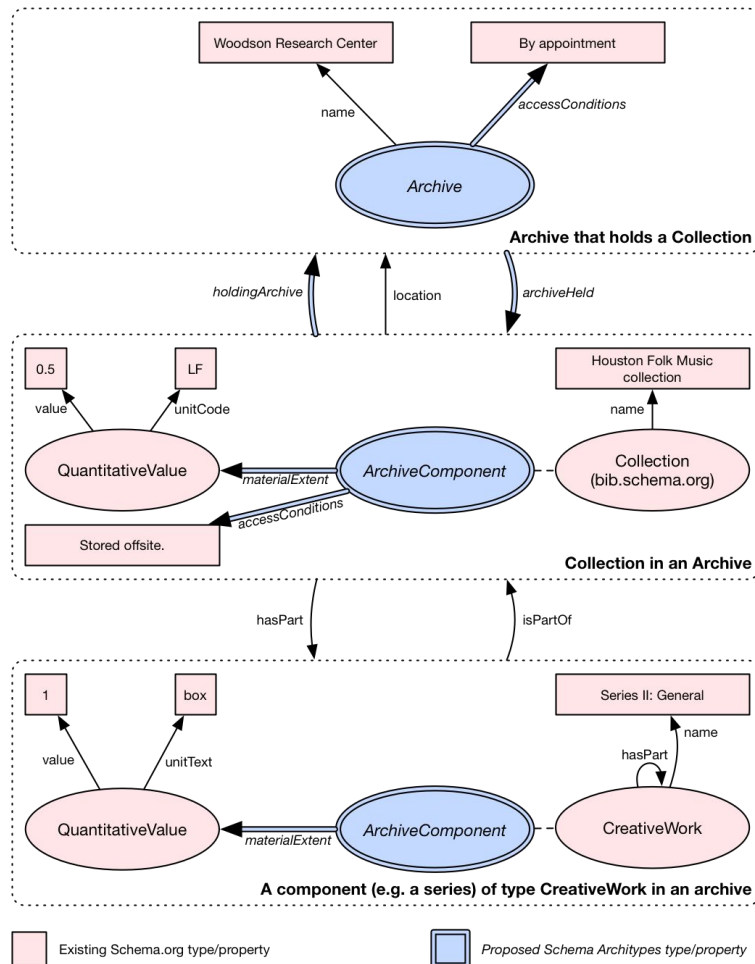
16 August 2018

Rationale and objectives

- Group of archivists and technologists interested in pragmatic approaches to linked data about archives

- Investigating use of Schema.org and its extensions as a minimally viable mechanism for publishing linked data

- Development of mappings from archival description standards to Schema.org/extensions

- Production of RDF-modeled archival description directly from archives management systems

Schema.org

- Created in 2011 by Bing, Google, and Yahoo to address structured data format proliferation for search engines

- Provides single schema across range of topics: people, creative works, places, etc. (589 types, 862 properties)

- Used on 1+ billion web pages & many popular websites

- Expressible as JSON-LD, RDFa, and Microdata

- Provides extension mechanism for both Schema.org-hosted and external extensions

Schema Architypes

- Schema.org extension to represent archives, proposed for inclusion in October 2017

- Introduces a minimal set of new types/properties

- Selected as the basis for our data modeling work

Mapping description to Schema.org

- Preliminary mappings from ISAD(G), ISAAR-CPF, DACS, and ArchivesSpace/Atom data models

- Mostly straightforward with notable exceptions:

  - Description control
  - Level of description
  - Reference code
  - Precision of note types

Elements of DACS and ISAD(G) mapped to both existing Schema.org properties/types and proposed properties/types from the Architypes Proposal

Example

```
{
  "url": "http://archives.library.rice.edu/repositories/2/resources/1038",
  "@context": "http://schema.org/",
  "@type": ["Collection", "ArchiveComponent"],
  "name": "Houston Folk Music collection",
  "inLanguage": "EN",
  "holdingArchive": {
    "@type": ["Archive", "LocalBusiness"],
    "name": "Woodson Research Center Special Collections & Archives",
    "address": "Fondren Library, Rice University, 6100 Main, Houston, TX 77005",
    "url": "http://library.rice.edu/woodson/"
  },
  "creator": [
    { "@type": "person", "name": "Townes Van Zandt", "@id": "http://viaf.org/viaf/56876870" },
    { "@type": "person", "name": "Guy Clark", "@id": "http://viaf.org/viaf/29717590" },
    { "@type": "person", "name": "Vince Bell", "@id": "http://viaf.org/viaf/78141250" },
    { "@type": "person", "name": "Lyle Lovett", "@id": "http://viaf.org/viaf/120632356" }
  ],
  "description": [
    "This collection contains items documenting the Houston Folk music scene from the 1970s-1980s.",
    "During the 1960s-1980s, Houston, TX had a vibrant folk scene. Local musicians and those from other parts of Texas and the
U.S. socialized and played at a variety of music establishments around the city, including Anderson Fair Retail Restaurant,
Sand Mountain Coffeehouse, Liberty Hall, The Old Quarter, Theodore's and Corky's. This scene spawned many musicians, Townes Van
Zandt, Guy Clark, Robert Earl Keen, Nanci Griffith, Lynn Langham, Wrecks Bell, Eric Taylor, John Vandiver, Danny Everitt, Vince
Bell, Richard Dobson, Don Sanders, Wheatfield/St. Elmo's Fire, Dogtooth Violet, Bill Cade, Lyle Lovett, and many more."
  ],
  "dateCreated": "1975-1985",
  "accessConditions": [
    "This material is open for research.",
    "Permission to publish from this material must be obtained from the specific copyright owner."
  ],
  "ownershipInfo": "Albums donated by Brendan Doss, III. Townes Van Zandt reel donated by Craig Calvert."
}
```

Expected benefits and future work

- Mappings allow for publication of linked data directly from archives management systems or discovery environments

- Development of a recommended profile for implementation in a wide variety of systems

- Further investigation and mapping to other ontologies and data models, and addressing existing gaps in modeling

Thank You!

Resources

- Archives and Linked Data Interest Group:
  https://archival.github.io/

- W3C Schema Architypes Community Group:
  https://www.w3.org/community/architypes/

- R.V. Guha, D. Brickley, and S. Macbeth, "Schema.org: Evolution of Structured Data on the Web." *ACM Queue* 13(9), 2015.
  http://queue.acm.org/detail.cfm?id=2857276

- M. Matienzo, E.R. Roke, and S. Carlson, "Creating a Linked Data-Friendly Metadata Application Profile for Archival Description." *DCMI 2017.* https://arxiv.org/abs/1710.09688

Acknowledgements

The Archives and Linked Data Interest Group:

- Scott Carlson (Rice University)
- Mark Custer (Yale University)
- Patrick Galligan (Rockefeller Archives Center)
- Dan Gillean (Artefactual Systems)
- Gloria Gonzalez (Zepheira)
- Maggie Hughes (UCLA)
- Mark Matienzo (Stanford University)
- Dave Mayo (Harvard University)
- Laney McGlohon (LYRASIS/ArchivesSpace)
- Evelyn McLellan (Artefactual Systems)
- Katy Rawdon (Temple University)
- Elizabeth Russey Roke (Emory University)
- Ruth Kitchin Tillman (Penn State)