

Cataloging Archives and Special Collections for the Semantic Web

Guy Frost

Valdosta State University

May 12, 2021, Georgia Users Group Meeting

Standards for Archival Description

Tools of the Trade (-2020)

- ***Archives, Personal Papers, and Manuscripts (APPM)*** and ***Describing Archives: A Content Standard (DACS)***
 - Change needed because of the "Web, XML, and EAD"
 - DACS appeared in 2004
 - Both APPM and DACS were developed to replace Ch.4 of AACR2
 - MARC-based
- ***Resource Description and Access (RDA)*** in 2012
 - Replaces AACR2
 - Still MARC-based, but with its eye toward a new non-MARC environment

Buzz words continue to emerge: Linked Data, Semantic Web, BIBFRAME, RDF, Triples, Quads, Graphs, and much, much more

- ***Resource Description and Access (RDA)*** - 3R Project (2020 Revision)
 - Designed to be applied to RDF built systems

Archives Processing

"More Product, Less Process" (MPLP) - 2005

"Cataloging is a function which is not working."

- Opening of article - 55 pages (Greene and Meissner)
- Article focused on addressing "massive backlogs of inaccessible collections"
- Research Found:
 - Archivists could not agree on what is important
 - More collections ingested than descriptions created - couldn't keep up
 - Hypothesis: archivists do too much that does not need to be done right away or every time or anytime

Findings

- Quarter-half of holdings unprocessed (1998); 34% more than half; 60% one-third unprocessed (2003/2004)
 - 10s of 1,000s of cubic feet
 - Fun fact: William F. Buckley, Jr.'s collection is the length of about 3 football fields according to his bibliographer
 - Applying "traditional approaches to a new problem"; unwilling to change
 - Researchers and Donors not pleased (51%) - bad in the eyes of resource allocators
- Proposal: Expedite collections to users; Adequate description; minimal steps; sufficient for promoting the collections (aka "Good Enough")

Other Handbooks and manuals

- Do not go folder level "unless there is a compelling benefit to be derived from such detailed work" (Brunton and Robinson)
- Do not go below series level; do not add series
- "Respect the organization imposed by the person who created the records" "Strive for simplicity" (St. Johnsbury Athenaeum Archives)

Scholars take notice...

Scholar Chas Clifton blogged about one of VSU's collections

"He has also written a lengthy historical note on the
Andersons and Feri."

The Finding Aid

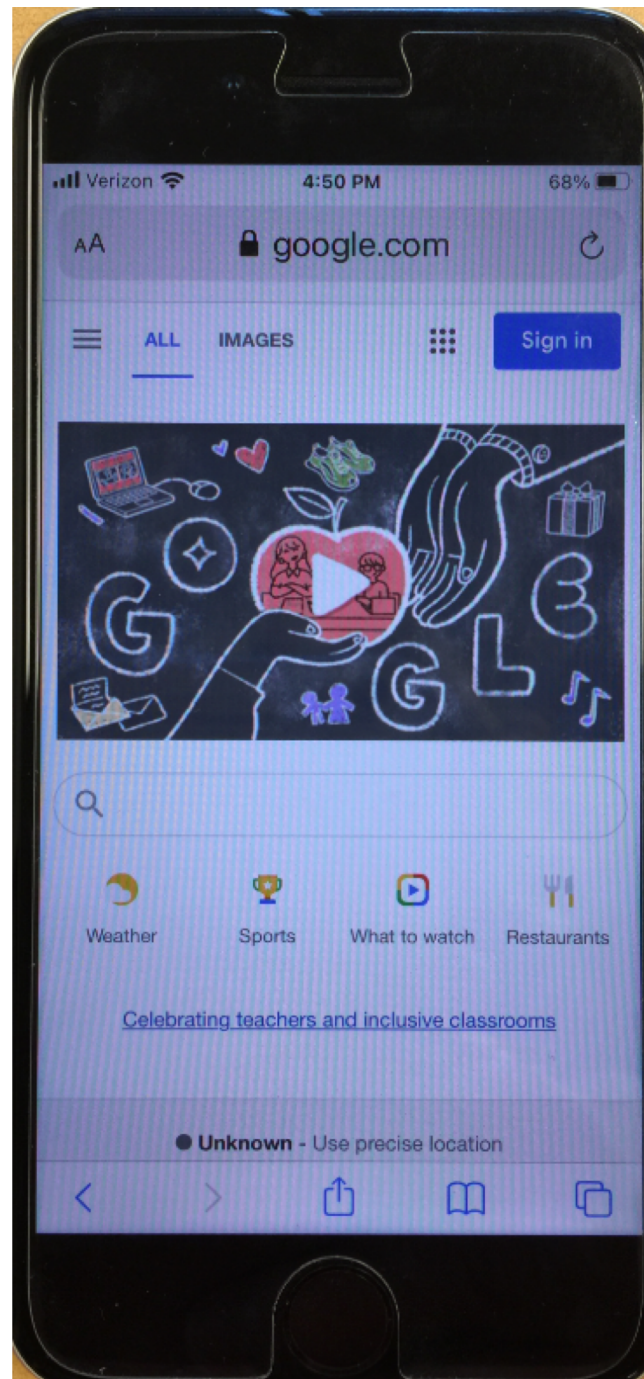
The Finding Aid is often "substantial, multilayered, descriptive" (philosophically, still old practice)

- describes the contents of series or folders
- Assists researchers coming to the archives by providing "something about its entire holdings"
- One shoe does NOT fit all
- APPM pushed the Finding Aid a new level of important
 - Chief source of information for MARC
 - Without a full Finding Aid, cataloging was not possible
- MPLP changed this
 - Stub records can be "Good Enough"
 - Exceptions to the Rule need to be made

But wait, there's more... What has happened in the last 20 years?

Smartphones

- Began to emerge 2002ish; touchscreens late 2000s-early 2010s
- Became a game changer for user expectations for information retrieval
- Many Users are discovering existence of Archival Collections through Google searching
- accessing ArchivesSpace
- RDA/BIBFRAME on the horizon
- *ArchivesGrid* (OCLC) is not on a regularly scheduled update
- Technology continues to evolve as well as discovery mechanisms



Perspective

Users needs

Fulfilled

- Webmaster of the official site living in California - item level description
- Economist/Parliamentarian from New Delhi, India - item level description
- Researcher on brother of collection subject - item level description

Semi-fulfilled

- Scholar in the UK - partial item level description; some full text in the institutional repository [word of mouth]

Not Fulfilled

- Documentary Film director in New England - unprocessed reel-to-reels [word of mouth]

What do these all have in common?

Researchers at a Distance

Advances in technology coupled with systems that are on Google's side of the web (like Archon, ArchivesSpace, and ArchivesGrid) will bring new users to Archives that expect (hope?) for services that do NOT require them to visit in person. Rather, they want the items digitized and sent to their inbox.

Archives and Special Collections Linked Data (July 2020)

- OCLC Research Position Paper
 - "Linked data forces you to structure your data" (e.g., less free-text notes and descriptions; more coded fields that include the use of URIs)
 - "Shift from notes to entities should influence descriptive practice and take the profession toward data that wants to be structured" ... "major shift in mindset"
 - "Potential for better discovery" - "identity markers" supporting user tasks (IFLA LRM concept): i.e. demographic data such as gender, residence, social groups
 - "Inclusion and increased visibility of 'item-specific information' e.g.: binding produced by X
 - End users will still need to adapt to changes ; [developers too]
 - "Education barrier that is difficult to overcome "the trouble with triples"
 - "A lot of labor with little commitment to financing the labor"
- More Product, Less Process has set the "Golden Minimum"
- "Work remains out of reach for most small to medium size archives ... "shift in mindset to think about how to structure archival descriptions 'natively'" using the tools and systems in place (i.e., ArchivesSpace) ; "balance local versus global IDs"

Archives Space Identifiers for Local Headings (Hypothetical)

Local Heading: Agriculture events

Identifier: <https://archivesspace.valdosta.edu/staff/subjects/1033>

- Local Headings Identifiers are unique to the institution
 - Does not work outside the Archives System (yet)

Currently, Local Authorities cannot be added to OCLC records that allows added identifiers

Local Application using the \$0 for authority control number. Note: only works if/when ArchivesSpace allows external linking or integration with the LMS

**690;__; \$a Agriculture events \$0
<https://archivesspace.valdosta.edu/staff/subjects/1033>**

How the Mindset Needs to Change, or (Old Thinking vs New Thinking)

OLD



Flat File

NEW



Linked Data

Genres, Thesauri, Registries (aka Points on the Star)

Structure Data - Fields with subfields for unique "values". This is the Age of Identifiers (i.e. CURIEs)

For example:

340;__; \$d daguerreotype process \$2 rdapm (registry URI)

<http://www.rdaregistry.info/termList/RDAproductionMethod/>

340;__; \$d daguerreotype process \$2 rdapm \$0

<http://www.rdaregistry.info/termList/RDAproductionMethod/1004>

Note: Registry URIs can be found in the new RDA Toolkit

385;__; \$m Occupational/field of activity group: \$n occ \$a Scholars \$2 lcdgt

<http://id.loc.gov/authorities/demographicTerms/dg2015060404> (authorities URI)

<http://id.loc.gov/authorities/classification/E445.G3> (Slavery; Antislavery movements in Georgia URI)

Explicit Descriptions (Isness vs. Aboutness)

- Old school subject strings are being separated into different Facets
- RDF systems require data to be explicit; will reduce the number of false hits
- Aboutness will only be for what the item is about.
- Subject heading values that reflect what something "is" will/should be moved elsewhere ; no more ambiguity

Field 385 - Audience Characteristics

- This demographic will allow catalogers the ability to add a target audience for an archival collection. Who would be interested in this?

Three most common for Archives

¶m Occupational/field of activity group: ¶n occ ¶a Scholars ¶2 lcdgt

¶m Occupational/field of activity group: ¶n occ ¶a Historians ¶2 lcdgt

¶m Occupational/field of activity group: ¶n occ ¶a Genealogists ¶2 lcdgt

Collection Specific: John C. Huxford Papers

Composer: Contains manuscript scores

¶m Occupational/field of activity group: ¶n occ ¶a Musicologists ¶2 lcdgt

Primo Audience Characteristics (385)

User searches are often generic

Browse by subject ▾

Modern dance

300 — |a vii, 189 pages : |b illustrations, music |c 28 cm.

Audience Facets pull from 521 note and 385

Bibliographic records provide additional linkages

Audience ^

11 & up (17)

9 & up (5)

Collaborative pianists (1)

Audience

Occupational/field of activity group: Dance teachers >

Occupational/field of activity group: Physical education teachers >

Occupational/field of activity group: Collaborative pianists >

ArchivesGrid (351,385,386)

Series: Series I. Organized by subseries: I. Ritual Scripts and Related Documents (Online); Spring Mysteries Festival; Esbat Ritual Scripts; Sabbat Ritual Scripts II. Correspondence. III. Physical Resources. IV. Miscellaneous items

Occupational/field of activity group: occ Scholars

Occupational/field of activity group: occ Historians

Religious group: rel Wiccans

Gender group: gdr Women

Nationality/regional group: nat Washingtonians (Washington State)

Nationality/regional group: nat Americans

Religious group: rel Wiccans

Occupational/field of activity group: occ Herbalists

Identifiers: ‡2 lcdgt, ‡2 lcsh

Associated Place (new field 370)

Creation of works and expression

Example: **Bill of sale for slaves (MS/1)**

\$c United States \$g Baker County (Ga.) \$2 naf \$t 1837
\$c United States \$g Pulaski County (Ga.) \$2 naf \$t 1839

Counties where sales originated (created)

Note: **\$s** would be beginning date and **\$t** ending date if that was appropriate

Note: This is *not the subject* but where created

Timespace for Aggregates (Field 388 and 046 \$o \$p)

Aggregates: two individual works or expressions issued together in a single work not originally created together

Example: Bill of sale for slaves (MS/1)

Contains two manuscript sales receipts for slaves, 1837 and January 2, 1839

046;__; \$o 1837 \$p 1839-01-02 \$ edtf

388;__; Nineteenth century \$2 lcsb [Best Practices: use only when dates not known (i.e., no **046 \$o \$p** used); OCLC's *BF&S* shows both

Timespan for significant events

- Current LCSH for the Civil War
651;_0; United States \$x History \$y Civil War, 1861-1865
- Transferring this LCSH event to the 388 is undesirable
FAST Heading
388;1_; \$a American Civil War (1861-1865) \$2 fast \$0
<http://id.worldcat.org/fast/1351658>

Genre (655) - the isness

Library of Congress Genre/Form Terms (lcgft)

655;_7; \$a Personal correspondence. \$2 lcgft

Common Genres in LCGFT

Records (Documents)

Personal correspondence

Business correspondence

Diaries

Ephemera

Personal narratives

Genre/Form ^

Personal correspondence (3)

Ephemera (3)

Records (Documents) (3)

Business correspondence (3)

Directories (2)

Are more options better?

True story: Students in a graduate research class asked:

Why was there more than one choice for the same search term?

Students chose the one with the most records and ignored the other two.
"One stop shopping"

I see a usability study for someone...

Newspapers	110 records
Newspapers (<i>aat</i>)	2 records
Newspapers (<i>fast</i>)	1313 records
Newspapers (<i>lcgft</i>)	872 records
Newspapers (<i>rbgenr</i>)	16 records

Teasing out "of interest" content: Relationships

- Cataloging certain "subunits", items, or components separately

Field 774 - Constituent Unit Entry (Show relationship to the a smaller collection)

774;08;\$i Container of (manifestation): \$t Cuneiform tablets collection. \$d [between 2112 and 2004 B.C.] \$w (OCoLC)809934105

Field 773 - Host Item Entry (Reciprocal Relationship) (008 BLvl: d)

773;08;\$i Contained in (manifestation): \$a Powell, Richard Holmes, 1875-1947. \$t Richard Holmes Powell papers \$w (OCoLC)17792438

RDA 2020; structured description (values in subfields) and identifiers (\$w)

Serials, serial issues

Serials

- Provides title access for "known items"
- Potential for increased use through ILL requests

Serial issues

- Lone serial issues may be contained in folders or may be stand alone for specific subject content (Playboy magazine)
 - Could choose to catalog the article alone as an analytic
- Allows more granularity than just the serial title
- Number of instances of this scenario unlikely to be frequent

Collocation

Serials, Serial Issues, and Monographs can be linked together a devised access point

Series (490/830; shared catalog fields 962/961) - UGA practice

490;1_;\$a [Cuyler (Ned) Brooks Fanzine Collection]
830;_0; \$a Cuyler (Ned) Brooks Fanzine Collection

Corporate Body Access Point (710; shared catalog field 953) - Valdosta practice

Unique items

710;2_;\$a Emily Hendree Park Memorial Library (Valdosta State University) \$5 GVaS

Non-Unique items (i.e. published)

953;2_;\$a Emily Hendree Park Memorial Library (Valdosta State University) \$5 GVaS

Both options work in a shared catalog when local fields are used; to work in a linked data environment \$0 URI would likely be required

Roles

Adding the specific role for creators would provide the linking mechanism needed

710;2_;\$a Jocu Playing Cards (Firm) \$e publisher. \$4 pbl

- Often used for rare books

\$e versus \$4

Disambiguation is a key factor: *PCC Task Group in Linked Data Best Practices Final Report* (2019)

- \$0 cannot differentiate \$a from \$e
- \$4 can be a code or a URI
- \$e optional; can include both

Versions (251)

A new field from 2018 (published and unpublished items)

- Allows for making a distinction between two items that are similar but different (mock ups, prototypes, drafts, one with plates one without, etc.)
- Displays in Primo in same place edition statements display



251;__; \$a Prototype version

Can be a controlled term with \$2, \$0, \$1 allowed; \$2 currently undefined

Bibliography

American Library Association Institutional Repository. Subcommittee on Faceted Vocabularies. *Best Practices for Recording Faceted Chronological Data in Bibliographic Records* (2020). <http://hdl.handle.net/11213/14591>.

Greene, Mark A. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist* 68, no. 2 (Fall/Winter 2005): 208-63. <https://www.jstor.org/stable/40294290>.

OCLC Research Archives and Special Collections, Linked Data Review Group. *Archives and Special Collections Linked Data: Navigating between Notes and Nodes* (Dublin, OH: OCLC Research, 2020). <https://doi.org/10.25333/4gtz-zd88>.

PCC Task Group on Linked Data Best Practices. *PCC Task Group on Linked Data Best Practices Final Report* (September 12, 2019). <https://www.loc.gov/aba/pcc/taskgroup/linked-data-best-practices-final-report.pdf>.

Thank You

Open for Questions