# Remodeling Archival Metadata Descriptions for Linked Archives
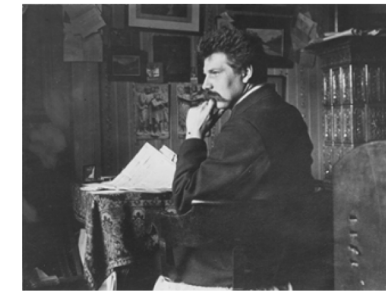
Brian Dobreski, Jaihyun Park, Alicia Leathers, & Jian Qin
School of Information Studies
Syracuse University
Syracuse, New York, USA

What are special collections?

- Cultural heritage objects
  - Archives
    - Text documents
    - Post cards
    - TV program videos
    - Photographs
    - Art and architecture drawings
  - Rare books
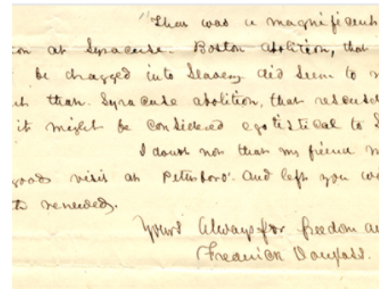 - Sound recordings
  - …
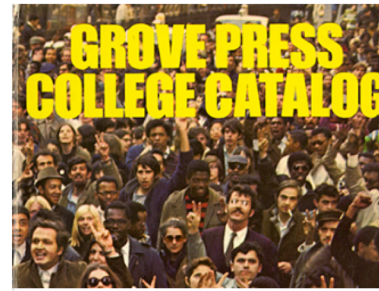
Grove Press Records

Albert Schweitzer Papers
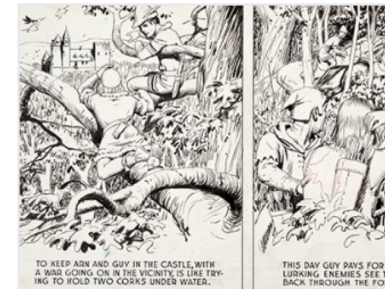

The Syracuse Hours

Gerrit Smith Papers

Gutenberg Bible Leaf

Harold R. (Hal) Foster Papers

Syracuse University Postcard Collection

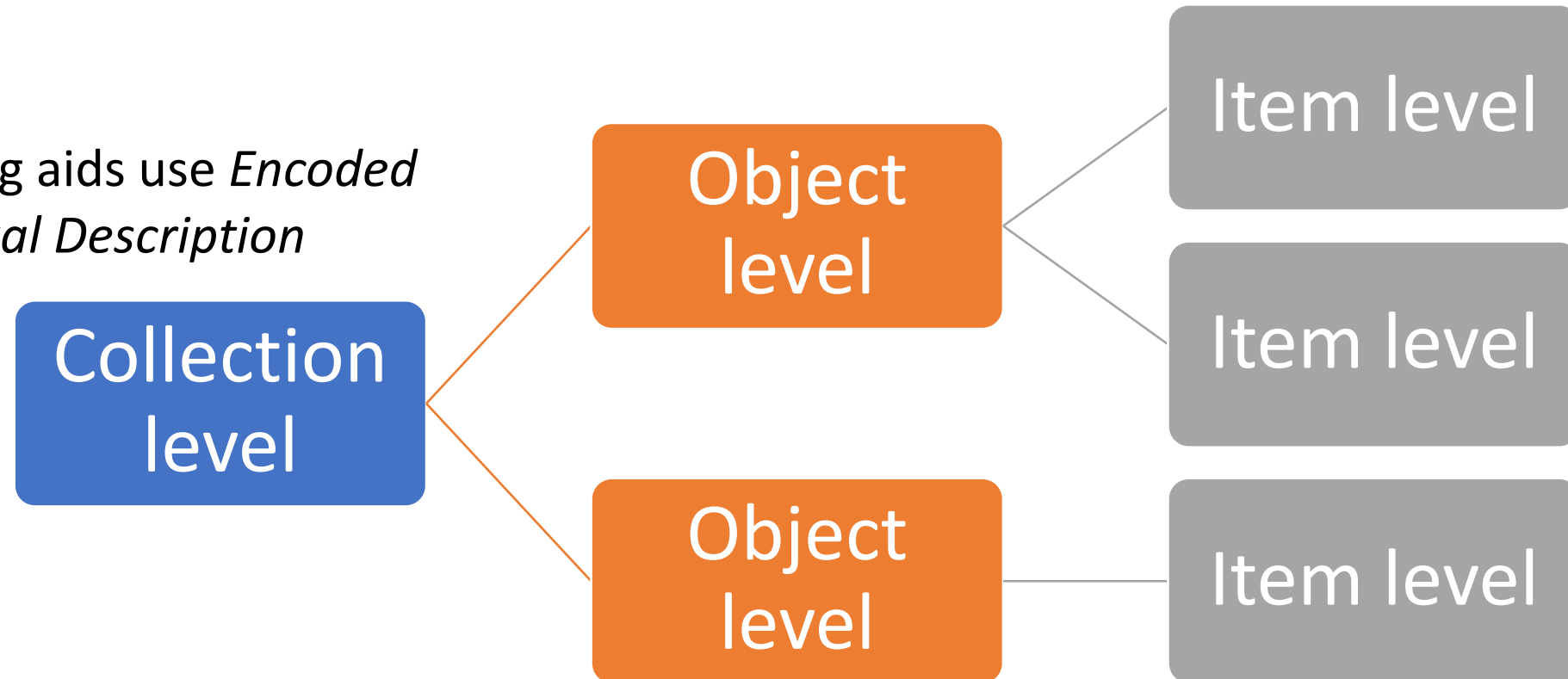

Marcel Breuer Papers

Metadata descriptions for special collections

Object and item level metadata descriptions use an in-house schema based on METS and MODS

Finding aids use *Encoded Archival Description (EAD)*

Syracuse University
Libraries
Special Collections Research Center

Collection Title: Marcel Breuer Papers ⬀

Creator: Breuer, Marcel, 1902-1981.

Inclusive Dates: 1921-2001

Abstract: Papers of the Modernist architect and designer, includes architectural drawings, photographs and slides, project files, writings, correspondence and other materials related to every phase of Marcel Breuer's career.

Type of Material:
Architectural drawings (visual works)
Blueprints (reprographic copies)
Change orders.
Clippings (information artifacts)
Contracts.
Correspondence.
Detail drawings (drawings)
Exhibition catalogs.
Exterior perspectives.
Interior perspectives.
Lantern slides.
Manuscripts for publication.
Memorabilia.
Photographs.
Schematic drawings.
Sketches.
Slides (photographs)
Specifications
Topographic surveys
Writings

Collection level metadata

- EAD record
- Relations to other special collections
  - Herbert Beckhard Papers
  - George Goodwin Collection Relating to Marcel Breuer
  - St. Luke's Church Building Committee Records
- A wide variety of material types
- A large number of people and corporate bodies

Object level metadata

mets:metsHdr "2017-12-07T16:09:13-05:00"
- mets:agent "EDITOR" Syracuse University Library
  - mets:name Syracuse University Library

mets:dmdSec "dmd1"
- mets:mdWrap "text/xml"
  - mets:xmlData 38
    - dc:identifier 38
    - dc:title Wire
    - dc:creator Roth, Gordon, B. (Author)
    - dc:date 1945-05-11
    - dc:subject Geller House I
    - dc:type Correspondence
    - dc:rights Images supplied by the Marcel Breuer Papers, Syracuse
    - dc:language English
    - dc:description Geller House I. Roth, Gordon, B. (Contractor). Roth,

mets:dmdSec "dmd2"
- mets:mdWrap "OTHER"
  - mets:xmlData
    - breuer:object 38

mets:dmdSec "dmd-image_1"
- mets:mdWrap "OTHER"

mets:dmdSec "ead"
- mets:mdRef "URL"

mets:fileSec
- mets:fileGrp "thumbnail image"
  - mets:file "fid1"
    - mets:FLocat "URL"
- mets:fileGrp "reference image"
  - mets:file "fid2"
    - mets:FLocat "URL"

mets:structMap
- mets:div "dmd1 dmd2"
  - mets:div "dmd-image_1"
    - mets:div "thumbnail image"
      - mets:fptr "fid1"
    - mets:div "reference image?"
      - mets:fptr "fid2"

5

Object and item level metadata

mets:metsHdr "2017-12-07T16:09:13-05:00"
- mets:agent "EDITOR" Syracuse University Lib…
  - mets:name Syracuse Universi…
- mets:dmdSec "dmd1"
  - mets:mdWrap…
    - …rdon, B. (Author)
    - …945-05-11
    - dc:subject Geller House I
      - dc:type Correspondence
      - dc:rights Images supplied by the Marcel Breuer Papers, Syracuse
      - dc:language English
      - dc:description Geller House I. Roth, Gordon, B. (Contractor). Roth,

mets:dmdSec "dmd2"
- mets:mdWrap "OTHER"
  - mets:xmlData
    - breuer:object 38
- mets:dmdSec "dmd-image_1"
  - mets:mdWrap "OTHER"
- mets:dmdSec "ead"
  - mets:mdRef "URL"
- mets:fileSec
  - …
    - mets:FLocat "URL"
  - mets:structMap
    - mets:div "dmd1 dmd2"
      - mets:div "dmd-image_1"
        - mets:div "thumbnail image"
          - mets:fptr "fid1"
        - mets:div "reference image?"
          - mets:fptr "fid2"

Nature of the object: Correspondence
Type of content: Wire to contact the gas company
Persons and firms: Role (author, contractor, recipient, etc.)

Digital counterpart: images of correspondence
Access and organizations: ID, location, sequence of order

Sample metadata record for object

collection_id: 227

IDs are local and internal

object_id: 78584

object_type: 1

time_stamp_created: 2017-03-01 16:50:23

time_stamp_updated: 2017-03-01 16:50:23

time_stamp_exported: 0000-00-00 00:00:00

object_draft: 0

object_deleted: 0

title: Das M'zuzele

title_alt:

Lots of administrative metadata

...tion:

date_original_range: 1917-

date_original_display: 1917

date_issued:

date_issued_display:

description: <p>Foreign Blue Amberol record release: Hebrew Series.</p>

coverage:

series: Edison Blue Amberol

series_id: 10051

media_type: Sound Recording

open_closed: Open

type: Music

internal_id: eba_10051

bibid: 1572810

alt_repo:

draft:

rights:

geo_code:

Lots of empty data cells

donor:

summon_content_type: Music Recording

language: Yiddish

subject: Songs, Yiddish

notes:

index:

collection_id	227	
object_id	78584	Linking collection, object, and item via IDs
item_id	80312	
object_type_id	2	
timestamp_created	2017-03-01 17:00:51	
timestamp_updated	2017-03-01 17:00:51	
timestamp_export	0000-00-00 00:00:00	
item_draft	0	
item_deleted	0	
title	eba100511edited	
date_digital	2016-02-04	
linked_objects	78584	
dimensions		Item file path important to include in transformation
duration	03:01	
internal_id	uEliPsh9p...	
file_display	scrc/belfer_cylinders/tinfoil2.gif	
file_archive	eba_10051_1_edited.mp3	
internal_file_path	scrc/belfer_cylinders/tinfoil2.gif	
file_size_display	5.53 MB	
item_html		
item_download	Checked	
checksum_archive	9772c709025ec3ffe3f13635d4835292	
dimensions_digital		

(Digital) item level metadata

dimensions_digital	
general_technical_information	"Compression: PPI: Quality: Good ScanHW: Benchm... ScanSW: ProTools... Sample Rate (Aud... Sample Rate (vide... Sampling Ratio: Codec: "
	Belfer Audio Archi...
	24 bits 44.1
digitized_by	
file_bit_depth	
sampling_rate_audio	
sampling_rate_video	
color_bw	
sampling_ratio	
codec	
file_quality_archive	
file_scan_hw_archive	
file_scan_sw_archive	Good
file_format	Benchmark ADC
sound	ProTools LE 7.4

Why ontological modeling?

- Enrich semantics in metadata by establishing a network of related people, events, subjects, places, times, material types, and other features important and meaningful for users

- Transform the structures of metadata to fully realize the technology potentials

- Enable the utilization of semantic data available externally and the publication and sharing of the semantic data about our collections

Metadata samples

- Belfer Cylinders Collection: music and spoken word recordings dating from 1890 to 1929.
  - 1729 physical item records and 1729 digital item records
  - 3000 name records for individuals, groups, and other entities in various roles.
- Ronald G. Becker Collection of Charles Eisenmann Photographs:
  - photographs of 19th century sideshows, circuses, and performers, most taken by Charles Eisenmann or his successor Frank Wendt, dating from 1836 to 1960
  - 1,414 physical item records, 1,416 digital item records, and 1,504 name records describing various roles.
- Ted Koppel Collection: videos of ABC News television programming with Ted Koppel, including approximately
  - 6,600 episodes of Nightline (March 1980-November 2005
  - 7416 physical item records, 13,610 digital item records, and 72,988 name records of individuals appearing in various roles.

| Category	Elements common to both	Physical items	Digital items
Administrative	collection_id, internal_id, object_id, object_type_id, time_stamp_created, time_stamp_export, time_stamp_updated	alt_repo, bibid, series_id date_issued, donor, draft, index, location, notes, object_deleted, object_draft, open_closed, related_items, rights, summon_content_type	item_id, Item_draft, item_deleted, digitized_by, linked_objects, item_html, item_download, checksum_display, checksum_archive, notes, orig_gen, orig_format, orig_notes, dig_notes
Descriptive	title	coverage, date_issued_display, date_orig_display, description, geo_code, language, media_type, series, subjects, subject_local, title_alt, type	date_digital, duration, color_bw, dimensions, dimensions_digital, physical_description, sound,
Technical (digital items only)	file_display, file_archive, general_technical_information, file_compression_archive, file_ppi_archive, file_quality_archive, file_scan_hw_archive, file_scan_sw_archive, file_type_archive, file_size_display, file_size_archive, internal_file_path, file_bit_depth, sampling_rate_audio, sampling_rate_video, sampling_ratio, codec, file_format, tech_info_file, tech_info_preservation		

Decision making issues

- Should every piece of metadata be included in the ontological model?
  - Can this ontological model replace the traditional relational database for all metadata functions (inventorying, managing, organizing, and curating special collections)?
- If not, what metadata elements should be included in the ontological model?
  - What is exactly the purpose of this ontological model?
    - Increase interactivity and discoverability of metadata and objects in collections
    - Administrative metadata are local and not necessarily included in the model, but there should be a way to link to them automatically when necessary

Approaches to ontological modeling

- Principles:
  - No creation of new metadata, but utilizing what have been created
  - Select only the elements that would facilitate interactivity and discoverability
  - Sustainable, meaning the process can be automated to reduce reinventing the wheel while maintaining consistency
  - Use an infrastructural approach whenever possible
- Steps:
  - Understand humanities scholars' needs and the ways of using special collections
  - Collect information about previous models
  - Analyze metadata elements and structures to identify classes and properties
    - Reusable classes and properties from existing models?
    - What new classes are absolutely needed?
  - Tune the model based on test with examples
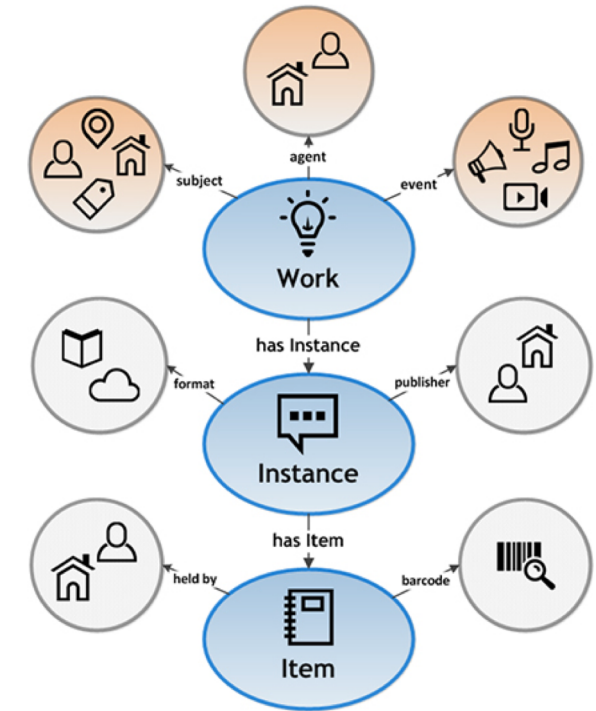
Models that already exist

CIDOC-CRM credit: https://slideplayer.com/slide/7984641/
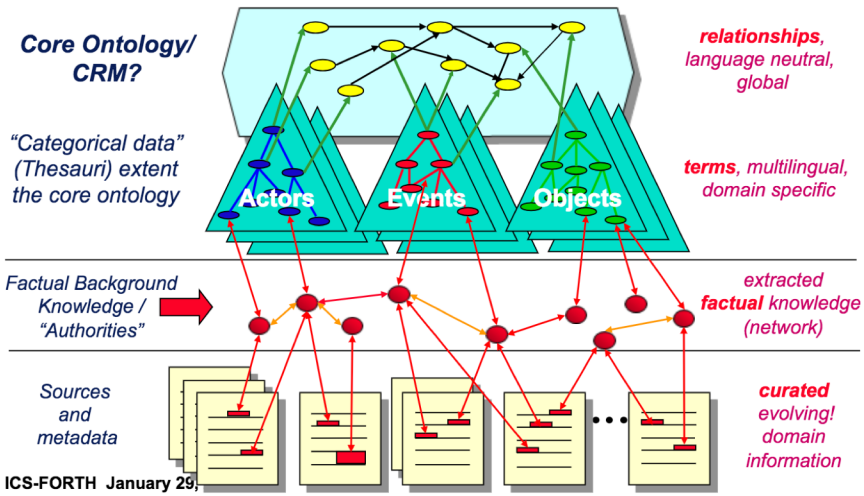
- Overly sophisticated models can make automatic metadata conversion very difficult
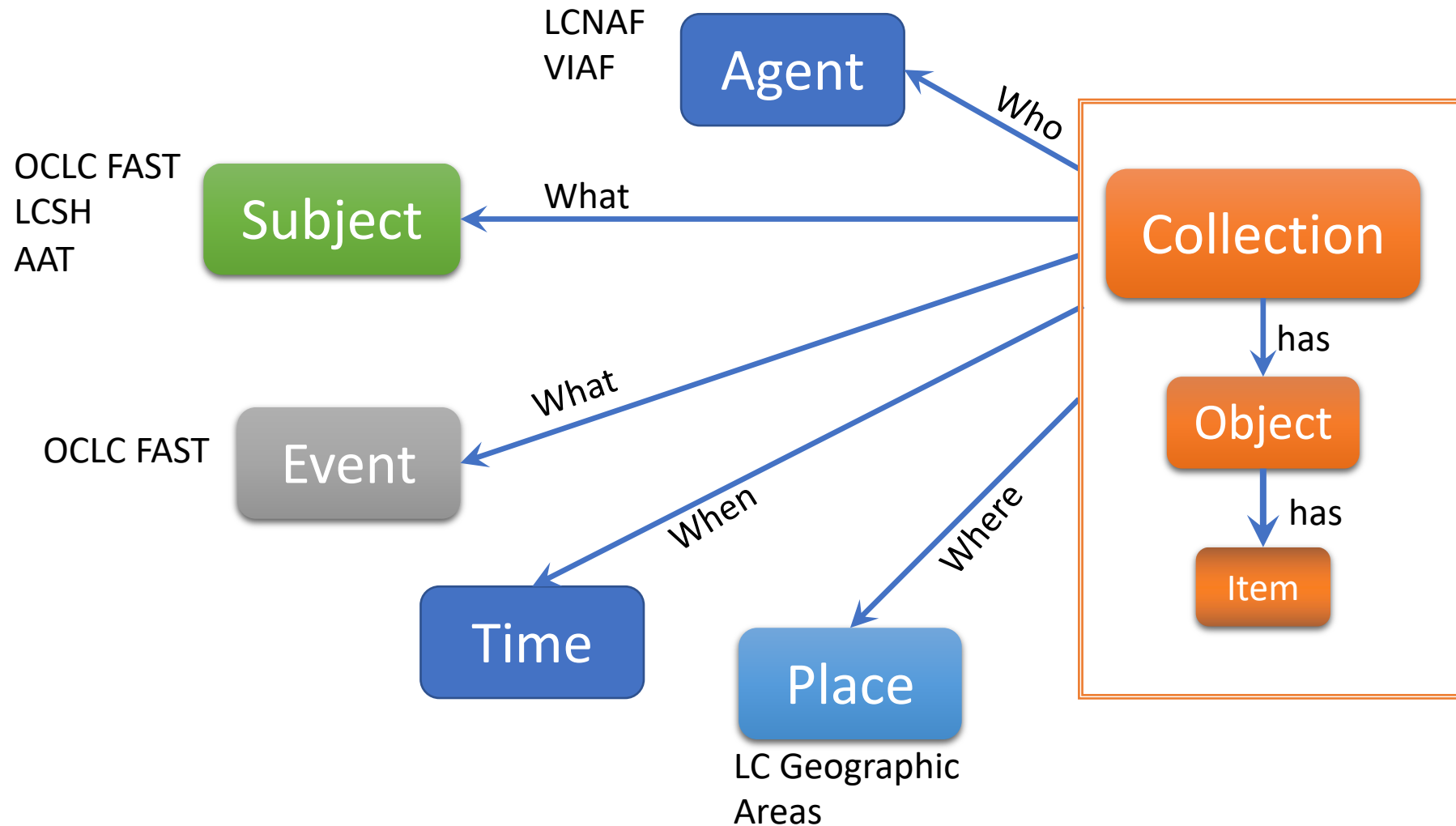- Stay focused: interactivity and discoverability
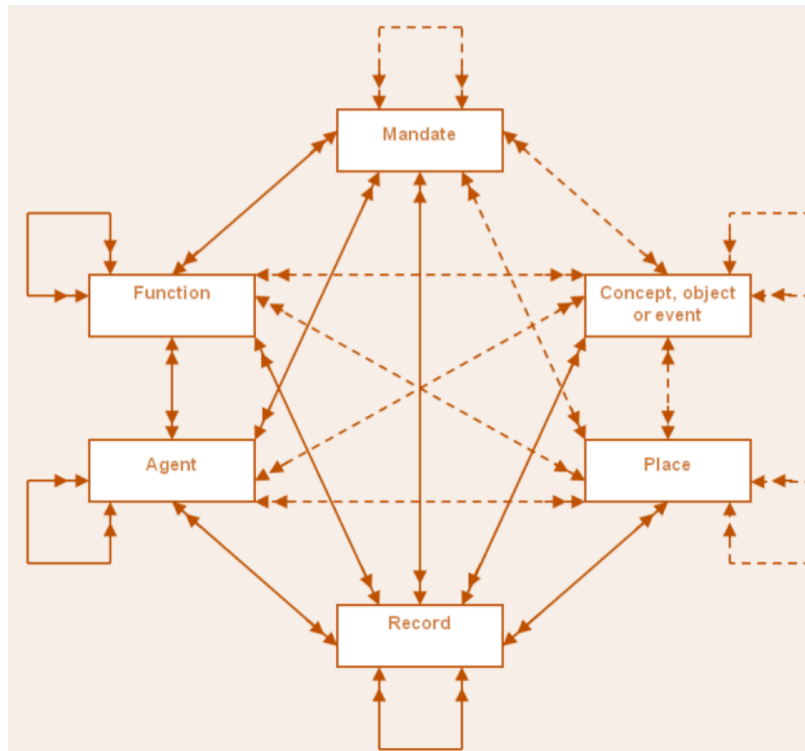
Record in Context – Conceptual Model (RiC-CM)

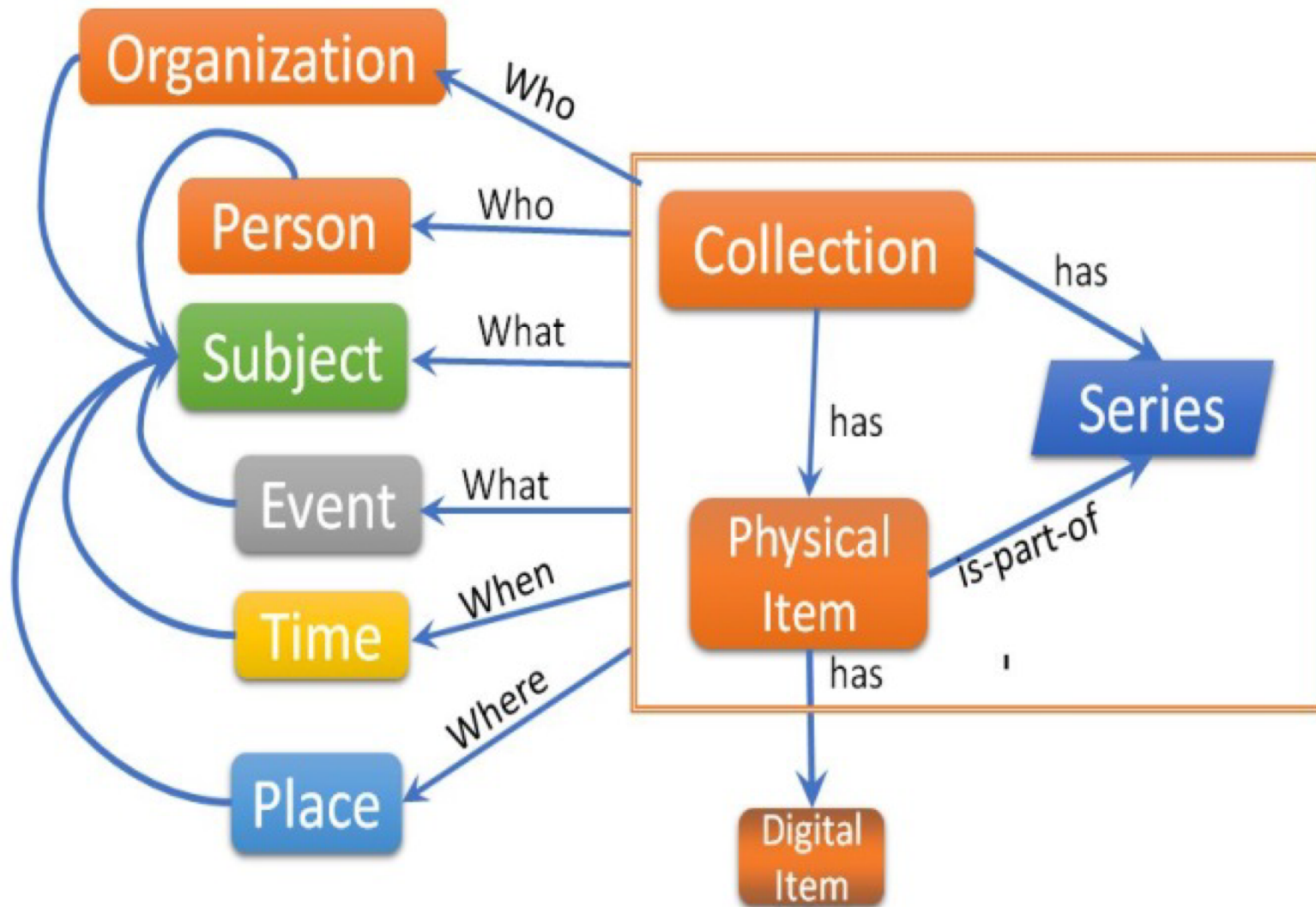https://www.ica.org/sites/default/files/session-7.8-ica-egad-ric-congress2016.pdf

14

Reusable classes and properties from existing models

- Criteria for external classes and properties to be "reusable"
  - Able to enrich semantics for content
  - Able to establish meaningful links between entities
  - Able to map to existing metadata descriptions

- How to keep the criteria consistent – documentation, documentation, documentation

The initial Linked Archive Model

LCNAF
VIAF

Agent

OCLC FAST
LCSH
AAT

Subject

Who

What

Collection

has

OCLC FAST

Event

What

When

Object

has

Item

Time

Where

Place

LC Geographic
Areas

Revised Linked Archives Model

Agent's role types determine relation types

Creator
Subject
Participant

Agent → creates → Collection/Object
Agent → participates in → Collection/Object
Agent → is_subject_of → Collection/Object

Collection level:
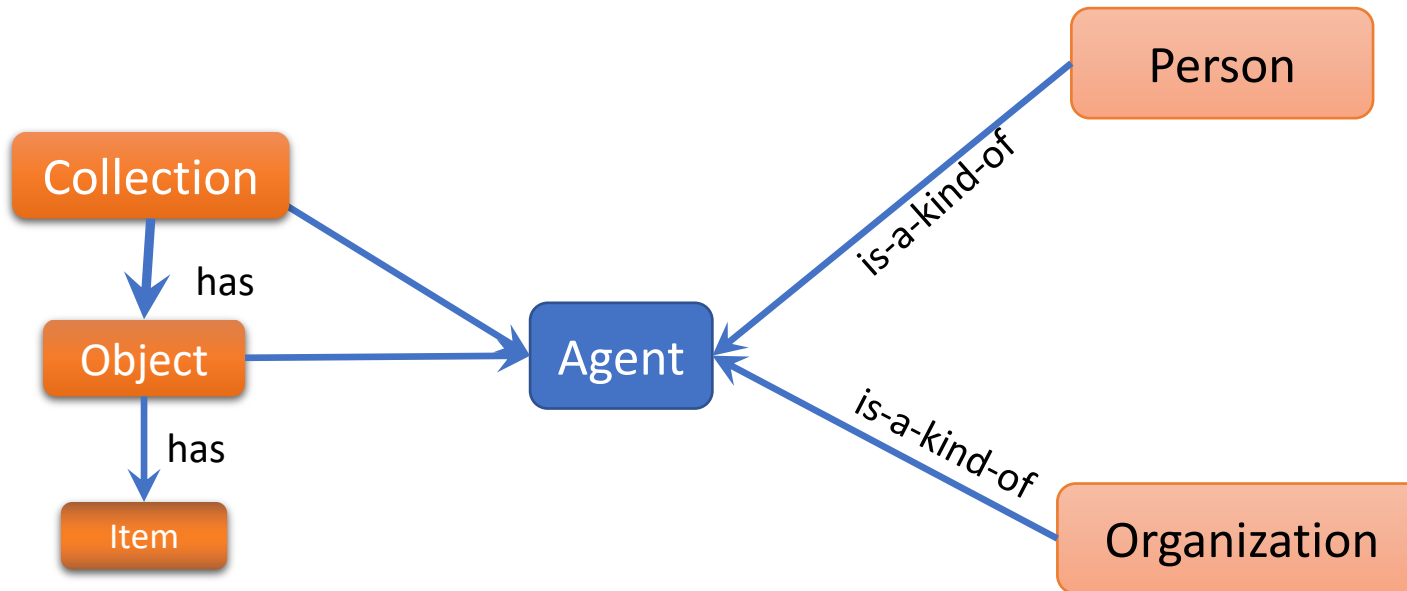- Author
- Subject

Object level:
- Author
- Interviewee
- Appearance
- Show host
- Photographer
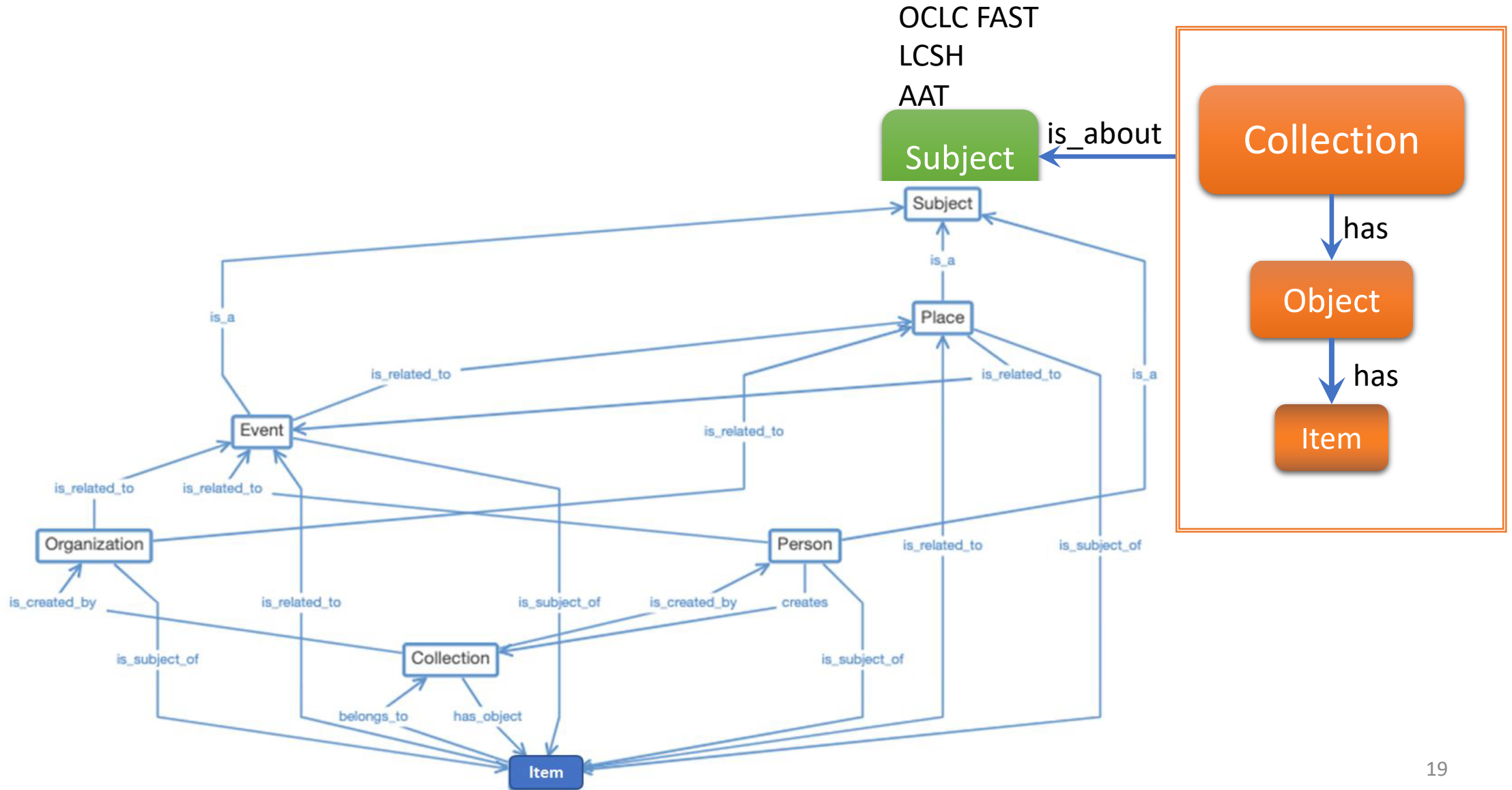- Composer
- Performer
- Contractor

Collection level:
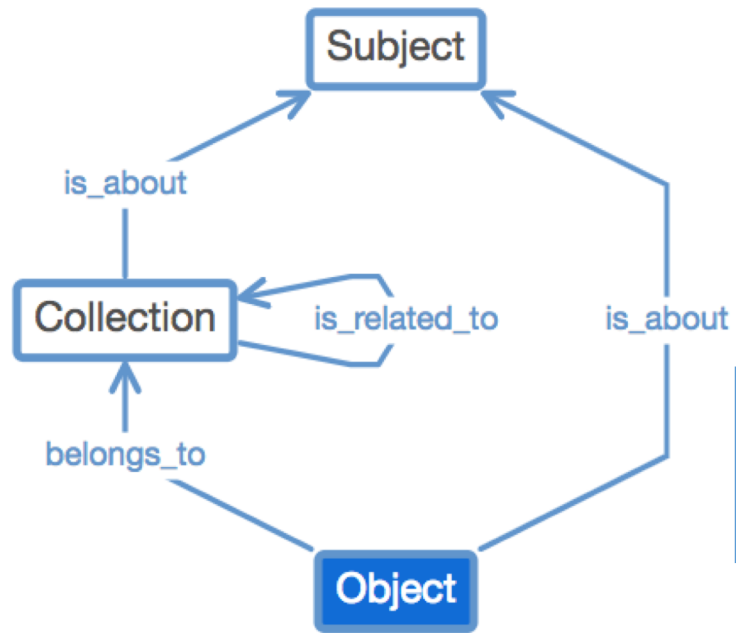- Repository
- Creator
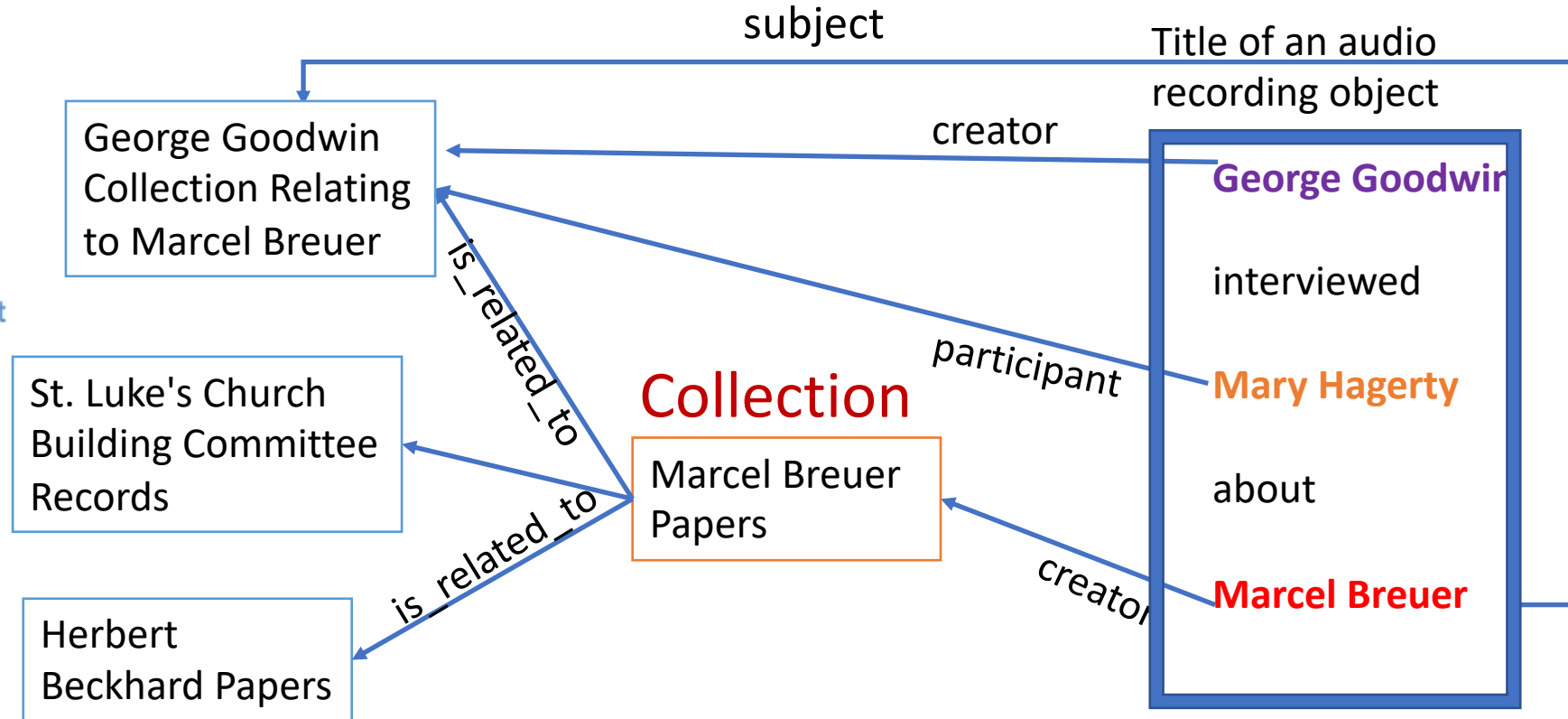- Processor

Object level:
- Subject
- Publisher

Collection — has → Object — has → Item

Collection → Agent
Object → Agent
Person — is-a-kind-of → Agent
Organization — is-a-kind-of → Agent

Any class in the model may serve as subject

OCLC FAST
LCSH
AAT

Subject

is_about

Collection

has

Object

has

Item

Mapping between model and metadata

Collections

Collection

Event as a type of subject

collection_id: 185
object_id: 56875
object_type: 1
time_stamp_created: 2015-07-13 09:16:15
time_stamp_updated: 2015-07-13 09:16:15
time_stamp_exported: 0000-00-00 00:00:00
object_draft: 0
object_deleted: 0
title: Nightline: Iran: Day 444
title_alt:
location:
date_original_range: 1981-01-19-
date_original_display: 1981-01-19
date_issued:
date_issued_display:
description: <p>"It is now day 444 for the hostages in Iran. Last minute financial problems still delay the release of the hostages. On this special edition of Nightline, we'll have the details of what's causing the hold-up; we'll have reports from Washington and West Germany; and we'll talk live to a former U.S. ambassador to Iran and with other experts about this final act in the hostage drama." Includes commercials.</p> Nightline

Event name: Iran hostage crisis
Date range: 1980-1981
Person: (lots)
Organization: (countries, national agencies, and other institutions)

100+ Nightline TV programs

Transforming metadata descriptions into Linked Data

- Person class
- Organization class } LC or local authority names enriched by role
- Subject class: OCLC FAST categories provide a great source for mapping subject terms in the metadata already created
- Geolocation class: can be enriched by GIS data and LC Geographic Areas terms
- Collection, object, and item class: the identifiers will be indicative of each of these levels

```
<!-- http://linkedarchive.syr.edu/collection/object/12983 -->
<owl:NamedIndividual rdf:about="http://linkedarchive.syr.edu/collection/object/12983">
    <rdf:type rdf:resource="http://linkedarchive.syr.edu/collection/object/"/>
    <rdfs:label>Nightline: Iran Hostage Crisis: Day 142</rdfs:label>
</owl:NamedIndividual>

<!-- http://linkedarchive.syr.edu/collection/object/56770 -->
<owl:NamedIndividual rdf:about="http://linkedarchive.syr.edu/collection/object/56770">
    <rdf:type rdf:resource="http://linkedarchive.syr.edu/collection/object/"/>
    <property:is_related_to rdf:resource="http://id.loc.gov/authorities/subjects/sh85067917"/>
    <rdfs:label>Nightline: Iran: Day 149</rdfs:label>
</owl:NamedIndividual>

    <!-- http://linkedarchive.syr.edu/person/12153 -->
    <owl:NamedIndividual rdf:about="http://linkedarchive.syr.edu/person/12153">
        <rdf:type rdf:resource="http://linkedarchive.syr.edu/person/"/>
        <property:is_related_to rdf:resource="http://id.loc.gov/authorities/subjects/sh85067917"/>
        <property:is_related_to rdf:resource="http://linkedarchive.syr.edu/collection/object/12983"/>
        <property:role>TV host</property:role>
        <rdfs:label>Koppel, Ted</rdfs:label>
    </owl:NamedIndividual>

    <!-- http://linkedarchive.syr.edu/person/12530 -->
    <owl:NamedIndividual rdf:about="http://linkedarchive.syr.edu/person/12530">
        <rdf:type rdf:resource="http://linkedarchive.syr.edu/person/"/>
        <property:is_related_to rdf:resource="http://id.loc.gov/authorities/subjects/sh85067917"/>
        <property:is_related_to rdf:resource="http://linkedarchive.syr.edu/collection/object/12983"/>
        <property:role>Reporter</property:role>
        <rdfs:label>Kashiwahara, Ken</rdfs:label>
    </owl:NamedIndividual>

<!-- http://linkedarchive.syr.edu/person/38696 -->
<owl:NamedIndividual rdf:about="http://linkedarchive.syr.edu/person/38696">
    <rdf:type rdf:resource="http://linkedarchive.syr.edu/person/"/>
    <property:is_related_to rdf:resource="http://id.loc.gov/authorities/subjects/sh85067917"/>
    <property:is_related_to rdf:resource="http://linkedarchive.syr.edu/collection/object/12983"/>
    <property:role>Interviewee</property:role>
    <rdfs:label>Morefield, Dorothea</rdfs:label>
</owl:NamedIndividual>
```

Video object identified with an IRI

"Iran Hostage Crisis, 1979-1981" Subject term from LCSH for the physical object

Local IRI assigned to the person

Subject "sh85067917"

is_about

Person 12530 — is_related_to → Object 12983

Person 38696 — is_related_to → Object 12983

23

Conclusion and next step

- There are various ontological models for linked data transformation, but not a single one can meet all of our requirements

- Reuse of classes and properties in relevant models needs further exploration

- Transformation from relational database to linked data format requires careful mapping, likely refining the model

- Linked archives is an effort of reusing components from existing models for enriching the semantics in metadata descriptions to achieve the goal of discoverability and interactivity

Thank you! Questions?

School of Journalism   Syracuse University