

Disambiguating Descriptions:

Mapping Digital Special Collections Metadata into Linked Open Data (LOD) Formats

Jacob Jett, Myung-Ja Han, Timothy W. Cole, Maria Janina Sarol

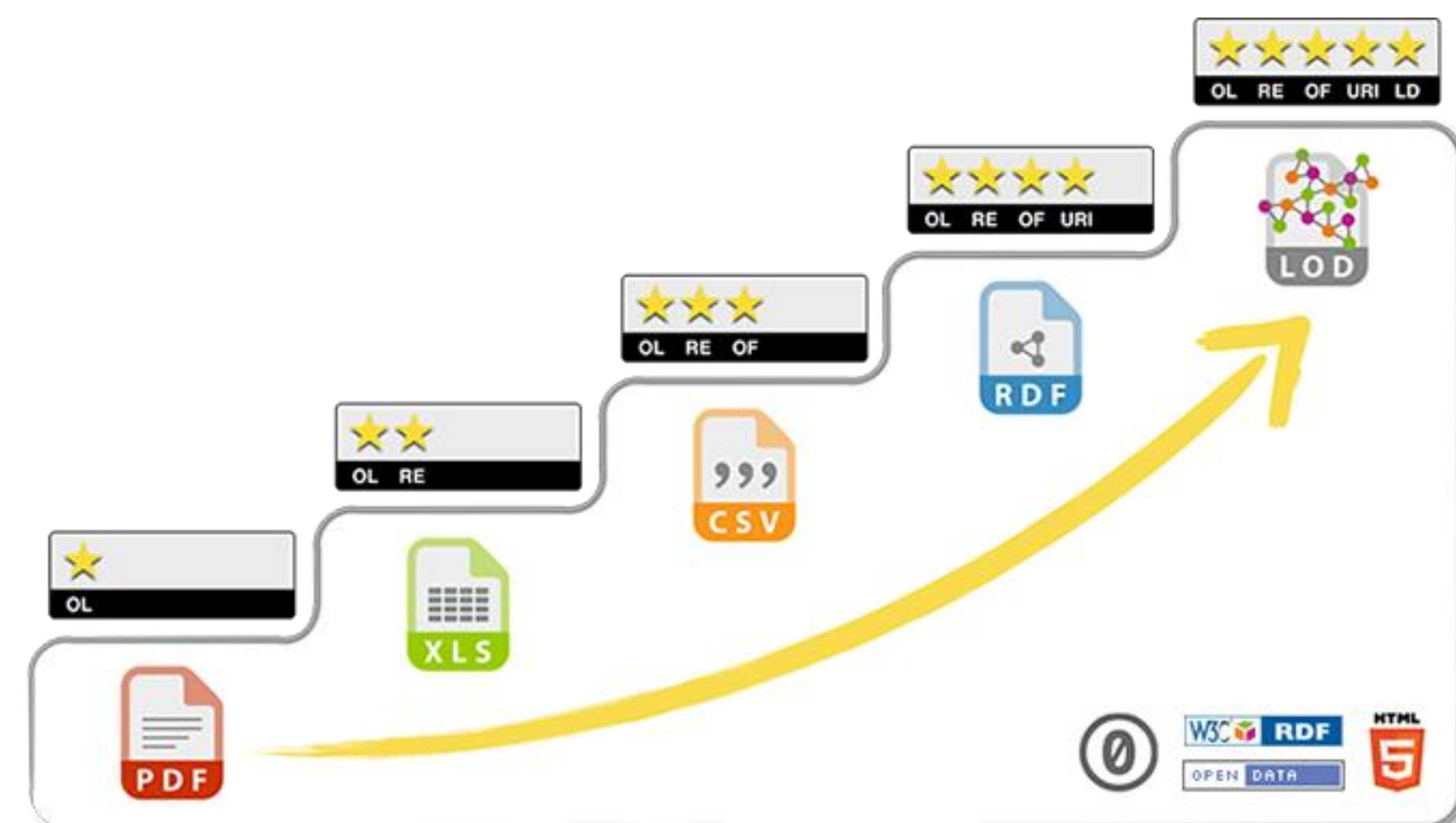
Center for Informatics Research in Science & Scholarship, School of Information Science, University of Illinois at Urbana-Champaign

LOD for Digital Special Collections

- Mellon-funded 20-month Exploratory Project
- Project Goals
 - Enrich existing metadata with links
 - Map metadata into LOD vocabularies
- Realize benefits of LOD
 - Enhance discoverability of special collection items
 - Allowing users access to contexts beyond the silo
 - Integrate special collections into semantic web
- Exemplar Case
 - Motley Collection of Theatre & Costume Design
<http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/motley>

Technological Context – LOD

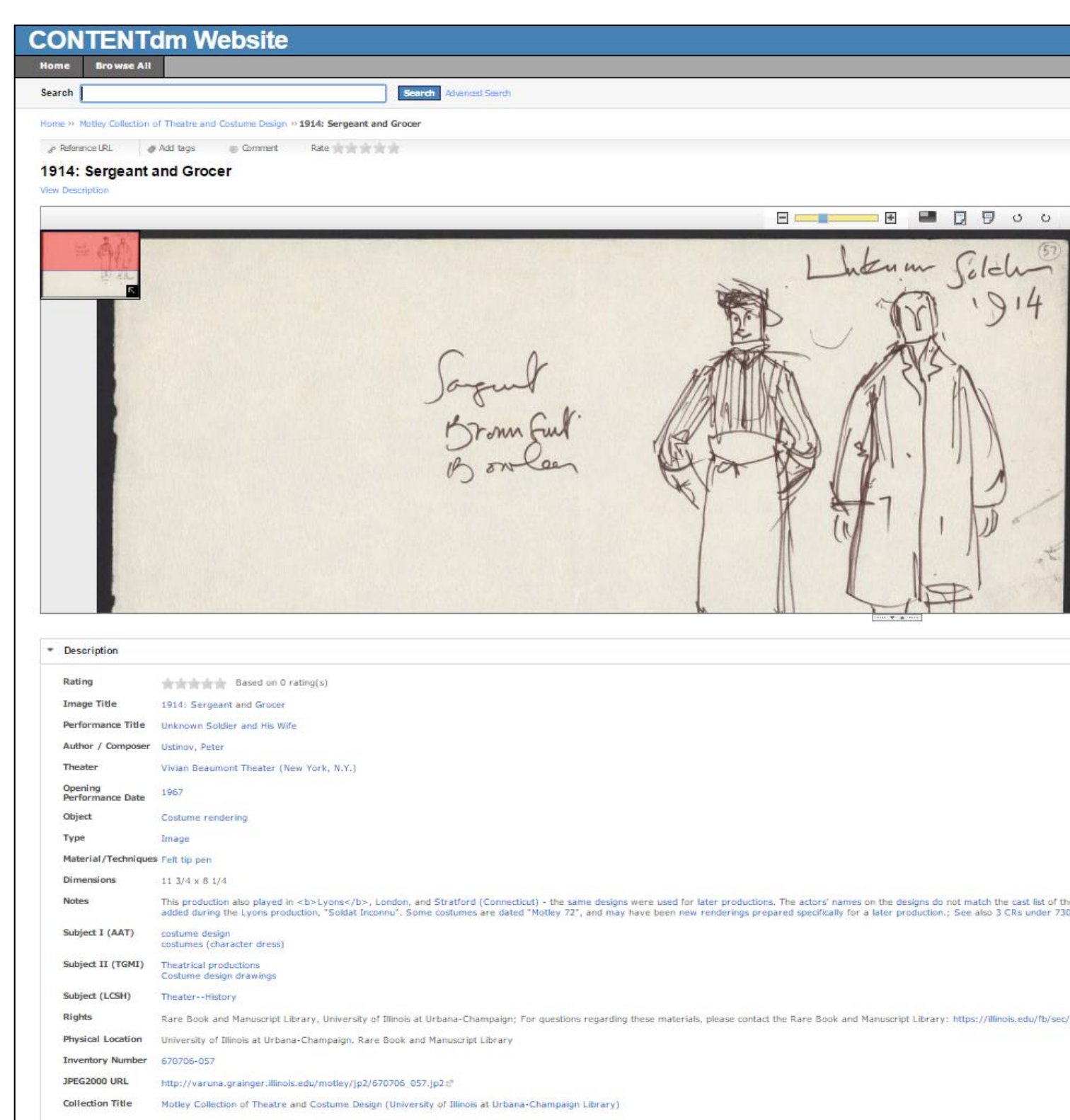
- Confluence of:
 - Linked Data (Berners-Lee 2006)
 - Open Data (Auer et al. 2007)
- Based on the Semantic Web Stack
 - Based on the W3C's Resource Description Framework standard
 - 5-Star Deployment Scheme (below)



Images courtesy of:

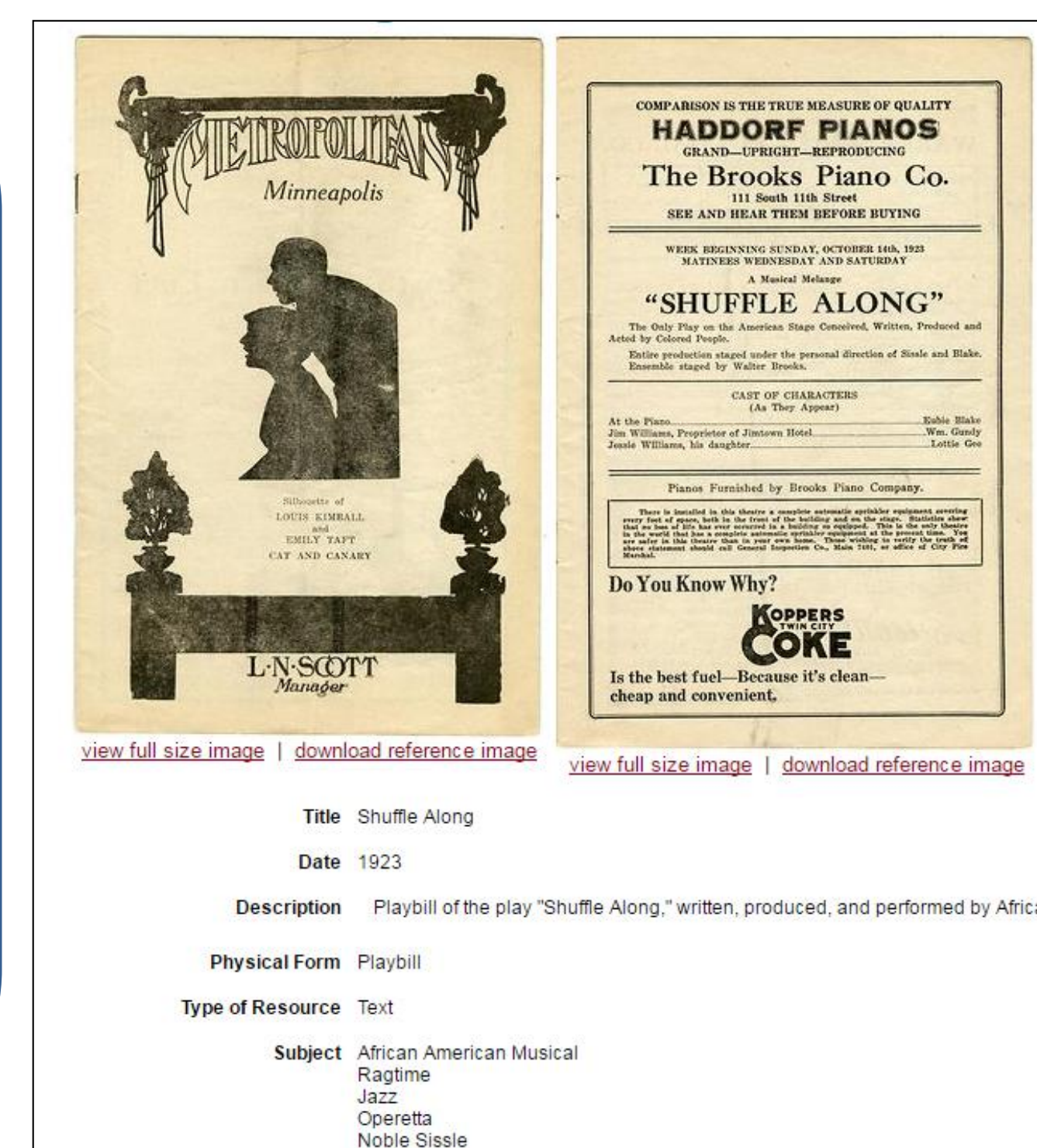
- 5-star Open Data (5stardata.info/)
- Motley Collection of Theatre & Costume Design (<http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/motley>)
- University of Minnesota (<http://umedia.lib.umn.edu/node/780429>)
- New York Public Library Digital Collections (<http://digitalcollections.nypl.org/items/510d47da-7956-a3d9-e040-e00a18064a99#>)

Illustrated Challenges



- Entity Conflation (i.e., violation of DC's 1-to-1 principle)
- Missing Entity Types (e.g., Stage Productions)
- Extensions of the target vocabulary needed.

- Property Conflation (e.g., the play's author is a topic of the play?)
- Missing metadata (i.e., not enough information from the resource has been transcribed by the cataloger)



A photograph of a woman, Miss Maie Ash, in a white, long-sleeved dress. To the right of the photo is a metadata sidebar with the following information:

- TITLE:** Miss Maie Ash.
- COLLECTION:** Cigarette cards; Actors and Actresses [Prize Crop Cigarettes].
- DATES / ORIGIN:** Place: U.K.
- LIBRARY LOCATIONS:** George Arents Collection; Shelf locator: Arents Cigarette Cards; Shelf locator: Arents Cigarette Cards 2.
- TOPICS:** Prize Crop Cigarettes; Actors.
- GENRES:** Cigarette cards; Trade cards.
- NOTES:** Citation/reference: Cartophilic reference books ZJ2-1 and ZJ3-1; Content: Mitchell; Prize Crop Cigarettes.

- Missing Provenance/Context (e.g., "item sold in / comes with cigarette packs")
- Missing Entities (e.g., the Cigarette Pack)
- Missing Properties (e.g., "includes" or "included-With")
- Extensions of the target vocabulary needed.

Preliminary Mapping for the Motley Collection of Theatre & Costume Design

Field Name	Mapping to schema.org – schema:VisualArtwork
Image Title	schema:name (Text)
Object	schema:genre (Text)
Type	schema:artform (Text or URL)
Material/Techniques	schema:artMedium (Text or URL)
Dimensions	schema:height & schema:width (schema:Distance or schema:QuantitativeValue)
Subject I (AAT)	schema:about (schema:Thing)
Subject II (TGMI)	schema:about (schema:Thing)
Subject III (LCSH)	schema:about (schema:Thing)
Rights	schema:copyrightHolder (schema:Organization or schema:Person)
Physical Location	schema:provider (schema:Organization or schema:Person)
Inventory Number	spc:standardNumber (Text or URL)
JPEG 2000 URL	schema:associatedMedia (schema:CreativeWork)
Collection Title	schema:isPartOf (schema:Collection)
[Design by]	schema:creator (schema:Organization) [always Motley in this case]
[is part of Stage Production]	schema:isPartOf (schema:CreativeWork, spc:StageWork)
Field Name	Mapping to schema.org – schema:CreativeWork
Performance Title	schema:name
Theatre	schema:locationCreated (schema:Place)
Opening Performance	schema:dateCreated (Date)
Date	
Notes	schema:description or schema:mainEntityOfPage
[additional type]	schema:additionalType (URL) [spc:StageWork]
[production of]	schema:exampleOfWork (schema:Book, fabio:Play)
Field Name	Mapping to schema.org – schema:Book
Author/Composer	schema:author (schema:Person)
[additional type]	schema:additionalType (URL) [http://purl.org/spar/fabio/Play]
[Published Work]	schema:name
[publication date]	schema:datePublished (Date)
[part of]	schema:isPartOf (schema:CreativeWorkSeries) [when true]
[adaptation of]	schema:exampleOfWork (schema:Book or schema:CreativeWork) [when true]

References:

- Auer, S. R., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A Nucleus for a Web of Open Data. The Semantic Web. *Lecture Notes in Computer Science* 4825, p 722. doi:10.1007/978-3-540-76298-0_52.
- Berners-Lee, T. (2006). Linked Data. Design Issues. [revised 2009]. Retrieved from: <https://www.w3.org/DesignIssues/LinkedData.html>