

Digital Scriptorium 2.0: Toward a community-driven LOD knowledge base and national union catalog for premodern manuscripts

L.P. Coladangelo, College of Communication and Information, Kent State University Lynn Ransom, Schoenberg Institute for Manuscript Studies, Penn Libraries Doug Emery, Penn Libraries

> Semantic Web in Libraries 2022 (SWIB22) December 2, 2022



Agenda

Background Manuscript Metadata Previous Database Overview of DS 2.0 Data Model Overview of DS 2.0 Workflows Decision-making toward Implementation Semantic Enrichment Next Steps



Digital Scriptorium and the Need for Redevelopment





Manuscript Description

Manuscripts are unique objects (Bair & Steuer, 2013)

Requires description of contents physical properties context around creation and provenance

Knowledge about manuscripts can be disparate and debated

Requires extensive scholarship Revisions in descriptions occur as more is learned

Imperfect reliance on bibliographic standards

No uniformly agreed upon standards for manuscript descriptions Barriers to institutions without expertise and resources to implement bibliographic description standards











Principles for the DS 2.0 Redevelopment

Minimal data entry standards

manuscript location in an institution is sufficient

Members manage their own metadata (formats and values)

DS 2.0 uses what is provided and does not correct or add to member metadata

DS 2.0 does not host images

primary goal is initial discovery and metadata aggregation but IIIF functionality is possible with provided URIs for manifests

DS 2.0 will provide semantic enrichment

metadata is enhanced, where possible, by reconciling values with external authorities as well as linking to internal Name and Manuscript ID Authorities

Open access

DS 2.0 data will be open and available for reuse



DS 2.0 Data Model

Backbone of main classes of entities (Koho et al., 2022)
 Manuscript (physical manuscript object)
 DS 2.0 Record (metadata description of a Manuscript)
 Holding (metadata description of holding information about a Manuscript)

Separates the physical object from its catalog record / metadata description Conceptualizes the institutional metadata record as a cataloger's observation of a manuscript object

Can be revised / overwritten as changes in description and scholarship warrant Physical object retains a unique and persistent DS ID despite description and holdings changes



DS 2.0 Data Model



(Koho et al., 2022)

approvolutione aquerto e corozante

"items": { "Manuscript": "Q1", "Holding": "02", "DS 2.0 Record": "Q3", "Current": "04", "Non-Current": "05", "Uniform Title (Authority File)": "06", "Actor (Name Authority)": "Q7", "Personal (Name Authority)": "Q8", "Corporate (Name Authority)": "Q9", "Role (Authority File)": "010", "Genre (Authority File)": "011", "Subject (Authority File)": "012", "Term (Authority File)": "Q13", "Language (Authority File)": "Q14", "Century (Authority File)": "Q15", "Datable": "Q16", "Non-datable": "017", "Place (Authority File)": "018", "Material (Authority File)": "019"

1. DS ID (P1) 2. manuscript holding (P2) 3. described manuscript (P3) 4. holding institution (P4) 5. holding institution as recorded (P5) 6. holding status (P6) 7. Institutional ID (P7) 8. shelfmark (P8) 9. link to institutional record (P9) 10. title as recorded (P10) 11. uniform title in authority file (P11) 12. uniform title as recorded (P12) 13. in original script (P13) 14. associated name as recorded (P14) 15. role (P15) 16. instance of (P16) 17. name in authority file (P17) 18. genre as recorded (P18) 19. subject as recorded (P19) 20. genre in authority file (P20) 21. subject in authority file (P21) 22. term in authority file (P22) 23. language as recorded (P23) 24. language in authority file (P24) 25. production date as recorded (P25)

26. production century in authority file (P26) 27. century (P27) 28. dated (P28) 29. production place as recorded (P29) 30. place in authority file (P30) physical description (P31) 32. material as recorded (P32) 33. material (P33) 34. note (P34) 35. acknowledgements (P35) 36. date added (P36) 37, date last updated (P37) 38. latest date (P38) 39. earliest date (P39) 40. place (P40) 41. start time (P41) 42. end time (P42) 43. external identifier (P43) 44. AAT ID (P44) 45. IIIF manifest (P45) 46. Wikidata QID (P46) 47. VIAF ID (P47) 48. External URI (P48) 49. equivalent property (P49) 50. formatter URL (P50)



DS 2.0 Data Model as implemented





DS 2.0 Workflow





Set swgLogos th the URL ith to your own logo image	Item Discussion	ntary o	n the Pauline	Read View hister	Dsadmin Talk Preferen ory ★ More ✓ S 878) (Q3522	ces Watchlist Contributions Log or Gearch wikibase-docker Q		
n page ent changes dom page	Manuscript metad	data collected	d by UPenn from University	y of Kansas (MS 9/1:B10)				
about MediaWiki	Language	Label		Description		Also known as		
S	English Comm		entary on the Pauline	Manuscript metadata collected by UPenn from				
ited changes cial pages table version	Statements							
e information cept URI her languages & Add links	language as rec	corded	 Latin language in authori 0 references 	redit Latin ces		ference		
					+ add re	nerence		
					+ add va	alue		
	physical descrip	otion	Extent: f. 1r-v; 325	x 237 mm.	sedit 🖍			

Mai Rec Bar

Too Wh Rel Spe Prir Per Pag Cor

and absonate la micta zeoula:

/ edit

+ add r

associated name as recorded Peter Lombard role Author name in authority file Peter Lombard • 0 references

Peter Lombard (Q315347)

medieval priest and theologian

Master of Sentences

✓ In more languages

Configure

Language	Label	Description	Also H
English	Peter Lombard	medieval priest and theologian	Maste
Spanish	Pedro Lombardo	No description defined	Petrus
Traditional Chinese	No label defined	No description defined	
Chinese	伦巴第人彼得	No description defined	彼得·
All entered language	S		
Statements			
instance of	🚊 human		sedit

▶ 1 reference

Peter Lombard (Q1015)

No description defined

In more languages

Configure

Language	Label	Description
English	Peter Lombard	No description defined

Statements

DS ID	DSNAME251
Wikidata QID	Q315347



LOD Vocabularies for Semantic Enrichment

Centuries

Getty Art and Architecture Thesaurus (AAT) Places Getty Thesaurus of Geographic Names (TGN)

Languages Wikidata

Materials AAT (controlled list)

People, Organizations Wikidata

Roles

Controlled list

Terms (Subject and Genre)

AAT OCLC Faceted Application of Subject Terminology (FAST) Ligatus Language of Binding Terms (LoBT)



Guiding Choices for Semantic Enrichment

Consider only vocabularies published as LOD

Leaving values unreconciled if an appropriate match cannot be found

Match only where appropriate as indicated by the data

e.g., MARC subfields for the source for genre terms

Use vocabularies known to GLAM communities

e.g., sources for places were not identified, but TGN was used

Leverage the value and flexibility of linked data

e.g., LC derived-strings and the choice to use FAST; Wikidata for languages and their ISO codes

Address limitations in existing authorities

e.g., use of Wikidata instead of LCNAF or VIAF

Implement in a manageable data processing workflow

Using reconciliation services available through OpenRefine



Open

Documentation

\leftrightarrow C github.com/DigitalScriptorium 🐠 🛈 🗯 🖬 🌏 🗄 10 \$ ⊉ + • 👧 • Search or jump to ... Pull requests Issues Marketplace Explore **Digital Scriptorium** Follow R 1 follower 🖉 https://digital-scriptorium.org/ 🔰 @dscriptorium2_0 Repositories 3 Projects Packages A Teams A People 5 🐼 Settings Overview Popular repositories O View as: Public ▼ You are viewing the README and pinned repositories ds-testing Public ds-open-refine Public as a public user. Repository for testing Digital Scriptorium instances Digital Scriptorium OpenRefine documentation and JSON recipes for data You can create a README file or pin repositories reconciliation visible to anyone. 2 2 \$1 Get started with tasks that most successful organizations complete. ds-data Public Digital Scriptorium repository for metadata and CSV files People Ruby 8 Invite someone Repositories Rew New Q Find a repository... Туре 👻 Sort -Language 👻

https://github.com/DigitalScriptorium



\leftarrow \rightarrow C \square github.com/DigitalScriptorium/ds-data/blob/main/terms/reconciled/places.csv

🖻 🖈 👎 🕕 🗯 🖬 👰 🗄

Metadata Repositories for Reconciled Values

	gittide.com/ergitalsenptonam/ds-d	ata/biob/main/terms/reconciled/place		
48	France	France	http://vocab.getty.edu/tgn/1000070	
49	France (?)	France	http://vocab.getty.edu/tgn/1000070	
50	France (Normandy)	France;Normandy	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7002886	
51	France (Paris)	France;Paris	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7008038	
52	France or Flanders	France;Flanders	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7024097	
53	France, Anjou?	France;Anjou	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7017069	
54	France, Brittany	France;Brittany	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7024267	
55	France, Dauphiné	France;Dauphiné	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7007764	
56	France, Franche Comté (?)	France;Franche-Comté	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7645440	
57	France, Normandy	France;Normandy	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7002886	
58	France, Northern	Northern France	http://vocab.getty.edu/tgn/4004036	
59	France, Picardy	France;Picardy	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7002890	
60	France, probably between 1425 and 1450	France	http://vocab.getty.edu/tgn/1000070	
61	France?	France	http://vocab.getty.edu/tgn/1000070	
62	France? Netherlands?	France;Netherlands	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7016845	
63	France?, Champagne?	France;Champagne	http://vocab.getty.edu/tgn/1000070;http://vocab.getty.edu/tgn/7658604	
64	Germany	Germany	http://vocab.getty.edu/tgn/7000084	
65	Germany (?)	Germany	http://vocab.getty.edu/tgn/7000084	
66	Germany (Nuremberg?)	Germany;Nuremberg	http://vocab.getty.edu/tgn/7000084;http://vocab.getty.edu/tgn/7004334	
67	Germany, Bavaria	Germany;Bavaria	http://vocab.getty.edu/tgn/7000084;http://vocab.getty.edu/tgn/7003669	
68	Germany, Bavaria or west Austria?	Germany;Bavaria;Austria	http://vocab.getty.edu/tgn/7000084;http://vocab.getty.edu/tgn/7003669;http://vocab.getty.edu/tgn/1000062	

https://github.com/DigitalScriptorium/ds-data/tree/main/terms/reconciled/places.csv



← → C 🔒 github.com/DigitalScriptorium/ds-data/blob/main/terms/reconciled/genres.csv

🖻 🖈 🐠 🕀 🗯 🗖 🥷

Metadata Repositories for Reconciled Values

genre_as_recorded	vocabulary	course authorit	the second s	
	,	source_authority_un	authorized_label	structured_value
Abridgements	aat		abridgements	http://vocab.getty.edu/aat/300202489
Accounts	aat		accounts	http://vocab.getty.edu/aat/300145802
Amulets	aat		amulets	http://vocab.getty.edu/aat/300266585
Annals and chronicles	lcgft		Annals and chronicles	http://id.worldcat.org/fast/1919892
Annotations	aat		annotations	http://vocab.getty.edu/aat/300026100
Anthologies	aat		anthologies	http://vocab.getty.edu/aat/300026037
Art	fast		Art	http://id.worldcat.org/fast/815177
Artificial grains	lobt		artificial grains	http://w3id.org/lob/concept/1507
Blind tooled (decoration)	lobt		blind-tooled (decoration)	http://w3id.org/lob/concept/2288
Books of hours	aat		books of hours	http://vocab.getty.edu/aat/300026457
Bosses	lobt		bosses	http://w3id.org/lob/concept/1230
Calendars (documents by form)	aat		calendars (documents)	http://vocab.getty.edu/aat/300026741
Choirbooks	aat		choirbooks	http://vocab.getty.edu/aat/300026422
Chronicles	aat		chronicles	http://vocab.getty.edu/aat/300026361
Clasps (components)	lobt		clasps (components)	http://w3id.org/lob/concept/1254
Codices	aat		codices (bound manuscripts)	http://vocab.getty.edu/aat/300224200
Codices (bound manuscripts)	aat		codices (bound manuscripts)	http://vocab.getty.edu/aat/300224200
Commentaries	fast		Commentaries	http://id.worldcat.org/fast/1423723
Commentaries	aat		commentaries	http://vocab.getty.edu/aat/300026098
Commentries	aat		commentaries	http://vocab.getty.edu/aat/300026098
	Accounts Amulets Annals and chronicles Annotations Anthologies Art Artificial grains Blind tooled (decoration) Books of hours Bosses Calendars (documents by form) Coloirbooks Chronicles Chonicles Clasps (components) Codices Codices (bound manuscripts) Codices (bound manuscripts) Commentaries Commentaries	AccountsaatAmuletsaatAnnals and chronicleslegftAnnotationsaatAnthologiesaatArtfastArtficial grainslobtBlind tooled (decoration)lobtBooks of hoursaatBooks of hoursaatCalendars (documents by form)aatChoribooksaatChorniclesaatCodices (bound manuscripts)abtCommentariesfastCommentariesaatCommentariesaatCommentariesfastCommentariesaatCommentariesaatCommentariesaatCommentariesaatCommentariesaatCommentariesaatCommentariesaatCommentariesaatCommentariesaat	AccountsaatatAmuletsaatInternationalAnnals and chronicleslogftInternationalAnnotationsaatInternationalAnthologiesaatInternationalArtfastInternationalArtificial grainslobtInternationalBlind tooled (decoration)lobtInternationalBooks of hoursaatInternationalBooks of hoursaatInternationalBooks of hoursaatInternationalBooks of hoursaatInternationalCalendars (documents by form)aatInternationalChoirbooksaatInternationalChoirbooksaatInternationalCodices (components)lobtInternationalCodices (bound manuscripts)aatInternationalCommentariesfastInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternationalCommentariesaatInternational <td< td=""><td>AccountsaataccountsAmuletsaatImmediateAnnals and chronicleslogftAnnals and chroniclesAnnotationsaatImmediateAntohologiesaatImmediateArthologiesaatImmediateArtificial grainslobtImmediateBilind tooled (decoration)lobtImmediateBooks of hoursaatImmediateBooks of hourslobtImmediateBooks of hourslobtImmediateChoirbooksaatImmediateChoirbooksaatImmediateChoirbooksaatImmediateCodices (bound manuscripts)lobtImmediateCodices (bound manuscripts)aatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediate</td></td<>	AccountsaataccountsAmuletsaatImmediateAnnals and chronicleslogftAnnals and chroniclesAnnotationsaatImmediateAntohologiesaatImmediateArthologiesaatImmediateArtificial grainslobtImmediateBilind tooled (decoration)lobtImmediateBooks of hoursaatImmediateBooks of hourslobtImmediateBooks of hourslobtImmediateChoirbooksaatImmediateChoirbooksaatImmediateChoirbooksaatImmediateCodices (bound manuscripts)lobtImmediateCodices (bound manuscripts)aatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediateCommentariesaatImmediate

https://github.com/DigitalScriptorium/ds-data/tree/main/terms/reconciled/genres.csv







Next Steps for DS 2.0

Move from Prototype to Beta Version

Taking in all legacy data (existing METS) and member data (MARC, TEI, CSV, etc.)

Development of a User Interface

Traditional online catalog interface for existing users

Ongoing analysis of Data Quality

Continuing to look at correctness and completeness of metadata **Providing Services to Users**

e.g., example SPARQL queries, workshops on Wikibase



Thank you!

https://digital-scriptorium.org/

Twitter: dscriptorium2_0

L.P. Coladangelo Icoladan@kent.edu @lpc359