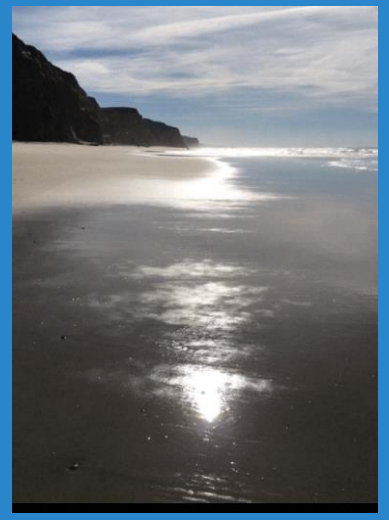
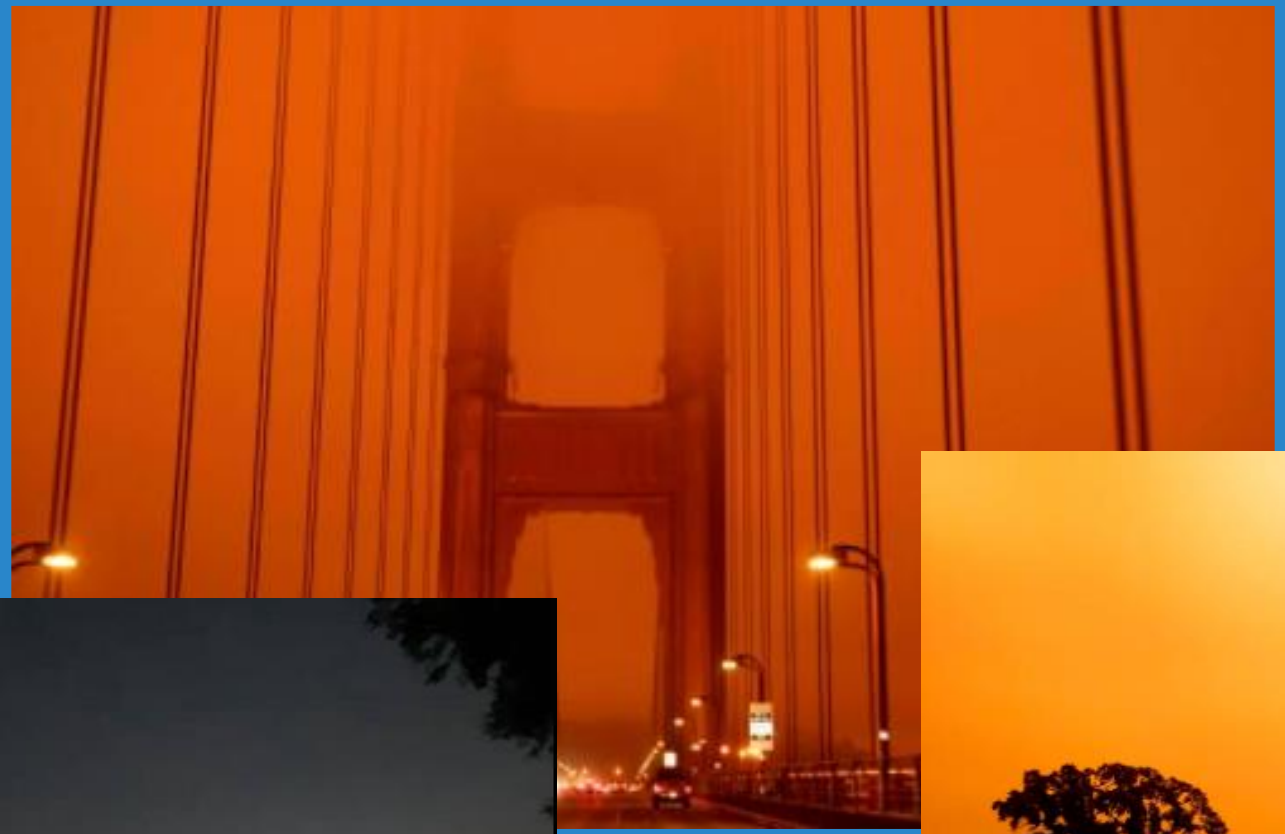
Widening our circles

Linked data, MARC, and the transformation of library metadata

Nancy Lorimer
Stanford University Libraries
ANZREG 2020



HISTORIC
MURPHY
AVENUE



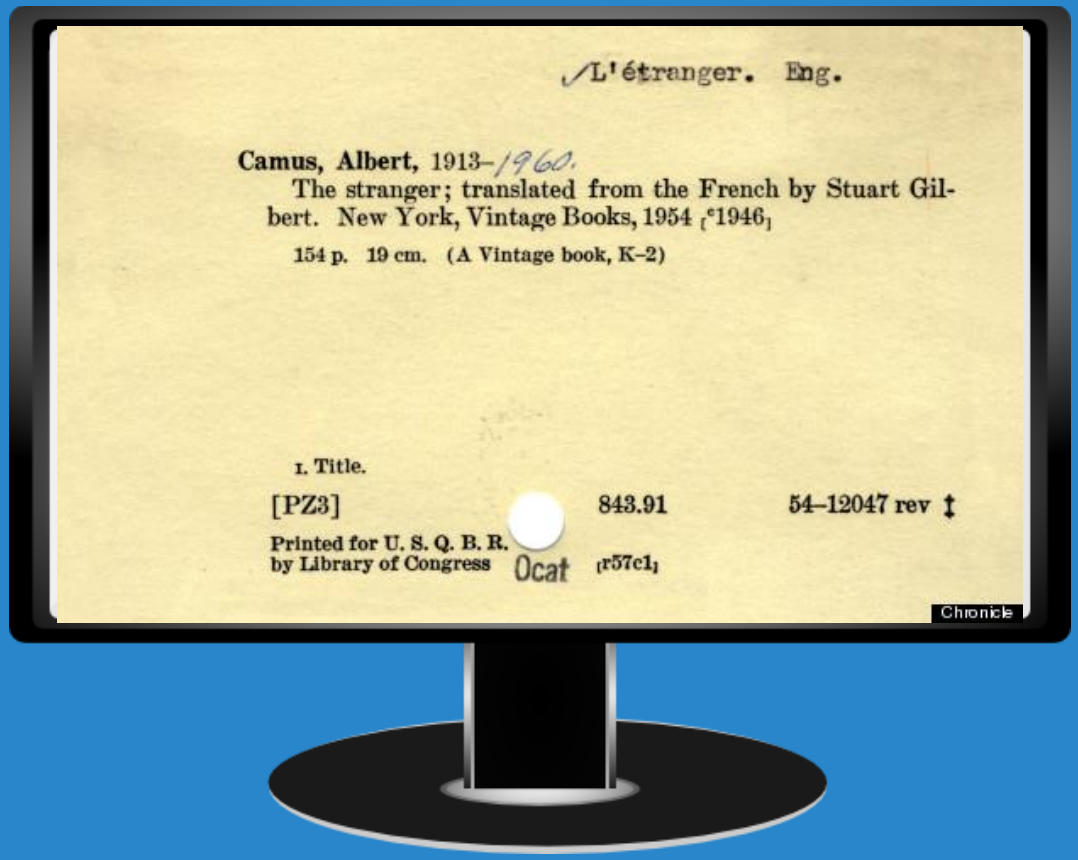
The world enriched with library data;

libraries enriched with the world's data.

LD4 Community Vision Statement

What About MARC?

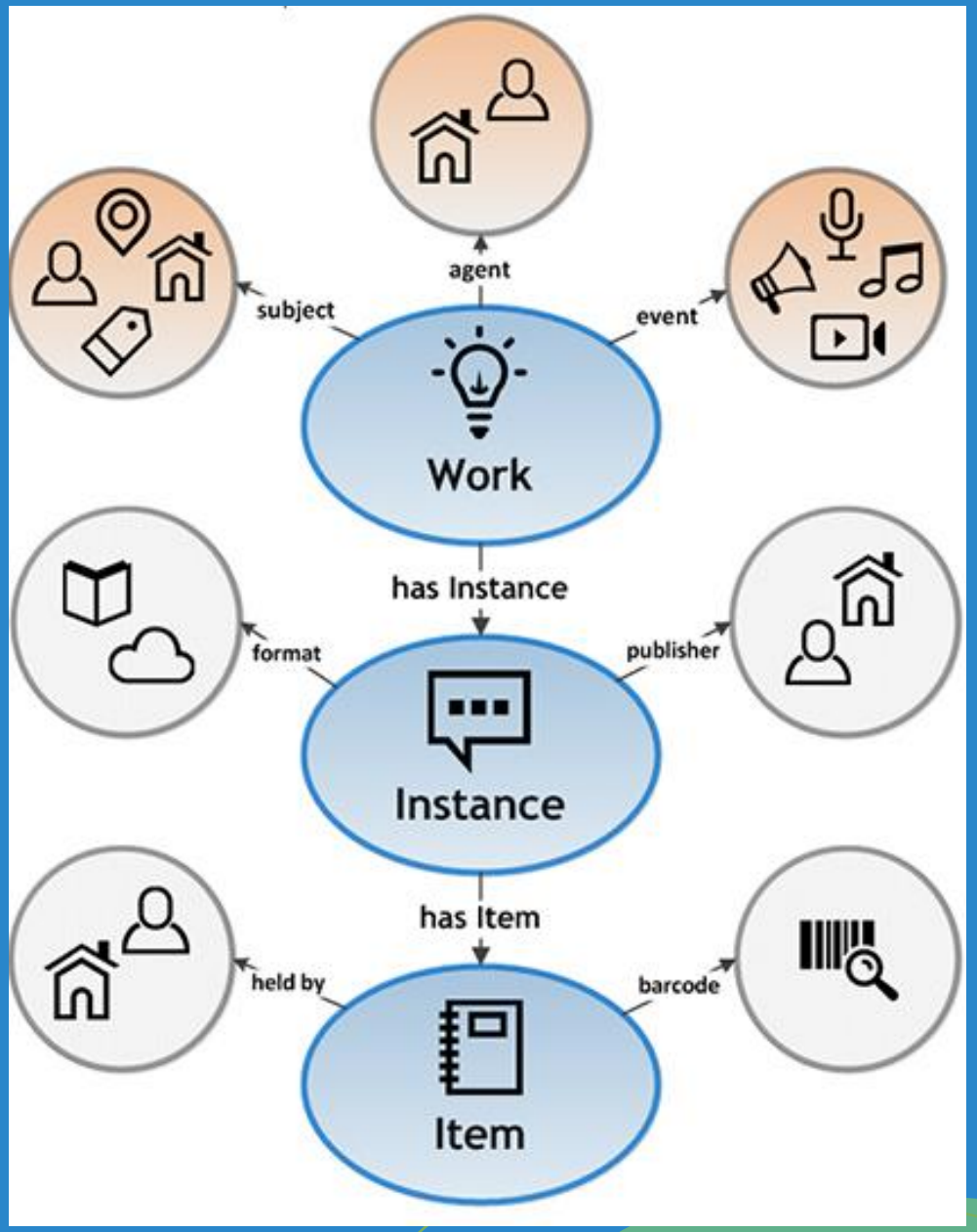
- *The MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form.* (https://www.loc.gov/marc/)
  - Developed in the 1960s by Henriette Avram for the Library of Congress as a metadata transmission standard for library metadata
  - 1971—became the standard for dissemination of bibliographic data in the US
  - By 1973—became an IOS standard
  - Several “flavours” have come and go
    - Currently MARC21 and UNIMARC dominate


/L'étranger. Eng.
Camus, Albert, 1913–1960.
The stranger; translated from the French by Stuart Gilbert. New York, Vintage Books, 1954 [c1946]
154 p. 19 cm. (A Vintage book, K-2)
1. Title.
[PZ3]
843.91
54-12047 rev
Printed for U. S. Q. B. R.
by Library of Congress
Ocat
[r57c1]
Chronicle

024	1		794881924264
024	7		00794881924264 ǂ2 gtin-14
028	0	2	HMC 801937.38 ǂb Harmonia Mundi
028	0	0	HMC 801937 ǂb Harmonia Mundi
028	0	0	HMC 801938 ǂb Harmonia Mundi
033	1		200603-- ǂa 200810-- ǂb 6299 ǂc B3 ǂb 6299 ǂc M9
033	1	0	200603-- ǂa 200810-- ǂb 6299 ǂc B3 ǂb 6299 ǂc M9
033	0	0	200603-- ǂb 6299 ǂc B3
033	0	0	200810-- ǂb 6299 ǂc M9
041	0		ǂd fre ǂb eng ǂb fre ǂb ger ǂe eng ǂe fre ǂe ger ǂn fre ǂg eng ǂg fre ǂg ger
042			lccopycat
045	0		ǂb d2000
050	1	4	M1500.S22 ǂb A46 2009
082	0	4	782.1 ǂ2 22
090			ǂb
049			STFA
100	1		Saariaho, Kaija, ǂe composer.
245	1	2	L'amour de loin / ǂc Kaija Saariaho.
264		1	Arles [France] : ǂb Harmonia Mundi, ǂc [2009]
264		4	ǂc ℗2009
300			2 audio discs (2 hr., 1 min.) : ǂb digital, stereo, SACD ; ǂc 4 3/4 in
306			020100
336			performed music ǂb prm ǂ2 rdacontent
337			audio ǂb s ǂ2 rdamedia
338			audio disc ǂb sd ǂ2 rdacarrier
500			Harmonia Mundi: HMC 801937--HMC 801938.
500			Opera in 4 acts.
546			Sung in French.
500			Libretto by Amin Maalouf.
511	0		Ekaterina Lekhina, soprano ; Marie-Ange Todorovitch, mezzo-soprano ; Daniel Belcher, tenor ; Rundfunkchor Berlin ; Deutsches Symphonie-Orchester Berlin ; Kent Nagano, conductor.
518			Recorded Mar. 2006, Teldex Studio, Berlin, and Oct. 2008, Bavaria Musik Studios, Munich.

586			Grammy for Best Opera Recording, 2010.
586			Grawemeyer Award, 2003.
500			Super audio compact discs.
538			"SACD. These hybrid super audio CDs (stereo and multi-channel) can be played on any standard compact disc player"--Container.
500			Program notes by Stephane Roth, synopsis, biographical notes on the composer, librettist, and performers, all in English, French, and German, and libretto with English and German translations (119 p. : ill., ports.) in accompanying booklet.
508			Ed. recorded: Chester Music.
505	0		CD 1: Premier acte -- Deuxieme acte -- Troisieme acte -- CD 2:Quatrième acte -- Cinquième acte.
600	0	0	Jaufré Rudel, ǂd active 12th century ǂv Drama.
600	0	7	Jaufré Rudel, ǂd active 12th century. ǂ2 fast ǂ0 (OCoLC)fst01826918
650		0	Operas.
650		7	Operas. ǂ2 fast ǂ0 (OCoLC)fst01046191
655		4	Compact disc.
655		7	Operas ǂ2 fast ǂ0 (OCoLC)fst01726688
655		7	Drama. ǂ2 fast ǂ0 (OCoLC)fst01423879
655		7	Operas. ǂ2 lcgft
700	1		Maalouf, Amin, ǂe librettist.
700	1		Lekhina, Ekaterina, ǂe singer.
700	1		Todorovitch, Marie-Ange, ǂe singer.
700	1		Belcher, Daniel, ǂe singer.
700	1		Nagano, Kent, ǂd 1951- ǂe conductor.
710	2		Rundfunkchor Berlin, ǂe performer.
710	2		Deutsches Symphonie-Orchester Berlin, ǂe performer

BIBFRAME 2.0


agent
subject
event
Work
has Instance
format
publisher
Instance
has Item
held by
barcode
Item

MARC vs Linked Data

Linked Data

- Data elements are linked together as independent statements and not tied to a record format
- Identifiers are URIs
- Relationships are expressed as triples in the form of URIS
- Triples are machine-actionable
- Entity-based descriptions combined through relationships defined by an ontology
- Links can be made to other linked data repositories

MARC

- Data elements are linked together by the context of the database record
- Identifiers are text strings
- Relationships are expressed in free text notes, or controlled text strings & require human interpretation
- Not machine-readable or –actionable
- Combines description of works, expressions, manifestations, events
- MARC databases are silos—they cannot make use of outside metadata

What BIBFRAME is and is not

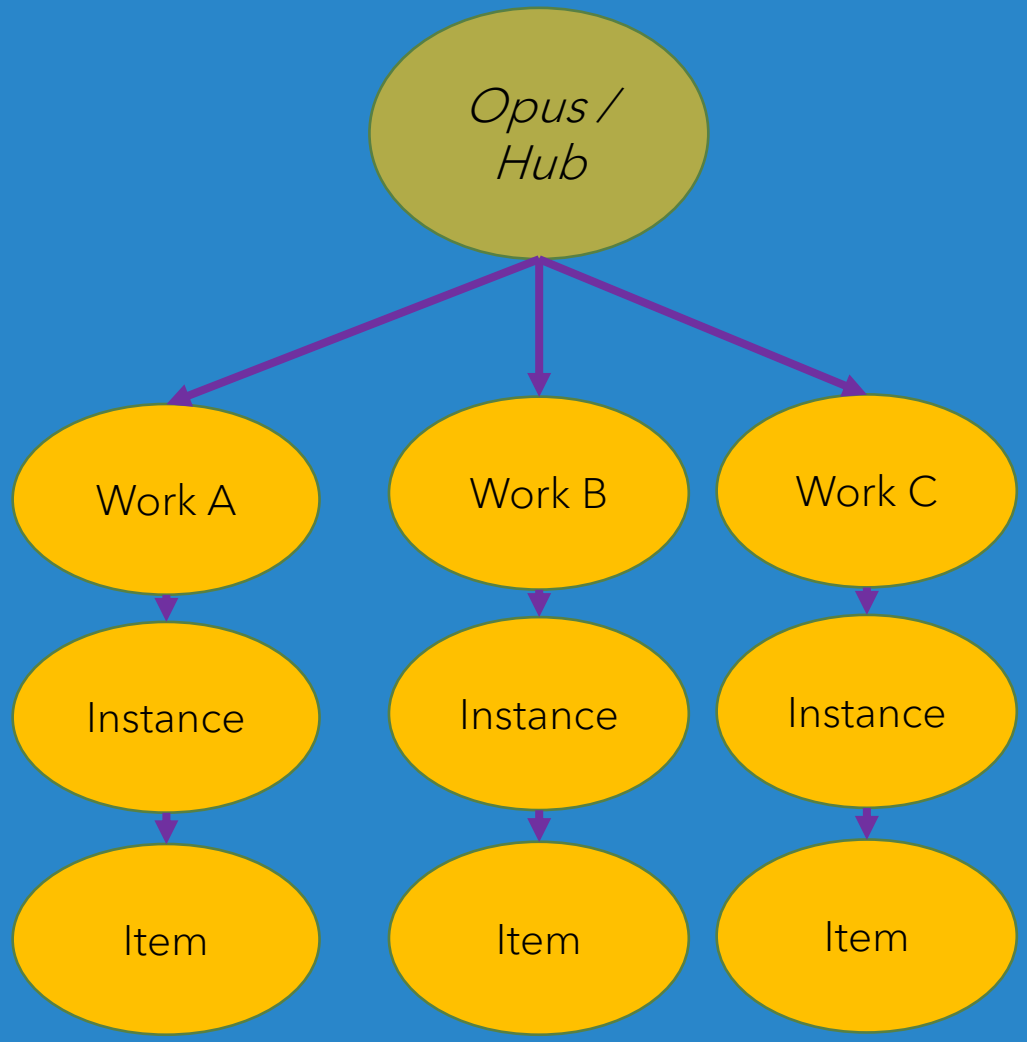
Is…

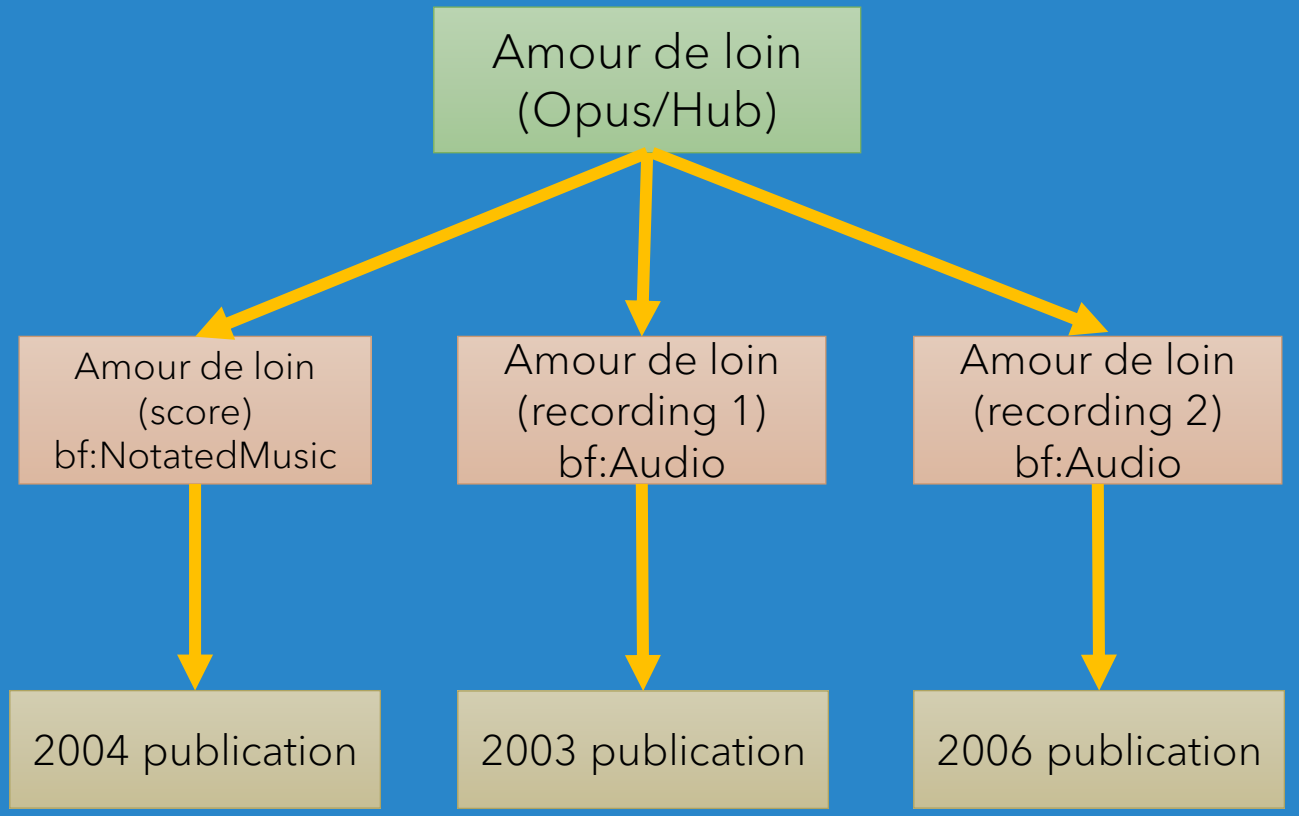
- “an initiative to evolve bibliographic description standards to a linked data model”
- a “core” ontology that covers the basics of bibliographic description as currently conceived
- a successor to MARC for cataloging
- a common framework for the creation of linked data by the library community

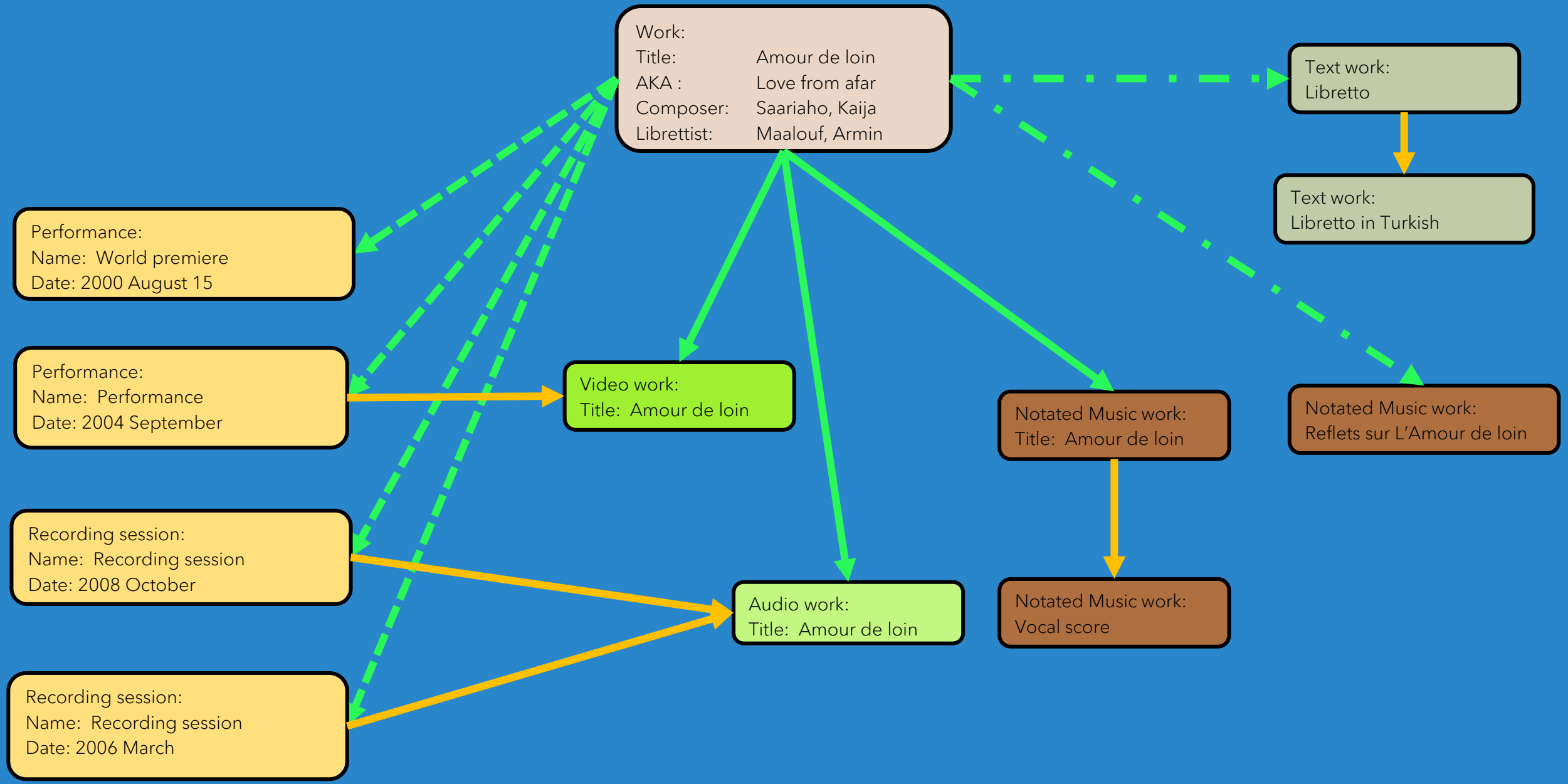
Is NOT…

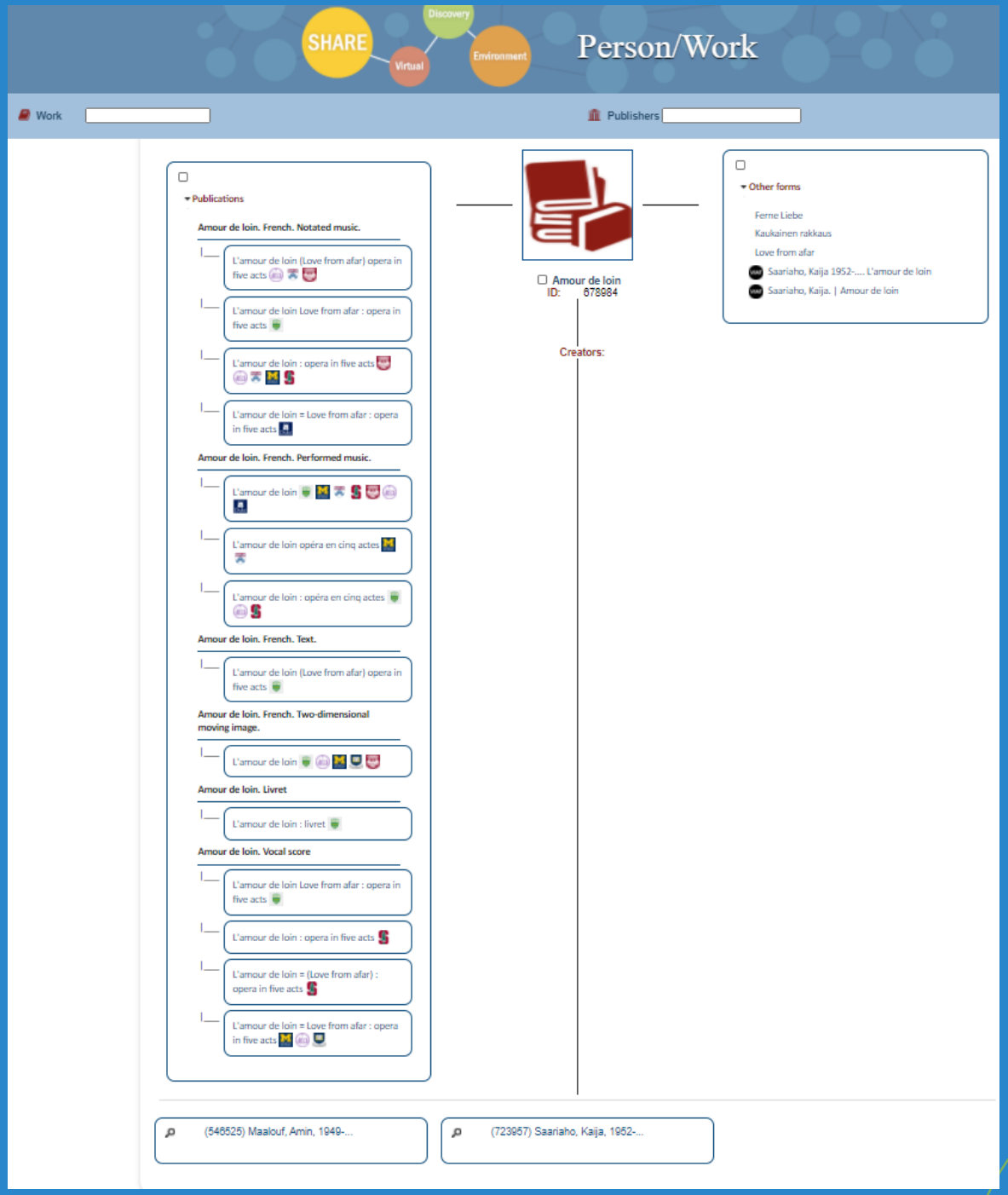
- a “replacement” for MARC
- a one-stop shop for linked data creation
  - requires additional vocabularies
  - requires extensions for domains
- your only choice…
  - linked data is all about choice!

Manipulating the Model (1)



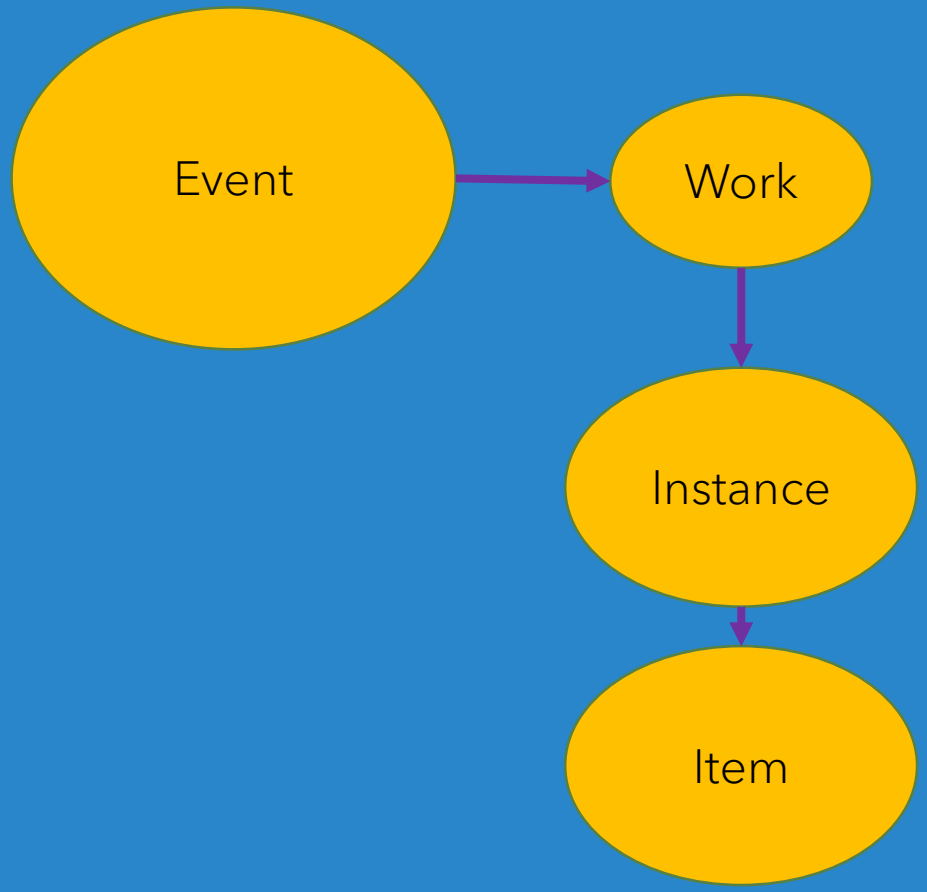


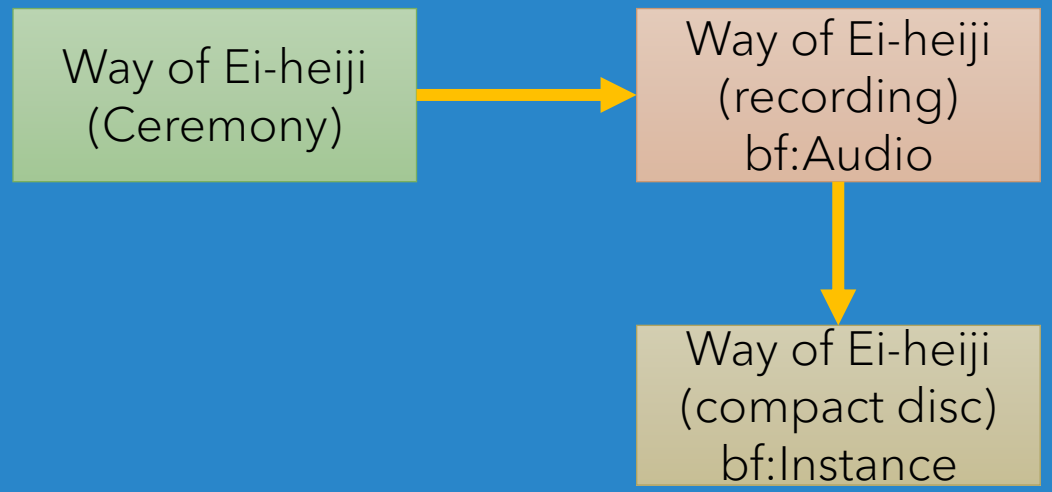

Work:
Title: Amour de loin
AKA : Love from afar
Composer: Saariaho, Kaija
Librettist: Maalouf, Armin
Text work:
Libretto
Text work:
Libretto in Turkish
Performance:
Name: World premiere
Date: 2000 August 15
Performance:
Name: Performance
Date: 2004 September
Video work:
Title: Amour de loin
Notated Music work:
Title: Amour de loin
Notated Music work:
Reflets sur L'Amour de loin
Recording session:
Name: Recording session
Date: 2008 October
Audio work:
Title: Amour de loin
Notated Music work:
Vocal score
Recording session:
Name: Recording session
Date: 2006 March



Adding another level to BIBFRAME
“Opus” or “Hub”

Manipulating the Model (2)





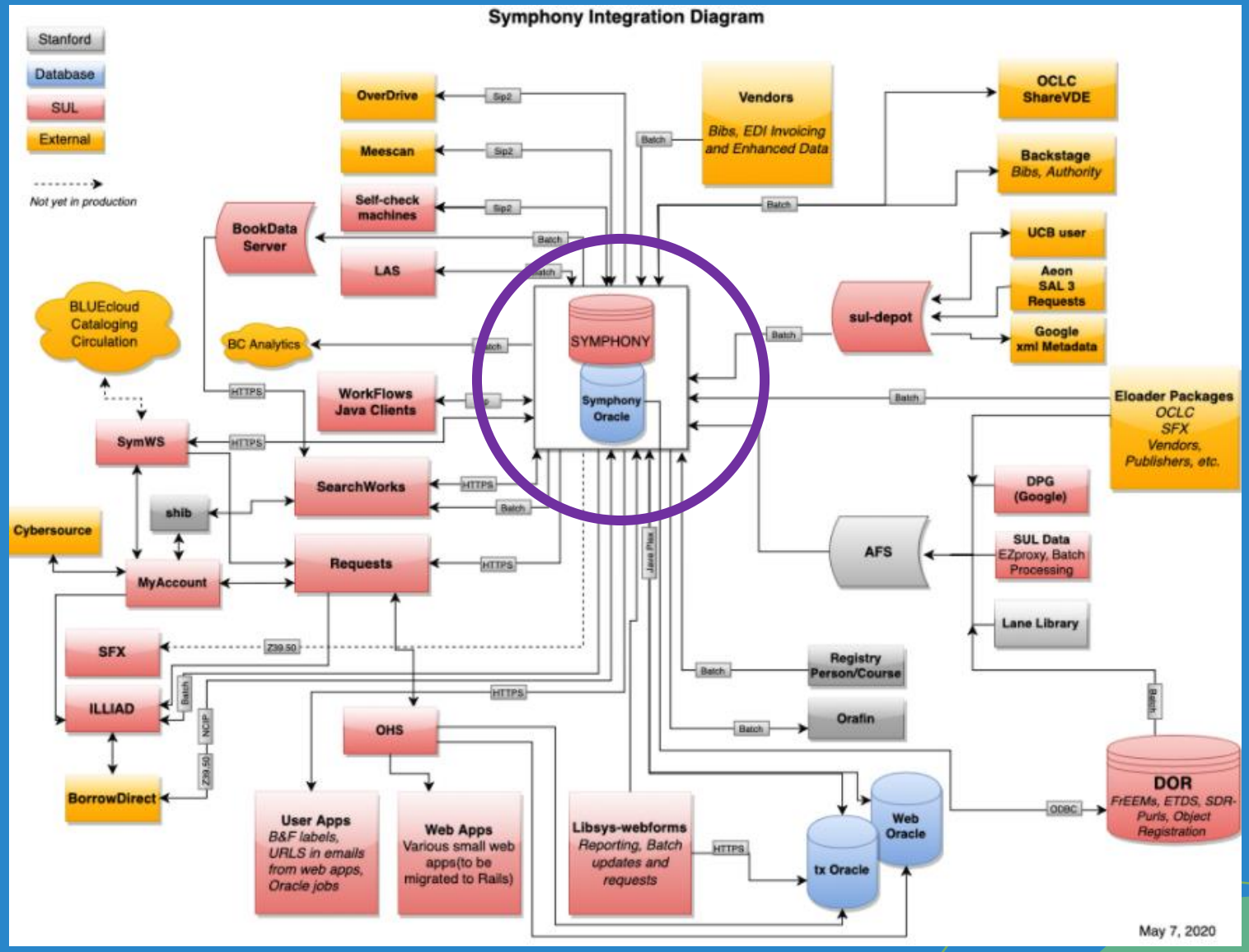
Extending the model

- BF extensions for specific domains
  - Art & Rare Materials Ontology
  - Performed Music Ontology
- Working in other data stores & referring to through BF
  - Program for Cooperative Cataloging Wikidata & ISNI pilots

Tools for a Linked Data Infrastructure

- An Ontology (BIBFRAME, RDA)
- An ILS
- MARC to BIBFRAME converter
- BIBFRAME to MARC converter
- RDF Editor
- Application Profile(s)
- Reconciliation & Entity Management Service

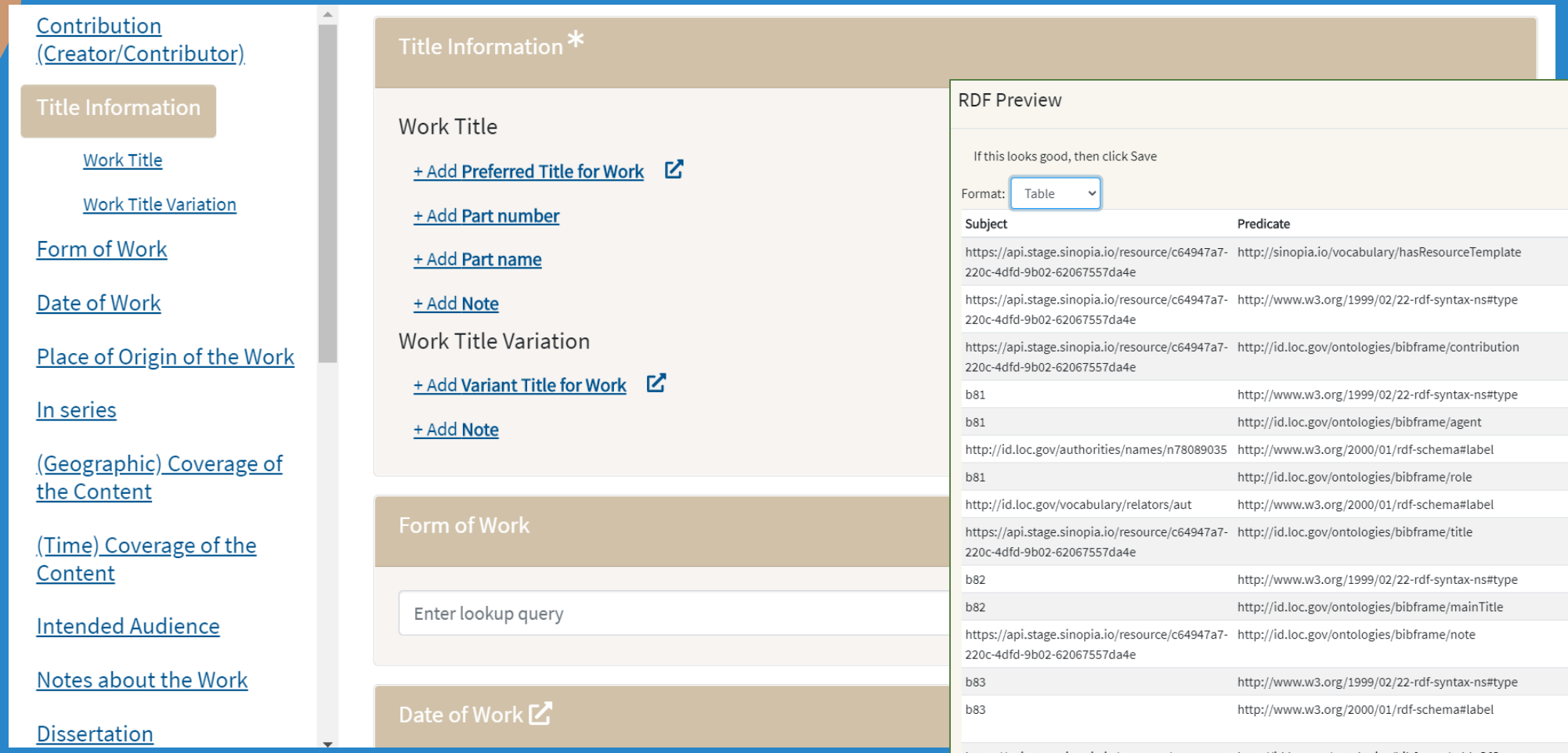
The ILS +


Symphony Integration Diagram
Stanford
Database
SUL
External
Not yet in production
OverDrive
Meescan
Self-check machines
BookData Server
LAS
Vendors
Bibs, EDI Invoicing and Enhanced Data
OCLC ShareVDE
Backstage
Bibs, Authority
UCB user
Aeon SAL 3 Requests
Google xml Metadata
sul-depot
BLUEcloud Cataloging Circulation
BC Analytics
SYMPHONY
Symphony Oracle
WorkFlows Java Clients
SymWS
SearchWorks
shib
Cybersource
MyAccount
Requests
Eloader Packages
OCLC
SFX
Vendors, Publishers, etc.
DPG (Google)
SUL Data
EZproxy, Batch Processing
AFS
Lane Library
SFX
ILLIAD
BorrowDirect
OHS
Registry Person/Course
Orafin
DOR
FrEEMs, ETDS, SDR-Purls, Object Registration
User Apps
B&F labels, URLS in emails from web apps, Oracle jobs
Web Apps
Various small web apps(to be migrated to Rails)
Libsys-webforms
Reporting, Batch updates and requests
tx Oracle
Web Oracle
Sip2
Batch
HTTPS
NCIP
Z39.50
ODBC
May 7, 2020

Converters

- MARC to BIBFRAME
  - Legacy data in our systems is in the MARC format
  - We will continue to receive MARC records for a long time to come
  - Conversion of single records and in bulk
- BIBFRAME to MARC or local format
  - Some ILS functions may require MARC data to operate for some time to come
    - Linking order & financial data
    - Local holding and circulation information
    - Connections to other automatic inputs

RDF editor & template maker



RDF Preview

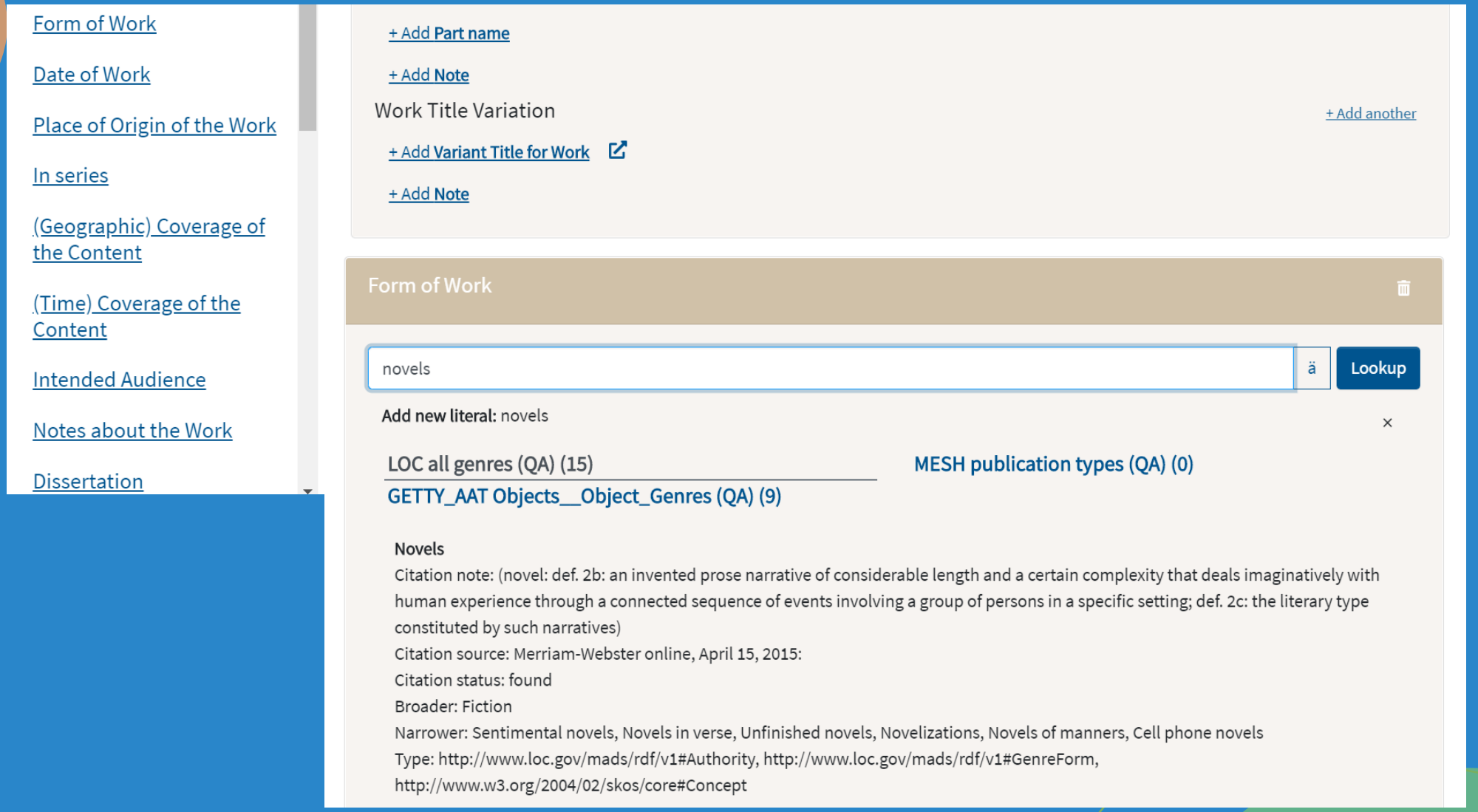
If this looks good, then click Save

Save

Format: Table

Subject	Predicate	Object
https://api.stage.sinopia.io/resource/c64947a7-220c-4dfd-9b02-62067557da4e	http://sinopia.io/vocabulary/hasResourceTemplate	pcc:bf2:Monograph:Work:Un-nested
https://api.stage.sinopia.io/resource/c64947a7-220c-4dfd-9b02-62067557da4e	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://id.loc.gov/ontologies/bibframe/Work
https://api.stage.sinopia.io/resource/c64947a7-220c-4dfd-9b02-62067557da4e	http://id.loc.gov/ontologies/bibframe/contribution	b81
b81	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://id.loc.gov/ontologies/bflc/PrimaryContribution
b81	http://id.loc.gov/ontologies/bibframe/agent	http://id.loc.gov/authorities/names/n78089035
http://id.loc.gov/authorities/names/n78089035	http://www.w3.org/2000/01/rdf-schema#label	Library of Congress
b81	http://id.loc.gov/ontologies/bibframe/role	http://id.loc.gov/vocabulary/relators/aut
http://id.loc.gov/vocabulary/relators/aut	http://www.w3.org/2000/01/rdf-schema#label	Author
https://api.stage.sinopia.io/resource/c64947a7-220c-4dfd-9b02-62067557da4e	http://id.loc.gov/ontologies/bibframe/title	b82
b82	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://id.loc.gov/ontologies/bibframe/Title
b82	http://id.loc.gov/ontologies/bibframe/mainTitle	Card catalog [eng]
https://api.stage.sinopia.io/resource/c64947a7-220c-4dfd-9b02-62067557da4e	http://id.loc.gov/ontologies/bibframe/note	b83
b83	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://id.loc.gov/ontologies/bibframe/Note
b83	http://www.w3.org/2000/01/rdf-schema#label	Includes bibliographical references (page 221) and index. [eng]
https://api.stage.sinopia.io/resource/c64947a7-220c-4dfd-9b02-62067557da4e	http://id.loc.gov/ontologies/bibframe/tableOfContents	b84
b84	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://id.loc.gov/ontologies/bibframe/TableOfContents
b84	http://www.w3.org/2000/01/rdf-schema#label	Origins of the card catalog -- The enlightened catalog --

Lookups to Vocabularies



Application Profiles

Top Class	Property	Property Label in Sinopia	Mandatory?	Repeatable?	Ordered?	Remark	Remark URL	Lookups added	RT linked
bf:Work									
	bf:contribution	Contribution (Creator/Contributor)	FALSE	TRUE					pcc:bflc:PrimaryContribution
									pcc:bf2:Contribution
	bf:title	Title Information	TRUE	TRUE					pcc:bf2:Title:WorkTitle
									pcc:bf2:Title:WorkVariantTitle
	bf:genreForm	Form of Work	FALSE	TRUE				urn:ld4p:qa:genres	
								urn:ld4p:qa:mesh:publication_type	
								urn:ld4p:qa:gettyaat:Objects__Object_Genres	
	bf:originDate	Date of Work	FALSE	TRUE			http://access.rdatoolkit.org/6.4.html		
	bf:originPlace	Place of Origin of the Work	FALSE	TRUE			http://access.rdatoolkit.org/6.5.html	urn:ld4p:qa:names:geographic	
								urn:ld4p:qa:geonames:area_and_place	
	bf:hasSeries	In series	FALSE	TRUE		URI for series as a work; lookup in Sinopia or use URI from another source		urn:ld4p:sinopia:bibframe:work	
	bf:geographicCoverage	(Geographic) Coverage of the Content	FALSE	TRUE			http://access.rdatoolkit.org/7.3.html	urn:ld4p:qa:names:geographic	
								urn:ld4p:qa:geonames:area_and_place	
	bf:temporalCoverage	(Time) Coverage of the Content	FALSE	TRUE			http://access.rdatoolkit.org/7.3.html		
	bf:intendedAudience	Intended Audience	FALSE	TRUE			http://access.rdatoolkit.org/rdachp7_rda7-634.h	https://id.loc.gov/vocabulary/maudience	
	bf:note	Notes about the Work	FALSE	TRUE					pcc:bf2:Note:GeneralNote
	bf:dissertation	Dissertation	FALSE	TRUE			http://access.rdatoolkit.org/7.9.html		pcc:bf2:Dissertation
	bf:tableOfContents	Contents	FALSE	TRUE			http://access.rdatoolkit.org/lcpschp25_lcps25-3		pcc:bf2:Note:Contents
	bf:summary	Summary	FALSE	TRUE			http://access.rdatoolkit.org/rdachp7_rda7-776.h		pcc:bf2:Note:Summary
	bf:subject	Subject of the Work	FALSE	TRUE			http://access.rdatoolkit.org/rdachp23_rda23-12	urn:ld4p:qa:subjects	
								urn:ld4p:qa:names	
								urn:ld4p:qa:mesh:subject	
								urn:ld4p:qa:oclc_fast:topic	
								urn:ld4p:qa:oclc_fast:geographic	
								urn:ld4p:qa:fast:event	
	bf:classification	Classification numbers	FALSE	TRUE					pcc:bf2:Classification:LCC
									pcc:bf2:Classification:DDC
									NLM:RT:bf2:NLMClassification
	bf:content	Content Type	TRUE	TRUE			http://access.rdatoolkit.org/6.9.html	https://id.loc.gov/vocabulary/contentTypes	
	bf:language	Language	FALSE	TRUE				https://id.loc.gov/vocabulary/languages	
	bf:notation	Script	FALSE	TRUE			http://access.rdatoolkit.org/7.13.2.html		pcc:bf2:Script
	bf:illustrativeContent	Illustrative Content	FALSE	TRUE			http://access.rdatoolkit.org/7.15.html	https://id.loc.gov/vocabulary/millus	
	bf:colorContent	Color Content	FALSE	TRUE			http://access.rdatoolkit.org/7.17.html	https://id.loc.gov/vocabulary/mcolor	
	bf:supplementaryContent	Supplementary Content	FALSE	TRUE					pcc:bf2:Note:SupplContent
	bflc:relationship	Related Work	FALSE	TRUE			http://access.rdatoolkit.org/rdachp25_rda25-65		pcc:bf2:RelatedWork
	bf:hasExpression	Related Expression	FALSE	TRUE			http://access.rdatoolkit.org/rdachp26_rda26-25		pcc:bf2:RelatedExpression
	bf:expressionOf	Expression of	FALSE	TRUE		Use to add URI for SHARE-VDE Opus, BIBFRAME Hub, or NAF name-title (with no expression info)			
	bf:hasInstance	Has BIBFRAME Instance	FALSE	TRUE					
	bf:adminMetadata	Administrative metadata	FALSE	TRUE		Use with a local AdminMetadata RT			

Reconciliation Services

- Take local URIs and match them to more broadly recognized URIs, e.g., VIAF, ISNI, Wikidata
- Can enhance your data with further data as specified
- Resembles a pumped-up authority service

Challenges

- Changing is HARD!
  - Deeply embedded infrastructures to change
  - “Disruptive innovation”
- A mixed environment creates its own challenges
  - Hard to move beyond MARC influences when you still need it
  - Discovery is more difficult
- Interoperability
  - BF is a very flexible ontology; some need for core modeling agreements
  - There will always be different flavours as we extend our data beyond the core
- ILS & Discovery
  - Many of our discovery systems are very MARC-dependent & will take time to change
  - Cataloging tools & data ingest need to allow for variations beyond a core BF to be brought into our discovery layers

Opportunities

- Enrichment of our metadata from other sources of our choice
- Sharing our own metadata with other libraries & the world
- Flexible modeling that can adapt to differing worldviews while retaining interoperability at a certain level
- Meaningful library connections with non-library LD respositories
- A dynamic presence in metadata on the Web

INSTEAD OF RISKING ANYTHING NEW, LET'S PLAY IT SAFE BY CONTINUING OUR SLOW DECLINE INTO OBSOLESCENCE.
TOM FISH BURNE
©marketoonist.com

The world enriched with library data;

libraries enriched with the world's data.

LD4 Community Vision Statement

Thank you!

nlorimer@stanford.edu