Creating Shareable Metadata

Pre-Conference at WebWise 2006: Inspiring Discovery: Unlocking Collections Los Angeles, CA February 15, 2006

Jenn Riley, Indiana University

Sarah Shreeves, University of Illinois at Urbana-Champaign

Liz Milewicz, Emory University

What does this record describe?

- identifier: http://name.university.edu/IC-FISH3IC-X0802]1004_112
- publisher: Museum of Zoology, Fish Field Notes
- format: jpeg
- rights: These pages may be freely searched and displayed. Permission must be received for subsequent distribution in print or electronically.
- type: image
- subject: 1926-05-18; 1926; 0812; 18; Trib. to Sixteen Cr. Trib. Pine River, Manistee R.; JAM26-460; 05; 1926/05/18; R10W; S26; S27; T21N

language: UND

source: Michigan 1926 Metzelaar, 1926--1926;description: Flora and Fauna of the Great Lakes Region

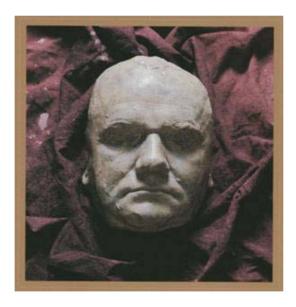
Dublin Core record retrieved via the OAI Protocol

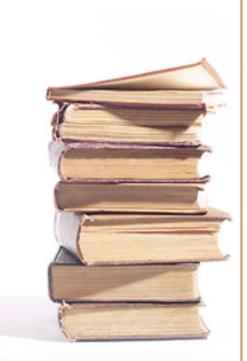
Orig. No. 460 Sta. No. Locality michigan: Trib to sixteen on Dec. 26-T212 R10W , Wefford B Trib. Pine River mainster rather floods vertas be Water: mostly spring - fed; clean spring Vegetation: mot much ab solutely nothing Bottom: sand; no med, nather deep. Temp. 470 ais 62° Shore; mostly cleared -> Jarma & same brush Current: The ft Distance from shore: an undth 4 ft Tide: Method of capture: Collected by Mekelaan Date: V:18:1926

Uninal life subnorme

And this one?

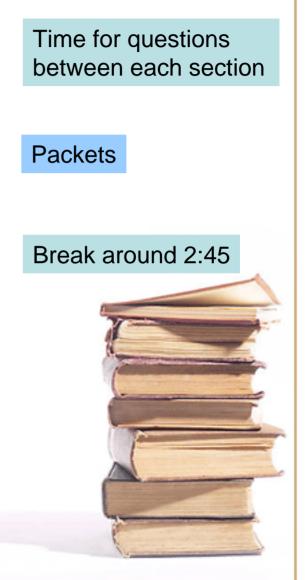
• <u>Sample Record</u> (adapted from an Illinois database)





Agenda

- 1. Introduction
- 2. Why share metadata anyway?
- 3. What is shareable metadata?
- 4. How do you create shareable metadata?
- 5. Communication and documentation
- 6. Before you share...



We hope that you will leave today and...

- Understand the need for interoperable or shareable metadata
- Understand the **impact** your metadata has on larger aggregations
- Understand the key components of shareable metadata
- Think critically about the shareability of their own metadata.

What we aren't covering...

- How to share metadata (whether via the OAI Protocol, SRW/U, FTP, CD-ROM, Cooperative Cataloging, Excel Spreadsheet!)
- Details of technical aspects of shareable metadata such as XML schema & namespaces
- How to use specific metadata formats

One Size Does Not Fit All!

Questions?

Why share metadata anyway?

- Benefits to users
 - Single search of a variety of digital resources
 - Aggregation of subject-specific resources
 - Higher quality resources
- Benefits to institutions
 - Increased user access to collection by allowing metadata to appear in other places
 - Exposure to broader audience, new users
 - Surfacing rare, unknown, or scattered collections



Different shapes and sizes...

- Range of different aggregations:
 - focused subject area v. comprehensive
 - specialized audience v. general audience
- Range of different displays



Metadata aggregators

- CIC Metadata Portal
 <u>http://cicharvest.grainger.uiuc.edu/</u>
 - Records and digital resources shared by consortium of institutions, *provided for*
 - Educators, researchers, and general public

Benefits:

Next

Single comprehensive search of multiple collections and a variety of disciplines

CIC Metadata Po	rtal	Click here to install
Search CIC Library Resour	rces	the Firefox plug-in Add Firefox search plugin
Simple search Bibliographic search Browse collectio	ns Project Details	Help
Advanced Search	CIC Search Portal	Browse by Institution
Search within particular fields (help) Use a word or phrase, e.g., diploma*, fancy dress Title: ar Author/ ar Creator: ar	 The CIC metadata portal provides access to a broad range of noteworthy and often singular information resources held by Midwestern research universities belonging to the CIC consortium (Big Ten Athletic Conference Universities, the University of Illinois at Chicago, and the University of Chicago). The portal allows you to search across many types of content: text, videos, sheet music, thesis and dissertations, Websites, software. Where available, links are provided directly to the information resource on the hosting CIC institution Website. 	University of Chicago University of Illinois at Chicago University of Illinois at Urbana-Champaign Indiana University University of Iowa University of Iowa University of Michigan Michigan State University Pennsylvania State University University of Wisconsin-Madison University of Minnesota The Ohio State University
Jary 15, 2006 Creatin	g Shareable Metadata:	

WebWise 2006 Preconference

Search CIC Lib	rary Resources	Click here to install the Firefox plug-in Add Firefox
Simple search Bibliographic search	Browse collections Project Details	search plugin Help
kw: "reading rooms" 196 records.		Refine search
Results by Collection	Sort by weighted hit frequency Group results by collection Jum	Next 10 Record np to Records: 1 11 21 31 41 51 61 71 81 91 101 11: 121 131 141 151 161 171 181 19:
Michigan State University: The Making of Modern Michigan 3 records	Record 1 of 196	
Ohio State University: The OhioLINK Digital Media Collection	Title American Library Association Archiv Service to Children (ALSC): Publica Yearbooks	
1 record	Subject Booklists; Children, Committee on L Children's Services; Foreign Born, L	
University of Chicago: Archival Photofiles 186 records	Type Text Collection American Library association finding	g-aid
University of Minnesota: War porters 1 record	Record 2 of 196	
University of Wisconsin-Madison:	Title American Library Association Archiv (IRO): American Library School in I Conference Subject File	
Portal to Asian Internet Resources 1 record	Subject American Legion; Paris, American L Type Archival Material	ibrary in
Portal to Asian Internet Resources	(IRO): American Library School in I Conference Subject File Subject American Legion; Paris, American L	Paris: American Legion .ibrary in q-aid IQ:

C	C Metada	ata Po	rtal		
	Search CIC Libra	ary Resour	rces		Click here to install the Firefox plug-in Add Firefox search plugin
Simp	le search Bibliographic search	Browse collection	is Project Details		Help
	earch was in the Subject field for "re ecords.	e ading rooms" and in	n the Resource Type field for "image" .	Refine search 🗖	Search
Result	s by Collection	Sort by weighted	hit frequency 💌 Jump to Record:	s: 1 11 21 31 41 51 61 121 131 1	Next 10 Records 71 81 91 101 111 41 151 161 171 181
Archiv	- rsity of Chicago: al Photofiles ecords	Author/Creator Subject Type	Joseph Regenstein Library Series II: Buildings and Grounds Reading Rooms 1 3 Skidmore, Owings and Merrill; Netsch, Walter, b University of Chicago. Library; Reading rooms Image Archival Photofiles		
		Record 2 of 184 Title	Joseph Regenstein Library		
			Series II: Buildings and Grounds		
Februci, i	0, 2000	Author/Creator	Reading Rooms 3 Skidmore, Owings and Merrill; Netsch, Walter, b Somanous moregana.	o. 1920; Feldman, Steven	
-			ise 2006 Preconference		

Metadata aggregators

- National Science Digital Library
 <u>http://nsdl.org</u>
 - Online resources and records pertaining to science & math education and research, vetted for inclusion, *provided for*
 - Educators, researchers, policy makers, and the general public

<u>Benefits:</u>

Single portal serving a range of resources on a specialized topic to a diverse audience



NSDL





NSDL Home

Home

Search

Browse

Resources For

- K12 Teachers
- Librarians
- NSDL Community
- University Faculty
- First Time Users

News

- New in NSDL
- Newsfeeds

Press

Publications

- Whiteboard Report
- Annual Report
- Newsletter Sign Up

AskNSDL About NSDL

SEARCH THE NATIONAL SCIENCE DIGITAL LIBRARY

evo devo

Search

NSDL is the Nation's FREE online library for education and research in <u>Science, Technology, Engineering</u>, and <u>Mathematics</u>.

News

Discovering A Lost World

February 8, 2006 Did you ever wonder if there were any places on earth left to be discovered? An international team of scientists from the Indonesian Institute of Science (LIPI), Conservation International, The South Australian Museum in Adelaide, and the U.S. had that question in mind as their helicopter touched down in a remote lake bed in the unexplored Foja Mountains of western New Guinea. Their month-long adventure began right away as they began spying new and rare species of plants and animals that included butterflies, enormous flowers, a new species of honeyeater bird, and a tree kangaroo (Matschie's tree kangaroo, New Guinea, is shown at right).



GREATE

Larger Type

Read More Read Previous Highlights

WebWise 2006 Preconference

Febru

THE NATIONAL SCIENCE		
NSDL Home > Search		Larger Typ
Home	NSDL Search Results	
Search Browse		
Resources For	Search Tips	
K12 Teachers	SEARCH	
Librarians	evo devo	
NSDL Community	(Displaying results:1 - 10 of 18)	Next Results →
 University Faculty 		<u>New Results</u>
 First Time Users 	Title/Description	Resource
News		Format
New in NSDL		
Newsfeeds	Bringing Evo Devo to Life Sean Carroll's book shows how the central black box of development t	hot
Press	maps phenotypes onto genoty	ilat.
Publications	View all related information - Include/Exclude results like this	
Whiteboard Report		
 Annual Report 	Trends in genomic 'evo-devo'	
Newsletter Sign Up	A report on the joint Spring meeting of the British Society of Developm	ental
AskNSDL	Biology and the Geneti	
About NSDL	View all related information - Include/Exclude results like this	

WebWise 2006 Preconference

SDL						
THE NATIONAL SCIENCE	D IGITAL LIBRAR	Y Explo	RE	LE A RN Larger Type		
Home Search	-Back to se	earch results		Laigei Type		
Browse	Title:					
Resources For	Bringing Evo	Devo to Life				
 K12 Teachers 	Description:					
 Librarians 						
 NSDL Community 	Sean Carroll's book shows how the central black box of development that maps phenotypes onto genotypes is revealing its secrets in many exquisite, and frequently unexpected, ways.					
University Faculty	Archives:					
 First Time Users 	Archives: Archived Version					
News	Item Level N		Collection I	evel Metadata:		
 New in NSDL 						
 Newsfeeds 	More In	formation	Brand Image Not Provided			
Press	NSDL OAI oai:plosbiology.org:		More Information			
Publications	identifier	10.1371/journal.pbio.0030340	More II	inormation		
 Whiteboard Report 	Title	Bringing Evo Devo to Life	NSDL OAI identifier	oai:nsdl.org:nsdl.nsdl:710541		
 Annual Report 	Author	Paul M. Brakefield	Stratitional and	DUED MERICE		
 Newsletter Sign Up 	Subject			PLoS Medicine		
AskNSDL	Keyword(s)	Development		MedicineStudy and teaching, Health, Biological and life sciences		
About NSDL	Description	Sean Carroll's book shows how the central black box of development that maps phenotypes onto genotypes is revealing its secrets in many exquisite, and frequently unexpected, ways.		PLoS Medicine is an open-access, peer-reviewed medical journal published monthly, online and in print, by the Public Library of Science (PLoS), a nonprofit organization. PLoS Medicine is an international, multidisciplinary medical journal that aims to publish		
	Publisher / Resource Provider	Public Library of Science		outstanding human studies that substantially enhance the understanding of human health and discass. We sim to promote translation		

WebWise 2006 Preconference

Febru

Metadata aggregators

- Collaborative Digitization Program
 <u>http://www.cdpheritage.org/</u>
 - Records and digital resources from archives, museums, and libraries, pertaining to cultural heritage of the American West provided for
 - Educators, researchers, and the general public

Benefits:

Single portal serving a range of resources on a specialized topic to a diverse audience

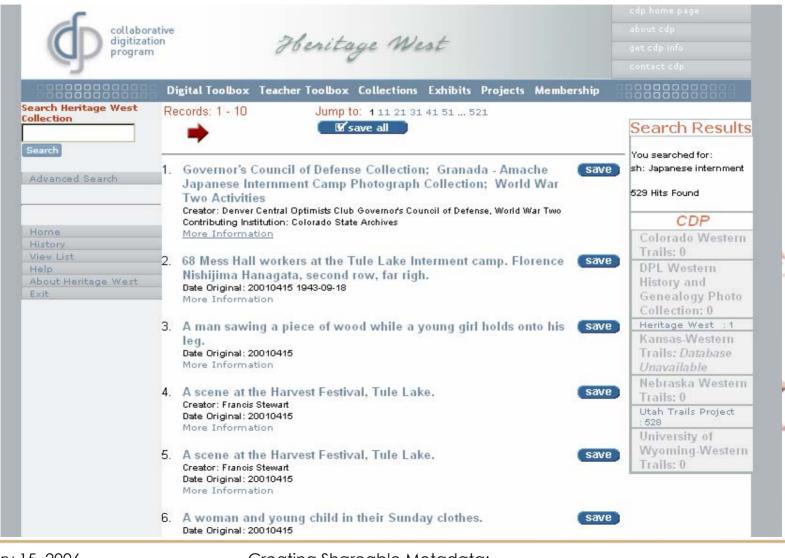


Collaborative Digitization Program



WebWise 2006 Preconference

Collaborative Digitization Program



February 15, 2006

Collaborative Digitization Program



February 15, 2006

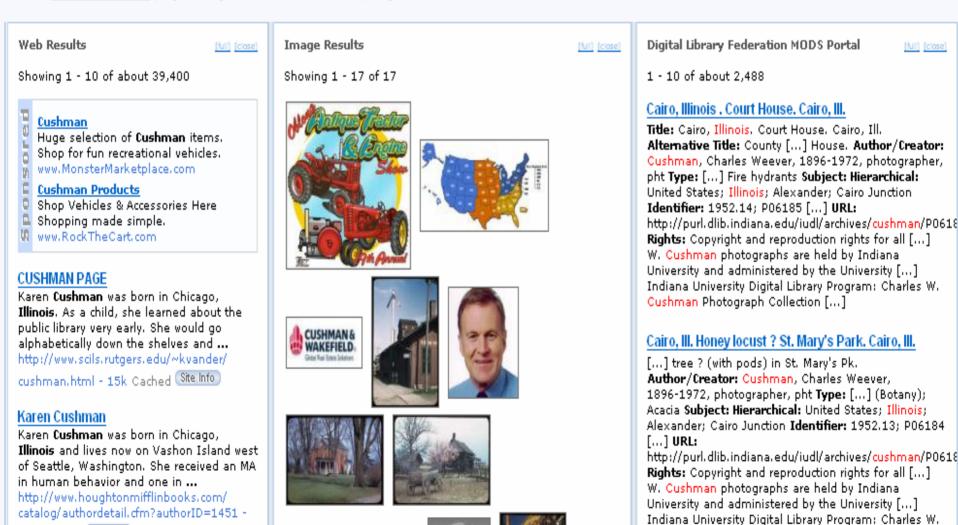
Not only stand alone portals...

http://www.a9.com/



Search:	illinois cushman				GO	Home	Prefs Toolbar Sign Out
	Advanced Web Search					Hello Sarah L	. Shreeves, <u>Clickhere</u> if this is not ye
Choices							
	🗹 Web	🔲 Books	🗹 Images	Yellow Pages	📃 Reference	Movies	
	People	📃 Blog Search	🔲 Wikipedia	Your History	🔲 Your Diary	🔲 Your Bookmarks	
	🗹 DLF MODS Po	rtal 🔲 More Choices	🔽				
		🗹 Web 📃 People	Advanced Web Search Choices Web Books People Blog Search	Advanced Web Search Choices Web Books Images People Blog Search Wikipedia	Advanced Web Search Choices Web Books Images Yellow Pages People Blog Search Wikipedia Your History	Advanced Web Search Choices Web Books Images Yellow Pages Reference People Blog Search Wikipedia Your History Your Diary	Advanced Web Search Hello Sarah L Choices Web Books Images Yellow Pages Reference Movies People Blog Search Wikipedia Your History Your Diary Your Bookmarks

Join the A9 Instant Reward program to get 1.57% off almost everything on Amazon.com



Common Problems with Metadata in Aggregation

- Consistency
- Sufficiency
- Compatibility

Consistency problems

- Appearance of data
- Application of format
- Granularity of records
- Vocabulary usage

<u>Result</u>:

Service Provider must normalize data (if can determine what "normal" is)

Sufficiency problems

 Too little info for understanding what resource is, especially outside of local context

<u>Result</u>: Users don't know whether a resource is relevant or not

Compatibility problems

Information in records is

- Erroneous
- Unnecessary
- Incompatible

<u>Result:</u>

Interferes with harvesting and indexing

Questions?

February 15, 2006

What is Shareable Metadata?



Shareable Metadata...

- Promotes search interoperability
 - "the ability to perform a search over diverse sets of metadata records and obtain meaningful results." (Priscilla Caplan)
- Is (at the least) human understandable outside of its local context (the field note!)
- Is useful outside of its local context (Can we build something off of it?)
- Preferably is machine processable!

The Four C's (and lots of S's) of Shareable Metadata

Consistency Coherence Context Conformance

Metadata standards (and not just DC) Vocabulary and encoding standards Descriptive content standards (AACR2, CCO, DACS) Technical standards (XML, Character encoding, etc)

Questions?

February 15, 2006

How do you create shareable metadata?

Appropriate representation of the resource through metadata

- Metadata as a *view* of the resource
- Standards promote interoperability
- Appropriate formats
- Appropriate content
- Appropriate context



Metadata as a *view* of the resource

- Metadata for the same thing is different depending on *use* and *audience*
- Affected by format, content, and context
- Harry Potter as represented by..
 - a public library
 - an <u>online bookstore</u>
 - a <u>fan site</u>



Same Object, Different Descriptions, Different Metadata

Illinois Digital Archives (alt)

University of Minnesota Libraries (alt)

Northwestern Poster Collection
(alt)

Teaching with Digital Content (alt)

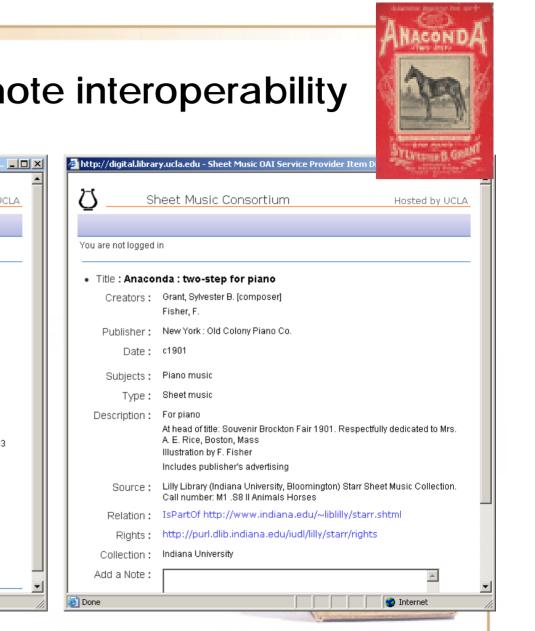
World War II Poster

February 15, 2006

Choice of vocabularies as a view

- Names
 - LCNAF: Michelangelo Buonarroti, 1475-1564
 - ULAN: Buonarroti, Michelangelo
- Places
 - LCSH: Jakarta (Indonesia)
 - TGN: Jakarta
- Subjects
 - LCSH: Neo-impressionism (Art)
 - AAT: Pointillism





Standards promote interoperability

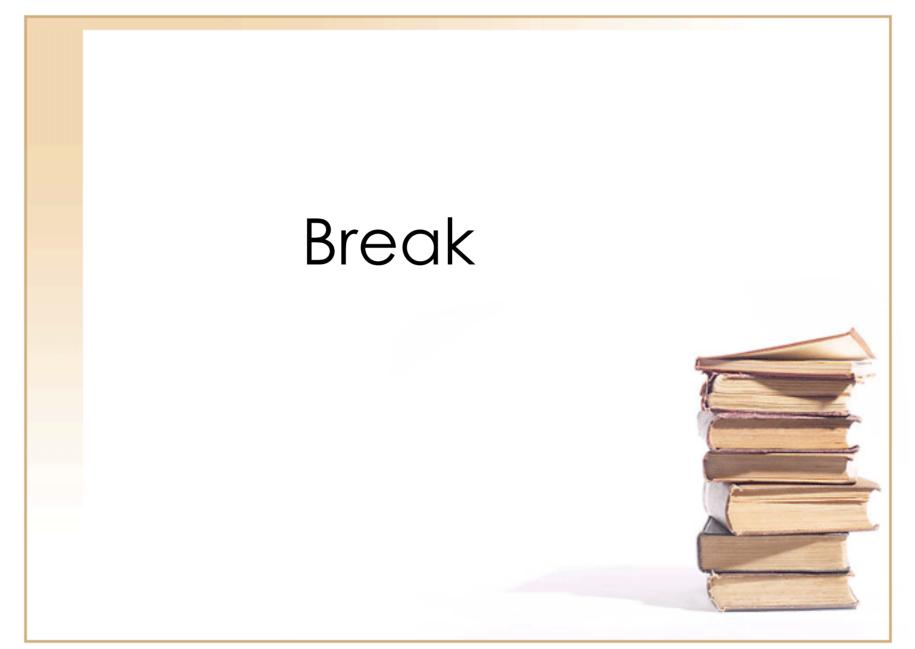
are not logged	in	
rare notroggeo		
Title : Anaco	nda Two-Step. Souvenir Brockton Fair 1901.	
Creators :	By Sylvester B. Grant. Fisher	
Publisher :	Brockton, Mass.: Compliments of Old Colony Piano Co., 4 Ma	iin St.
Date :	1901	
Subjects :	Horses Horse racing	
Description :	Respectfully Dedicated to mrs. A.E. Rice, Boston, Mass. ads on back cover for Old Colony Piano Co.	
Source :	Johns Hopkins University, Levy Sheet Music Collection, Box 0	128, Item 003
Collection :	Johns Hopkins University Levy Collection	
Add a Note :		
		T
	add - add/edit a note for this record	
	[access online] [add - add to virtual collection]	

Types of standards to consider

- Data structure standards
 - "Buckets" of information (fields)
 - Both label and scope important
 - e.g., MARC, MODS, Dublin Core
- Data content standards
 - Selection, structure and formatting of value within a field
 - e.g., AACR2, DACS, CCO

Appropriate formats

- Depends upon:
 - nature of materials
 - nature of holding institution
 - depth of description needed
 - community practice
 - need for distinction between versions
 - need for repeating elements
- MARC, MODS, Dublin Core, EAD, and TEI may all be appropriate for a single item



Appropriate content

- Choose appropriate vocabularies
- Choose appropriate granularity
- Make it obvious what to display
- Exclude unnecessary "filler"
- Make it clear what links point to
- Handout has some guidance for "classes of metadata elements"

Common content mistakes

- No indication of vocabulary used <u>Example</u>
- Shared record for a single page in a book
- Link goes to search interface rather than item being described
- "Unknown" or "N/A" in metadata record



Appropriate context

- Thinking about shareability
 - Include information not used locally
 - Exclude information *only* used locally
- Current safe assumptions
 - Users discover material through shared record
 - User then delivered to your environment for full context
- Context driven by intended use

Common context mistakes

- Leaving out information that applies to an entire collection ("On a horse")
- Location information lacking parent institution
- Geographic information lacking higher-level jurisdiction
- Inclusion of administrative metadata



What are you describing?

<u>Physical object w/</u> <u>links to the digital?</u> (Digital surrogate <u>approach)</u>

Both digital and physical in the same flat record?

Content but not the carrier? Both digital and physical in the same record but in a hierarchy?

A record for the analog and the digital item with linkage?

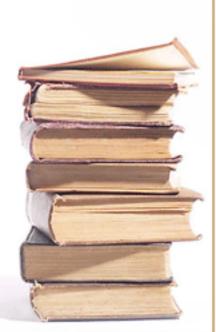
(one to one principle)



Questions?

Communication

- Metadata providers can learn from aggregators
- Aggregators can learn from metadata providers
- Providing supplemental information to make records more intelligible



Metadata providers can learn from aggregators

- Crosswalking methods and rules
- Information to include and exclude
- Choice of standards
 - Metadata formats
 - Vocabularies
- Where to spend normalization effort
- Use can drive future priorities
 <u>Example from the CIC Portal</u>... (alt)

Aggregators can learn from metadata providers

- Where to spend normalization effort
- Context, importance, and primary uses of resources shared
- Variety of resource types and descriptive practices encountered
- Local, robust metadata formats



Providing supplemental information to make records more intelligible

- Method for creating shared records
- Vocabularies and content standards used in shared records
- Record updating practices & schedules
- Accrual practices and schedules
- Existence of analytical or supplementary materials
- Provenance of materials



Before you share...

- Check your metadata
 - Appropriate view?
 - Consistent?
 - Context provided?
 - Does the aggregator have what they need?
 - Documented?

Can a stranger tell you what the record describes?



Questions?