Shared Services for Community Metadata Improvement

Ted Habermann₁, Matthew B. Jones₂, Sean Gordon₁, Ben Leinfelder₂, Bryce Mecum₂, Peter Slaughter₂ and Lindsay A. Powers₃ 1. The HDF Group, 2. National Center for Ecological Analysis and Synthesis 3. United States Geological Society

Metadata Dialects

Many communities use the term "standard" when they describe their metadata and, as a result, there are many existing "standards". This approach focuses attention on differences between communities. We use the term "dialect" to focus attention on common concepts and goals.

Recommendations and Dialects:

Recommendations reflect community experiences and documentation needs. Sharing recommendations is an important mechanism for sharing those experiences and community knowledge. Many communities share documentation needs, so recommendations overlap, particularly for the discovery use case.



Metadata Recommendations and Dialects

Metadata recommendations change as new communities and needs emerge. Metadata management tools are driven by dialects. Changing those tools and training people are difficult, so adoption of new dialects is relatively slow. This leads to gaps between existing organizational capabilities (dialects) and recommendations.



DataCite Recommendation / CSDGM Dialect

RECOMMENDATION ANALYSIS DASHBOARD

The Recommendation Analysis Dashboard is an exploratory metadata evaluation tool. It enables metadata for a single dialect to be easily evaluated using multiple recommendation, such as OGC Catalogue Services for Web (CSW) or Data Citation (DataCite).



A second community creates a dialect (D_2) with recommendations at 2 levels (R_4, R_5) .

Recommendations can also be purely conceptual, i.e. independent of dialects (e.g. Catalog Services for the Web)

Catalog Services for the Web

Additional Queryables

Dashboard: Provides desktop-based capability to test complete collections against user-selected recommendations.



Collection	Concept	Occurance	Concept	Guidance	Links	
Concept	CSDGM	Recommendation Profile	Incomplete use in the Collection	Unused in Collection	Missing from Dialect	
(evword	86%	CSWCoreQuervables	Keyword	Recourse Format	Motodata Identifier	
Resource Title	100%	CSWCoreQueryables	Bounding Box	Resource Type	Association	
Abstract	100%	CSWCoreQueryables	Author / Originator	Coordinate Reference System (CRS)	Resource Revision Date	
Posource Format	10076	CSWCoreQueryables	Publisher	Related Resource Citation	Resource Identifier	
Actodata Identifier	100%	CSWCoreQueryables	Contributor Name	Security Constraints	Parent Identifier	▼
vietadata identifier	-100%	CSWCoreQueryables	Resource Creation/Revision Date			
vietadata Modified Date	100%	CSWCoreQueryables	Publication Date			
Resource Type	0%	CSWCoreQueryables	Organization Name			
Bounding Box	71%	CSWCoreQueryables	Keyword Type			
Coordinate Reference System (CRS)	0%	CSWCoreQueryables				list all of the alon
Association	-100%	CSWCoreQueryables				List all of the elem
Resource Title	100%	CSWCoreReturnables				
Author / Originator	98%	CSWCoreReturnables				l recommendation a
Keyword	86%	CSWCoreReturnables				
Abstract	100%	CSWCoreReturnables				l often they occur in
Publisher	18%	CSWCoreReturnables				
Contributor Name	54%	CSWCoreReturnables				
Metadata Modified Date	100%	CSWCoreReturnables				collection. Identity
Resource Type	0%	CSWCoreReturnables				
Resource Format	0%	CSWCoreReturnables				I missing from dialect
Metadata Identifier	-100%	CSWCoreBeturnables				
Source Citation	2%	CSWCoreBeturnables				collection complet
Metadata Language	100%	CSWCoreBeturnables				
Related Resource Citation	0%	CSWCoreReturnables				
Bounding Box	71%	CSWCoreReturnables				
Dights	100%	CSWCorePeturnables				
Resource Revision Date	100%	CSWAdditionalQuanablas				
Resource Revision Date	-100%	CSWAdditionalQueryables				
Resource Creation/Revision Date	18%	CSWAdditionalQueryables				
Publication Date	18%	CSWAdditionalQueryables				
Drganization Name	98%	CSWAdditionalQueryables				
Security Constraints	0%	CSWAdditionalQueryables				
Metadata Language	100%	CSWAdditionalQueryables				
Resource Identifier	-100%	CSWAdditionalQueryables				
Parent Identifier	-100%	CSWAdditionalQueryables				
Keyword Type	80%	CSWAdditionalQueryables				
		-100% – Concer	t Not in Dialect			
		-100% = concep	t NOT IT DIALECT			Soo https://www.clidach
						See <u>mups://www.sndesm</u>
		0% = Concept N	ot in Collection			
	8 - 3		of the concetton			
		100% = Concept	in All Records			
	-					

Evaluation Tools

A Library of Metadata Checks:

Commnities can develop several types of checks and implement those checks in java, R, or python. Currently checking metadata, congruency between metadata and data, and data.

A Library of Recommendations:

Recommendations are collections of checks that communities believe are important for some documentation goal.

Web Services:

Provide web-based capability to test single records or complete collections against user-selected recommendations (sets of checks).

Check ID	Check Name	Check	Туре
M1	Descriptive Title	Title exists, > 7 words Metadata	Metadata
M2	Unique Attribute Names	Attribute names unique	Metadata
M3	Valid Units	Units assigned from controlled vocabulary	Metadata
C1	Checksum matches	Data checksums match metadata	Congruency
C2	Data links live	All URLs return data	Congruency
D1	Duplicate data rows	Count duplicate rows	Data

Recommendation	Checks
LTER Best Practice	M1, M2, C2, C3, D3,
Attribute Conention for Data Discovery (ACDD)	M2, M3, M4, C1, C2, D3,
Arctic Data Center	M3, M4, M5, C6, C8, D1, D2, D3,

Data Support About

Metadata Quality Report



54% = Concept in Some Records











News Participate Resources	Education Data	
EARCH: Search Summary	Jump to: DOI or ID Go	Sign in or
knb	Datasets 1 to 5 of 2,666 1 2 3 534 Next Sort by Knb Gregory Goldsmith. 2016. Data from: Plant-O-Matic: A dynamic and mobile guest the Americas. KNB Data Repository. knb.909.8.	y Most recent \$
KNB Data Repository		
Member Node The Knowledge Network for Biocomplexity (KNB) is a national network intended to facilitate ecological and environmental research on biocomplexity.	 Environmental Laboratory, US Army Engineer Research & Development Center, Lemasson. 2016. A sensory-driven tradeoff between coordinated motion in spredator's visual confusion. KNB Data Repository. knb.865.15. 	and Bertrand social prey and a
4 years, 7 monthsDataONE Member Node since 20124,540contributions2,503,786downloads	LTER BP ADCC $10\overline{0}ext$ 100 80 60 40 20 20 20 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 1000 1000 1000 1000 1000 1000 1000 10000	Type Discovery Identification Evaluation



ments in the and show how the metadata fields that are t, missing from ete, or partial.

CONCEPT GUIDANCE

Links to online guidance in the ESIP wiki for incomplete, unused or missing concepts. See http:// wiki.esipfed.org/index.php/ Category:Documentation_Connec tions

nare.net/secret/D6ZbE1t55FMzyK for more information

NSF-DIBBS Award 1443062