

Supporting data production description: an exploration of metadata standards use

Tiffany C. Chao {tchao @ illinois.edu} Center for Informatics Research in Science and Scholarship | Graduate School of Library and Information Science | University of Illinois, Urbana-Champaign

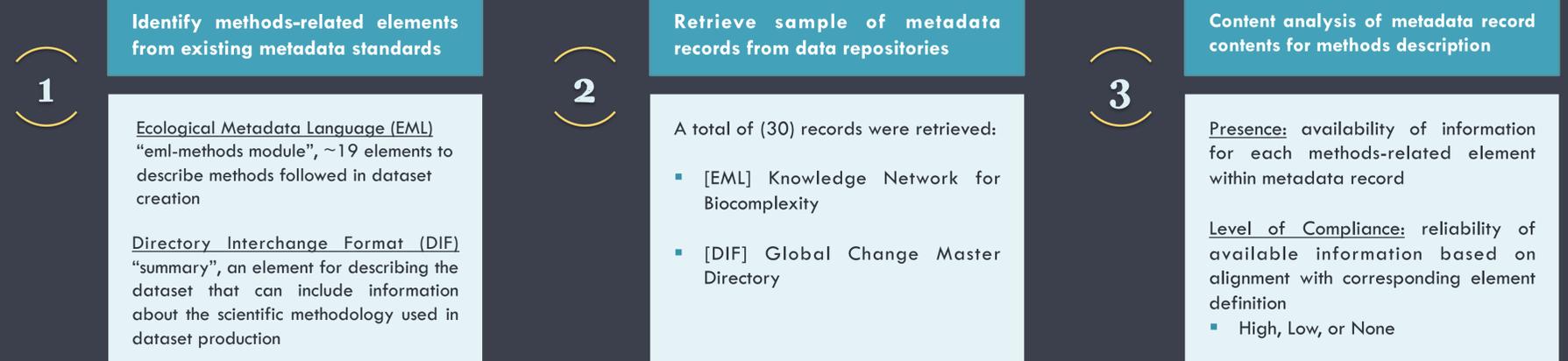
STUDY OVERVIEW

Robust metadata is integral to the reuse of research data. The description of methods used to generate data has a significant role for researchers in reviewing and assessing the quality of data for new inquiry^{1,2} yet little is known about how such information is represented in dataset metadata. This study investigates the availability and content of information related to methods described in metadata records for publically available scientific datasets.

Research Questions

- How are metadata standards used for methods description?
- What methods information is provided in metadata records?

RESEARCH DESIGN



PRELIMINARY FINDINGS

Repository records analysis

Overview of methods-related elements represented in metadata records

Methods-related metadata elements and definitions		Presence	Level of Compliance		
			High	Low	None
EML (15 records)	description* "allows for a textual description of the specific sampling area, the sampling frequency (temporal boundaries, frequency of occurrence), and groups of living organisms sampled (taxonomic coverage)."	12 (80%)	6	5	1
	instrumentation "allows the description of any instruments used in the data collection or quality control and quality assurance. The description should include vendor, model number, optional equipment, etc."	1 (6.7%)	0	1	0
	spatialSamplingUnits "A spatial sampling unit describes the specific geographic areas sampled...can either be described by filling out the structured coverage element or by reference to the values in a data"	2 (13.33%)	1	1	0
	samplingDescription "allows for a text-based/human readable description of the sampling procedures used in the research project. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article."	2 (13.33%)	1	0	1
DIF (15 records)	Summary* A brief description of the data set along with the intended use of the data, including "scientific methodology or analytical tools" and "data processing".	15 (100%)	9	0	6

(*) Elements designated as "required" by metadata scheme

(% Presence) determined by xx/15 records from sample | (Level of Compliance) xx/# of records from Presence

High: fully-aligned with element definition | **Low:** partially-aligned with definition | **None:** not aligned with definition; additional clarification needed to comprehend information

OBSERVATIONS

- Information is more readily available for *required* methods-related elements.
 - EML has many more methods-related elements that were not visible in the metadata records.
- Level of information compliance varied across methods-related elements.
 - Some element definitions more open-ended (i.e. "Summary") while others were more explicit (i.e. "Instrumentation") in what information should be included.
- Available information includes references to project websites and published articles for methods detail.
 - Represented in record content as external links or statements on direct use of article content.

SYNOPSIS

- Mixed availability of methods information suggests metadata standards are not used to their full potential and more comprehensible element definitions and support for use may be needed.
- Recognition of external sources such as journal articles and project websites in the metadata record content presents a potential information source that can be leveraged for metadata generation to enhance standards compliance.

FUTURE WORK

Metadata standards provide an initial basis for description of methods but greater attention is needed in addressing those methods-related elements that were not reported on or had minimal content provided. Additional inquiry on making the data production process more visible to foster data reuse includes:

- examining those referenced resources such as journal articles and project sites for applicable descriptive information to address these metadata elements
- investigating what refinements should be made to methods-related element definitions or documentation to better support metadata generation

