Are Current Metadata Ready for the FORCE11 Software Citation Guidelines?

Ted Habermann – Metadata Game Changers - ted.habermann@gmail.com

Communities develop metadata recommendations for specific use cases and/or data types. Understanding how these recommendations are implemented, identifying high-quality examples, and providing guidance for improvement are important elements of the metadata improvement and adoption process. Software citation is an important use case with an existing recommendation from FORCE11. Metadata collections in several dialects are evaluated for the content recommended by FORCE in order to test readiness for the software citation use case.

① Develop a conceptual recommendation and map it to implementations:

FORCE11 is a community of scholars, librarians, archivists, publishers and research funders that recently developed a recommendation for software citation (Smith AM, Katz DS, Niemeyer KE, FORCE11 Software Citation Working Group. (2016) Software Citation Principles. *PeerJ Computer Science* 2:e86. DOI: 10.7717/peerj-cs.86). The concepts were then mapped to the DataCite Schema. We used crosswalks to map to two other dialects (JATS and ISO 19115-2).

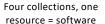| FORCE11 Concept	DataCite (4.1)	FORCE11 Concept	DataCite (4.1)	FORCE11 Concept	DataCite (4.1)
Unique Identifier	Identifier	Description	Description with descriptionType 'TechnicalInfo' Description with descriptionType 'Abstract'	Location/ Repository	Publisher or Contributor/contributorType 'HostingInstitution'
Unique Identifier Type*	'DOI'	Keywords (Keyword Vocabulary*)	Subject (Subject/scheme)	Software License	Rights
Software Name	Title	Version Number	Version	Contributor	Contributor
Author(s)	Creator	Release Date	Publication Year	Contributor Role	contributorType

Important content related to FORCE principles but not included in the original list.

② Search metadata collections:

We know the locations of the concepts recommended by FORCE11 in several dialects. We can search collections that use those dialects and determine how many records in those collections include the recommended content.

DataCite — Four collections, one resource = software

JATS — Two collections from open science publishers

ISO — One collection of software metadata

R E S U L T S

Concept / Dialect	DCITE	DCITE	DCITE	DCITE	DCITE (software)	JATS	JATS	ISO
Collection	CDL.NCAR	CDL.NOAA	CERN.ZENODO	FIGSHARE.ARS	CERN.ZENODO	ELIFE	HINDAWI	Unidata
Date	3/28/18	3/28/18	3/28/18	3/28/18	5/4/18	3/28/18	5/24/18	10/7/17
Unique Identifier	100%	100%	100%	100%	100%	100%	100%	0%
Unique Identifier Type	100%	100%	100%	100%	100%	100%	100%	0%
Software Name	100%	100%	100%	100%	100%	100%	100%	100%
Author(s)	100%	100%	100%	100%	100%	100%	100%	100%
Contributor	96%	1%	7%	0%	2%	78%	99%	100%
Contributor Role	96%	1%	7%	0%	2%	100%	100%	100%
Version	92%	2%	14%	0%	96%	6%	16%	73%
Release Date	100%	100%	100%	100%	100%	78%	99%	100%
Location / Repository	100%	100%	100%	100%	100%	100%	100%	100%
Software License	95%	23%	100%	100%	100%	100%	100%	100%
Description	99%	56%	100%	81%	100%	99%	98%	100%
Keyword	91%	16%	73%	81%	12%	100%	100%	100%
Theme Keyword	91%	16%	73%	81%	12%	100%	100%	100%
Keyword Vocabulary	54%	15%	0%	45%	0%	0%	0%	100%
Average	94%	52%	70%	71%	66%	83%	87%	84%

% of records in 8 metadata collections that include the FORCE 11 Software Citation concepts.
Green cells indicate 100%, yellow cells indicate 0%.

③ Conclusion:

All three dialects studied include the concepts recommended by FORCE for software citations. In 57% of the cases, all records in the collections include content for these concepts (Green). In 6% of the cases, there is no content (Yellow).

Yes, current metadata dialects are ready to use the FORCE11 Software Citation Guidelines.