



Best Practices for Software Metadata: A report from the Software
Preservation Network

International Conference on Dublin Core & Metadata Applications
October 27, 2017

SPN Metadata Working Group

Elizabeth Roke, Emory University

Glynn Edwards, Stanford University

Wendy Hagenmaier, Georgia Institute of Technology

Eric Kaltman, Carnegie Mellon University

Daniel Noonan, Ohio State University

Katherine Thornton, Yale University

Tim Walsh, Canadian Centre for Architecture

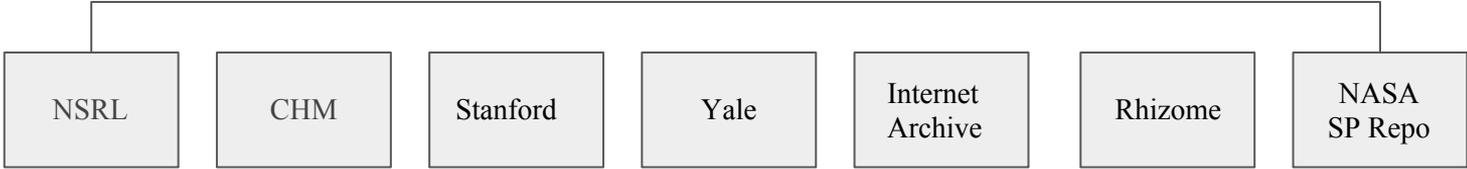
softwarepreservationnetwork.org
@softpresnetwork
<https://osf.io/a7uea/>

Introduction to SPN

- *Saving Software Together*
- **Mission:** Preserving software through community engagement, infrastructure support, and knowledge generation
- **Values:**
 - Community
 - Sustainability
 - Access
 - Transparency
 - Advocacy

SPN in Context (of software preservation)

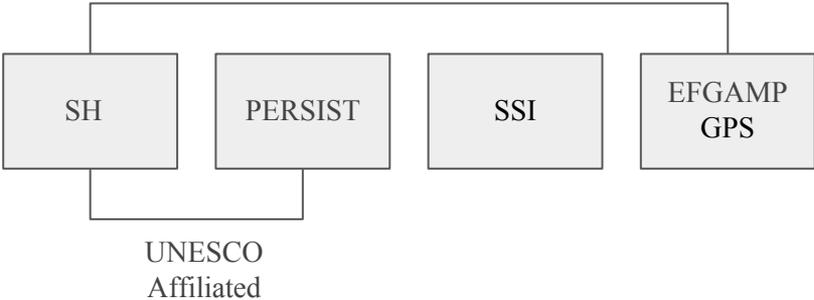
Software Collections (not exhaustive)



Projects (U.S.)



Programs/Efforts (international)



Commercial Services



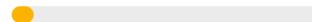
SPN Metadata Survey

The Software Preservation Network (SPN) is an initiative to explore and establish partnerships, collaborations, and best practices for software preservation. The SPN Metadata Working Group is currently conducting a survey of institutions with digital preservation programs to gain insight into metadata practices for software and other digital objects. The results of this survey will be used to establish a baseline for metadata best practices for software.

Please consider completing this very short survey about metadata practices in use at your institution. We are looking for responses both from institutions collecting software and those who have yet to begin this kind of work. This survey will be open through April 28, 2017



NEXT



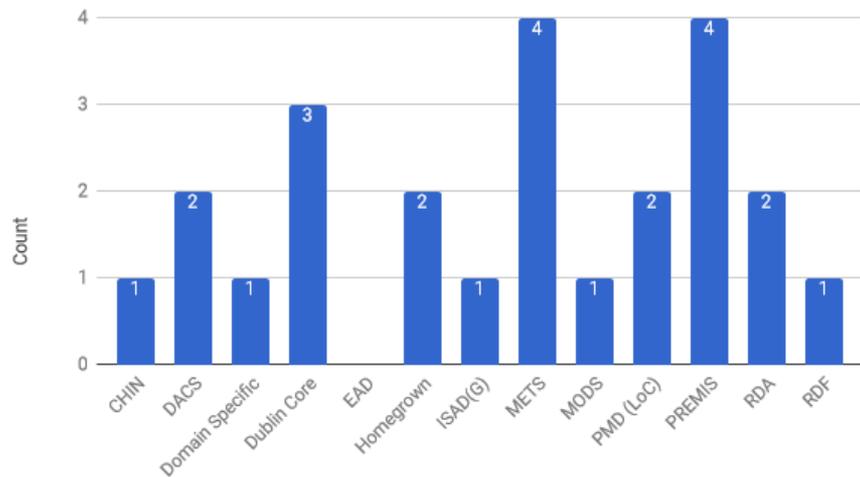
Page 1 of 13

Never submit passwords through Google Forms.

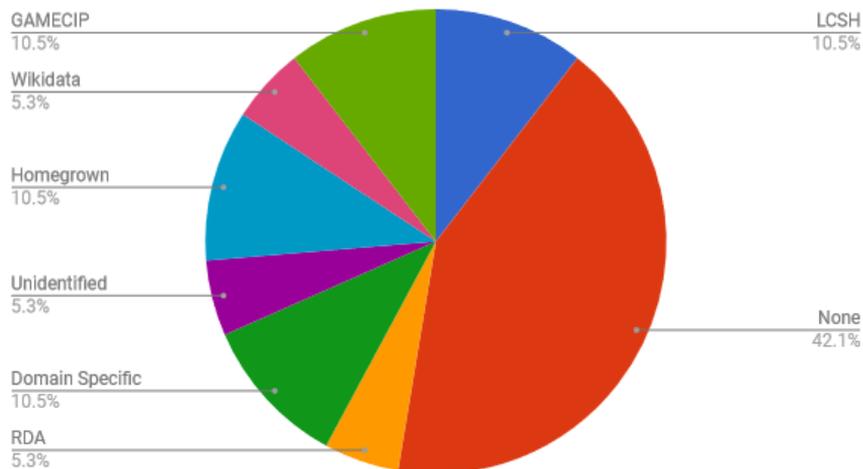


MDWG Survey: Preserving Software? Yes

What metadata standards are you using?

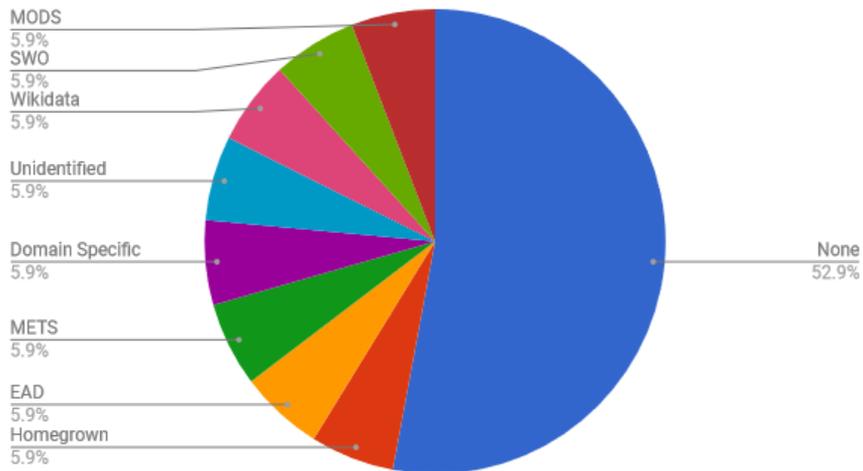


What controlled vocabularies are you using?

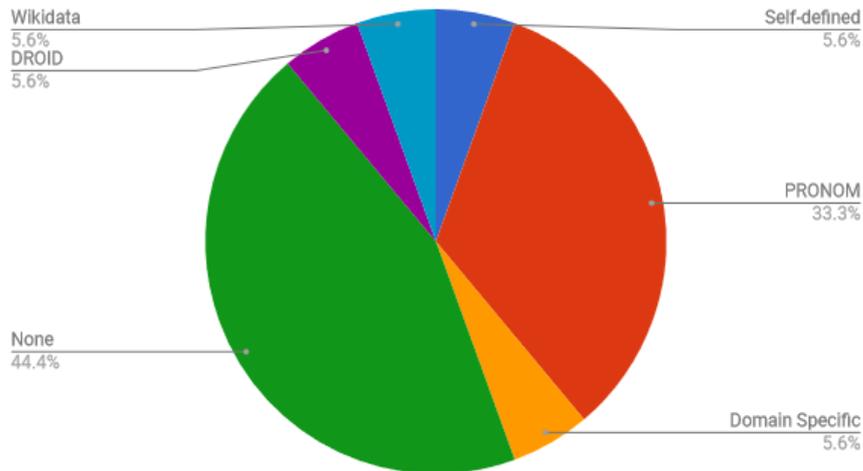


MDWG Survey: Preserving Software? Yes

What schemas/ontologies are you using?

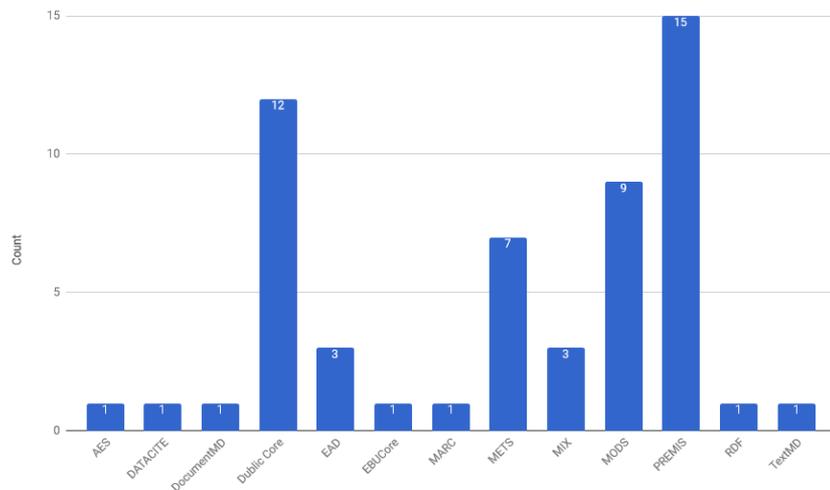


What file format registries are you using?

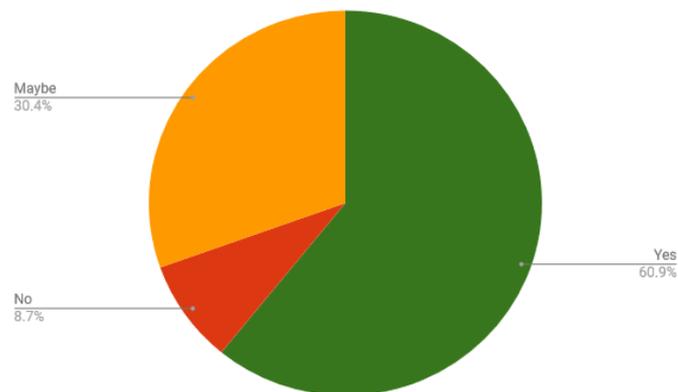


MDWG Survey: Preserving Software? No

What metadata standards are you using?



Will this metadata be useful?



Takeaways

- Limits to existing standards and tools
- Need for Real World examples of existing metadata solutions for software preservation
- Considerations for machine-actionable metadata to support emulation as a service and other applications

Full survey: <https://osf.io/4878z/>

Metadata Schemas for Software (...so far)

Registries

GDFR

PRONOM

Wikidata

NSRL

UDFR

Projects

TOTEM

CodeMeta

SEON

Software Heritage

Standards

PREMIS

GameCIP

MARC/RDA

Dublin Core

MODS

LoC Software Formats

DOAP

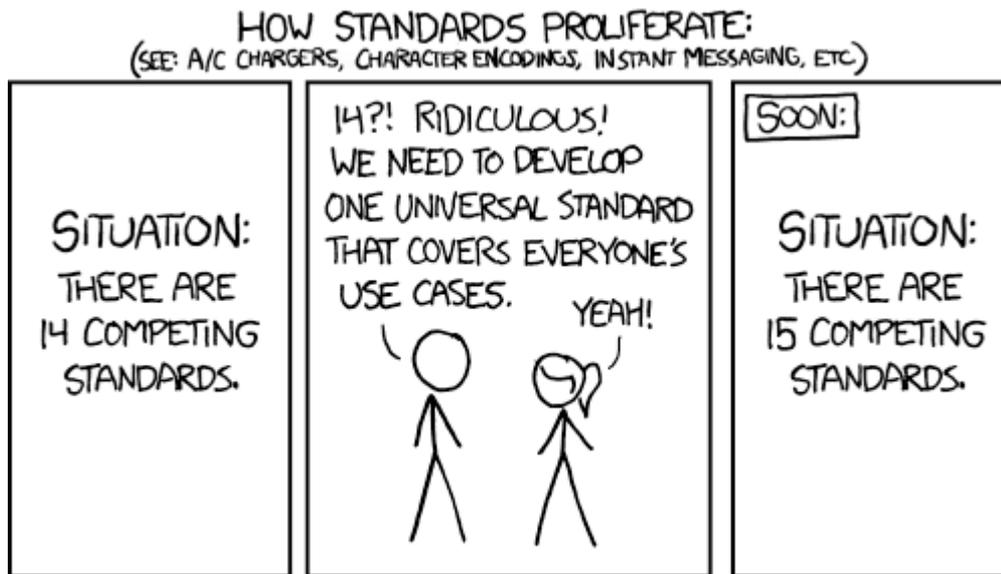
Schema.org

softwarepreservationnetwork.org

@softpresnetwork

<https://osf.io/a7uea/>

Metadata Schemas for Software (...so far)



<https://xkcd.com/927/>

Documenting the Metadata Landscape for Software

Software Metadata crosswalk-draft ☆

File Edit View Insert Format Data Tools Add-ons Help Last edit was made yesterday at 1:38 PM by Elizabeth Roke

noonan.37.osu.edu@gmail.com

Comments Share

100% \$ % .0 .00 123 Arial 10 B I U A

Semantic unit

	A	B	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Semantic unit	Definition of Semantic Unit	Software Heritage https://www.softwareheritage.org/	Wikidata PID	Wikidata property name	National Software Reference Library	GAMECIP Schema/BPG	MARC/RDA	DC	MODS	Library of Congress software recommended formats	UDFR (http://www.udfr.org/onto/onto.rdf)		Counts
41	mime type	Two-part identifier for file formats and format contents transmitted on the Internet. Also called media type.		P1163	media type									2
42	Media Format (of physical media)			P1163			Media Format	338 computer chip cartridge \$2rdacarrier	format	media_type1 (repeatable)				7
43	Memory requirements	Minimum memory requirements												
44	Name	term used for identification of a resource		no property is used for this, it is just called label	label	ProductName	Title ; Standard Title ; Variant Title	245 00 Camp Lazio : Leaky Lake games.// 500 Title from cartridge label.	title	title	title			13
45	Namespace	In computing, a namespace is a set of symbols that are used to organize objects of various kinds, so that these objects may be referred to by name												1
	Operating system	An operating system (OS) is system software that manages		P306	operating system	OpSystemCode	Platform	538 System Requirements: Playstation 2 with the NTSC-U/C designation.// 753 Playstation 2	relation	note	Operating system			15

Sheet1 Semantic unit definitions

Explore

Semantic Units

Operating system (15)

Version (14)

Name (13)

Manufacturer (10)

Identifier (10)

Publisher (9)

Contributor (9)

Language-Programming (8)

Language-Natural (8)

Form of Content (8)

Programmer (8)

Copyright date (8)

Subject (7)

Software dependencies (7)

Publication date (7)

Phys. Media format (7)

System requirements (6)

Rights (6)

Phys. Media extent (6)

Content type (6)

Storage repository (5)

Maintainer (5)

Bug tracker (5)

Screenshots (4)

Release/revision info (4)

File type (4)

Download page (4)

Creator (4)

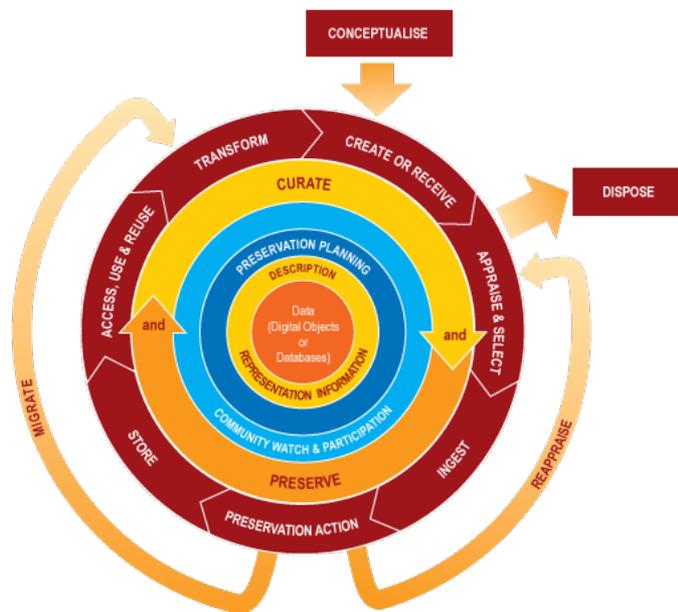
Checksum (4)

Application type (4)

Abstract (4)

DRAFT

Metadata from Creation to Preservation



The DCC Curation Lifecycle Model. Digital Curation Center.
<http://www.dcc.ac.uk/sites/default/files/documents/publication/DCCLifecycle.pdf>

softwarepreservationnetwork.org
@softpresnetwork
<https://osf.io/a7uea/>

Activities Ahead

- Refinement of full set of semantic units for software preservation
- Definition of minimal metadata for various software preservation use cases
- Exploration of WikiData as a tool for translating software metadata standards



Participate!

- Follow us in our OSF repository and on the SPN website
- Contribute to the crosswalk on Github
<https://github.com/softwarepreservationnetwork/metadata-wg>
- Join the effort!



Thank You!

Elizabeth Russey Roke

Elizabeth.roke@emory.edu

 [@elizabethroke](https://twitter.com/elizabethroke)