
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

softwarepreservationnetwork.org
@softpresnetwork
https://osf.io/a7uea/
SPN in Context (of software preservation)

Software Collections (not exhaustive)

- NSRL
- CHM
- Stanford
- Yale
- Internet Archive
- Rhizome
- NASA SP Repo

Projects (U.S.)

- DASPOS
- Building SF
- DCN
- PresQT

Commercial Services

- Code Ocean
- GitHub
- Intact Digital

Programs/Efforts (international)

- SH
- PERSIST
- SSI
- EFGAMP GPS

UNESCO Affiliated
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.
MDWG Survey: Preserving Software? Yes

What metadata standards are you using?

What controlled vocabularies are you using?

- GAMECIP: 10.5%
- LCSH: 10.5%
- Wikidata: 5.3%
- Homegrown: 10.5%
- Unidentified: 5.3%
- Domain Specific: 10.5%
- RDA: 5.3%
- None: 42.1%
MDWG Survey: Preserving Software? Yes

What schemas/ontologies are you using?
- MODS: 5.9%
- SWO: 5.9%
- Wikidata: 5.9%
- Unidentified: 5.9%
- Domain Specific: 5.9%
- METS: 5.9%
- EAD: 5.9%
- Homegrown: 5.9%
- None: 52.9%

What file format registries are you using?
- Wikidata: 5.6%
- DROID: 5.6%
- PRONOM: 33.3%
- None: 44.4%
- Domain Specific: 5.6%
- Self-defined: 5.6%
MDWG Survey: Preserving Software? No

What metadata standards are you using?

Will this metadata be useful?

Yes 50.9%
No 6.7%
Maybe 30.4%
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/

softwarepreservationnetwork.org
@softpresnetwork
https://osf.io/a7uea/
# Documenting the Metadata Landscape for Software

<table>
<thead>
<tr>
<th>Semantic Unit</th>
<th>Definition of Semantic Unit</th>
<th>Software Heritage</th>
<th>Wikidata PID</th>
<th>Wikidata property name</th>
<th>BNF/ISO</th>
<th>National Software Reference Library</th>
<th>GAMA/P</th>
<th>MARC/RDA</th>
<th>DC</th>
<th>MOOS</th>
<th>Library of Congress software recommended formats</th>
<th>UFR/ISNI</th>
</tr>
</thead>
<tbody>
<tr>
<td>mime type</td>
<td>Two-part identifier for file formats and format contents transmitted on the Internet. Also called media type.</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
</tr>
<tr>
<td>Media Format</td>
<td>(of physical media)</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
</tr>
<tr>
<td>Memory</td>
<td>Requirements</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
</tr>
<tr>
<td>Name</td>
<td>Name used for identification of a resource</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
</tr>
<tr>
<td>Namespace</td>
<td>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.</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
</tr>
<tr>
<td>Operating</td>
<td>Operating system</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
<td>11621</td>
</tr>
</tbody>
</table>

Counts: 2, 7, 13, 1
## Semantic Units

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating system</td>
<td>15</td>
</tr>
<tr>
<td>Version</td>
<td>14</td>
</tr>
<tr>
<td>Name</td>
<td>13</td>
</tr>
<tr>
<td>Manufacturer</td>
<td>10</td>
</tr>
<tr>
<td>Identifier</td>
<td>10</td>
</tr>
<tr>
<td>Publisher</td>
<td>9</td>
</tr>
<tr>
<td>Contributor</td>
<td>9</td>
</tr>
<tr>
<td>Language-Programming</td>
<td>8</td>
</tr>
<tr>
<td>Language-Natural</td>
<td>8</td>
</tr>
<tr>
<td>Form of Content</td>
<td>8</td>
</tr>
<tr>
<td>Programmer</td>
<td>8</td>
</tr>
<tr>
<td>Copyright date</td>
<td>8</td>
</tr>
<tr>
<td>Subject</td>
<td>7</td>
</tr>
<tr>
<td>Software dependencies</td>
<td>7</td>
</tr>
<tr>
<td>Publication date</td>
<td>7</td>
</tr>
<tr>
<td>Phys. Media format</td>
<td>7</td>
</tr>
<tr>
<td>System requirements</td>
<td>6</td>
</tr>
<tr>
<td>Rights</td>
<td>6</td>
</tr>
<tr>
<td>Phys. Media extent</td>
<td>6</td>
</tr>
<tr>
<td>Content type</td>
<td>6</td>
</tr>
<tr>
<td>Storage repository</td>
<td>5</td>
</tr>
<tr>
<td>Maintainer</td>
<td>5</td>
</tr>
<tr>
<td>Bug tracker</td>
<td>5</td>
</tr>
<tr>
<td>Screenshots</td>
<td>4</td>
</tr>
<tr>
<td>Release/revision info</td>
<td>4</td>
</tr>
<tr>
<td>File type</td>
<td>4</td>
</tr>
<tr>
<td>Download page</td>
<td>4</td>
</tr>
<tr>
<td>Creator</td>
<td>4</td>
</tr>
<tr>
<td>Checksum</td>
<td>4</td>
</tr>
<tr>
<td>Application type</td>
<td>4</td>
</tr>
<tr>
<td>Abstract</td>
<td>4</td>
</tr>
</tbody>
</table>
Metadata from Creation to Preservation

The DCC Curation Lifecycle Model. Digital Curation Center.

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
@elizabethrroke