# Metadata for Web Archives

Rebecca Guenther
NDSR-NYC Symposium
April 28, 2016

# Development of NYARC metadata profile

- Part of NYARC's Web archiving initiative

- Provide rich metadata for valuable resources that are being archived

- Consider how to integrate with workflow and other records for born digital resources

- Support discovery and use of archived websites
  - Provide enhanced metadata

# Web archiving description

- Describe using archival or bibliographic tradition?

- Limited experience with best practices; many institutions don't have resources to do full descriptions

- Integrate with current workflow in NYARC based on MARC records contributed to OCLC and used in Arcade
  - Creation of a rich record based on MARC fields that integrate into a library catalog

# Flexibility in implementing metadata profile

- Elements have generic names not tied to MARC

- Provides information on encoding in MARC, but also mappings to other metadata schemes
  - Dublin Core
  - MODS
  - BIBFRAME
  - VRA Core
  - Schema.org
  - (need to add EAD)

# Core fields

- Title
- Alternative title
- Names
- Type of resource (Text + additional)
- Genre (Web site)
- Content type
- Media type
- Carrier type

- Language
- Abstract
- Subject/keywords
- URL (live site)
- URL (archived site)
- Rights statement
- Source of description
- Preservation actions

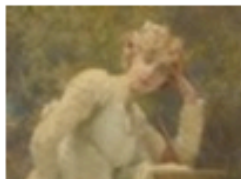| Element Name | Electronic Location (URL) of Live Site | |
|---|---|---|
| **MARC Field** | 856 40 $u $z | |
| **Definition** | The URL of the live site that was archived. | |
| **Expected values** | URI | |
| **Repeatability** | Repeatable | |
| **Obligation** | Mandatory | |
| **MARC Creation/maintenance notes** | Extract URL from Wayback machine in $u<br>Default: $z Current site | |
| **MARC Usage notes** | Repeat if the live site changes its URL and content is considered the same. | |
| **Mappings** | Property/element name | Notes on mapping |
| **DC Property** | Identifier | No way to differentiate live vs archived |
| **MODS Element** | <location><url> | Can use displayLabel to differentiate live vs archived |
| **BIBFRAME Property** | url | To indicate live vs archived site, use structured value: Identifier with identifierValue (url) and label (current site) |
| **VRACore Element** | refid (attribute) | Website might be considered collection or work or image; refid is attribute (need to verify) |
| **Schema.org Property** | url | Archived website would be described separately from live website and have separate url for each |

# THE ART LOSS ■ REGISTER™

www.artloss.com



## The world's largest database of stolen art

Experts around the world use our services to check the provenance of items before they buy or handle them and to record items they own or that have been lost to maximise their chances of recovery. Learn more.

## 4.2. Restitution of Lost or Looted Art

```
OCLC 904558401
Rec stat n      Entered 20150309      Replaced 20150309
Type a          ELvl K          Srce d  GPub   Ctrl     Lang eng
BLvl i          Form o          Conf 0 Freq |  MRec  Ctry xx
S/L 2   Orig    EntW   Regl | Alph
Desc i SrTp w           Cont   DtSt c  Dates 20uu , 9999
006 m o d
007 c ǂb r ǂd u ǂe n
040 FXM ǂb eng ǂe rda ǂc FXM
110 2 Art Loss Register.
245 1 4 The Art Loss Register.
300 1 online resource : ǂb color illustrations
336 text ǂb txt ǂ2 rdacontent
336 still image ǂb sti ǂ2 rdacontent
337 computer ǂb c ǂ2 rdamedia
338 online resource ǂb cr ǂ2 rdacarrier
583 capture ǂc 2014 ǂh New York Art Resources Consortium ǂ5 NyNyARC $2 pet
588 Description of the resource based on live site viewed on March 9, 2015, and archived
site; title from home page.
520 The ALR is the world's largest private database of lost and stolen art, antiques and
collectibles. Its range of services includes item registration, search and recovery services
to collectors, the art trade, insurers and worldwide law enforcement agencies. Users may
request a search of the Art Loss Register's database to discover whether an item has been
registered as missing, fake or stolen. They mays also report the theft or loss of an artwork
or valuable item to the Art Loss Register for registration on the database.
506 1 Registration required.
650 0 Art thefts.
650 0 Art ǂx Expertising.
655 7 Web sites. ǂ2 aat
856 4 0 ǂu http://www.artloss.com/en ǂz Live site
856 4 0 ǂu https://wayback.archive-it.org/org-484/*/http://www.artloss.com/ ǂz
Archived site
```

# Recommendations for further exploration

- Integrating bibliographic vs archival description

- Extending the use of the Metadata Application Profile

- Moving to a Linked Data environment

- Understanding user discovery needs

- Repurposing existing metadata

- Consider metadata needs for long-term preservation of archived websites

# Links

- **Metadata profile:**
http://www.nyarc.org/sites/default/files/web-archiving-profile.pdf

- **Recommendations for further exploration:**
http://www.nyarc.org/sites/default/files/Recommendations%20for%20further%20exploration-final.pdf

- **Questions?**
rguenther52@gmail.com