Describing web archives: a standard with an identity crisis?

Jessica Cebra
Metadata Management Librarian
Basic descriptive metadata requirements for web archives in SearchWorks
- Title of website
- Name of website creator(s) and/or publishers
- Language(s) used in the site
- Brief abstract describing the site
- Subject terms for the site's content
- Website collector name

**Metadata Object Description Schema (MODS)**

- Type of resource (text)
- Genre (archived website)
- Date of website capture
- Form (electronic)
- Digital origin (born digital)
- Internet media type (text/html)
- Original site URL
- "Archived by" note
- Web archiving service note
- Repository (Stanford University Libraries)
- SWAP URL
- PURL
- Record administrative info
**Descriptive Metadata for Web Archiving: Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group**

**DATA ELEMENTS AND USAGE GUIDELINES**

<table>
<thead>
<tr>
<th>Data Element</th>
<th>Usage Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collector</td>
<td>Extent</td>
</tr>
<tr>
<td>Contributor</td>
<td>Genre/Form</td>
</tr>
<tr>
<td>Creator</td>
<td>Language</td>
</tr>
<tr>
<td>Date</td>
<td>Relation</td>
</tr>
<tr>
<td>Description</td>
<td>Rights</td>
</tr>
<tr>
<td>Source of description</td>
<td>Subject</td>
</tr>
<tr>
<td></td>
<td>Title</td>
</tr>
<tr>
<td></td>
<td>URL</td>
</tr>
</tbody>
</table>
# Descriptive Metadata for Web Archiving: Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group

## Data Elements and Usage Guidelines

<table>
<thead>
<tr>
<th>Element</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collector</td>
<td>✓</td>
</tr>
<tr>
<td>Contributor</td>
<td>✓</td>
</tr>
<tr>
<td>Creator</td>
<td>✓</td>
</tr>
<tr>
<td>Date</td>
<td>✓</td>
</tr>
<tr>
<td>Description</td>
<td>✓</td>
</tr>
<tr>
<td>Extent</td>
<td>✓</td>
</tr>
<tr>
<td>Genre/Form</td>
<td>✓</td>
</tr>
<tr>
<td>Language</td>
<td>✓</td>
</tr>
<tr>
<td>Relation</td>
<td>✓</td>
</tr>
<tr>
<td>Rights</td>
<td>✓</td>
</tr>
<tr>
<td>Source of description</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>✓</td>
</tr>
<tr>
<td>Title</td>
<td>✓</td>
</tr>
<tr>
<td>URL</td>
<td>✓</td>
</tr>
</tbody>
</table>
African Politics Web Archive

Type of resource: Text
Date captured: 2007-2015

Digital collection preview

Explore this collection (172 digital items)

Description

Creators/Contributors

Collector: Stanford University. Libraries. Humanities and Area Studies Resource Group

Subjects

Genre: Archived website
<table>
<thead>
<tr>
<th>Title</th>
<th>Picturing the Americas: landscape painting from Tierra del Fuego to the Arctic.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher</td>
<td>Chicago, Illinois: The Terra Foundation for American Art, [2017]</td>
</tr>
<tr>
<td></td>
<td>©2017</td>
</tr>
<tr>
<td>Description</td>
<td>1 online resource: color illustrations</td>
</tr>
</tbody>
</table>

More research tools

Connect to:
- Live site
- Archived site

https://arcade.nyarc.org
| Note | Online supplement to exhibition held from June 20 - October 20th, 2015 at the Art Gallery of Ontario in Toronto, from November 7, 2015 - January 18, 2016 at the Crystal Bridges Museum of American Art in Bentonville, Arkansas, and from February 27 - May 29, 2016 at the Pinacoteca do Estado de São Paulo. Title from caption (viewed on May 13, 2017). Archived websites may not be exact copies of the original website at the time of the web crawl due to technical or privacy considerations. |
| Summary | Online exhibition featuring an interactive map, timeline, artist biographies, and listing of themes, complementing the eponymous traveling exhibition and catalog. |
| Language note | Parallel versions of site in English, French, Spanish, and Portugese. |
| Description source | Description of the resource based on live site viewed on May 13, 2017, and archived site; title from home page. |
| Genre/Form | Web sites. |
| WorldCat no. | 987459783 |

Permanent link to this page: https://arcade.nyarc.org:443/record=b1415935~S16
### Note
Online supplement to exhibition held from June 20 - October 20th, 2015 at the Art Gallery of Ontario in Toronto, from November 7, 2015 - January 18, 2016 at the Crystal Bridges Museum of American Art in Bentonville, Arkansas, and from February 27 - May 29, 2016 at the Pinacoteca do Estado de São Paulo.

Title from caption (viewed on May 13, 2017).
Archived websites may not be exact copies of the original website at the time of the web crawl due to technical or privacy considerations.

### Summary
Online exhibition featuring an interactive map, timeline, artist biographies, and listing of themes, complementing the eponymous traveling exhibition and catalog.

### Language note
Parallel versions of site in English, French, Spanish, and Portuguese.

### System details

### Description source
Description of the resource based on live site viewed on May 13, 2017, and archived site; title from home page.

### LC subject
- Landscape painting -- 19th century -- Exhibitions.
- America -- In art -- 19th century -- Exhibitions.
- Landscape painting -- 20th century -- Exhibitions.
- America -- In art -- 20th century -- Exhibitions.

### Genre/Form
Web sites.

### Related corporate bodies
- Terra Foundation for American Art, issuing body.
- Art Gallery of Ontario, host institution.
- Crystal Bridges Museum of American Art, host institution.
- São Paulo (Brazil : State). Pinacoteca do Estado, host institution.

### WorldCat no.
987459783

Permanent link to this page:
https://arcade.nyarc.org:443/record=b1415935~S16

https://arcade.nyarc.org
bibliographic  archival
bibliographic
identification
discovery
context

archival
origin, authenticity
accountability
provenance, processing
● selecting what and what not to keep

● using a technological tool (which may have limitations of its own) to capture the content, and almost always results in an ‘incomplete’ copy of the original

● crawl configuration

● playback environment and its display behavior
UC Guidelines for Born-Digital Archival Description

Note to readers: Version 1.0 of these guidelines are available on GitHub, where pull requests are welcome and encouraged. Please submit any written feedback to: Annalise Berdini, Charles Macquarie, Kate Tasker, & Shira Peltzman.

Table of Contents:

INTRODUCTION
  Background and Methodology
  Goals
  How to Use These Guidelines
  Systems of Record
  Level of Description

https://github.com/uc-borndigital-ckg/uc-guidelines
DESCRIPTIVE ELEMENTS

- Processor
- Physical Description and Extent
- Abstract
- Conditions Governing Access and Restrictions
- Conditions Governing Reproduction and Use
- Physical Characteristics and Technical Requirements
- Immediate Source of Acquisition

Table of Contents

INTRODUCTION
- Background
- Goals
- How to Use These Guidelines
- Systems of Record
- Level of Description

https://github.com/uc-borndigital-ckg/uc-guidelines
“Decisions made during processing can greatly affect who, what, where, when, why and how researchers access and understand the digital material within a given collection. For us, the Processing Information section is therefore one of the most important aspects of any finding aid that describes born-digital materials. In particular, processing legacy born-digital material can often involve changing the nature of the data to preserve it and make it accessible.”

Berdini et al., “Describing Digital,” p.8
Consider the following questions while describing a group of seeds:

What is this collection for? Who is it for?
What is the general scope and content?
Why these seeds and not others? What was the selection criteria?

A collection-level record is recommended to describe a curated group of web archive seeds. A descriptive overview to highlight a theme, topic(s), event, or other criteria can provide much needed context for a collection of individual websites. Recent literature about the perspective of researchers using web archives collections has identified the user-end need for transparency and understanding the curator’s rationale for the selection of websites to be archived, contributing to an emerging concept of ‘web archives provenance’.
Inspiring works

NYARC Documentation, Metadata Application Profile for Description of Websites with Archived Versions Version 2 (August 2018)

UC Guidelines for Born-Digital Archival Description

Descriptive Metadata for Web Archiving. Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group


Emily Maemura. “What’s Cached is Prologue: Reviewing Recent Web Archives Research Towards Supporting Scholarly Use,” University of Toronto, 2018

Yasmin AlNoamany. “Using Web Archives to Enrich the Live Web Experience through Storytelling,” Old Dominion University dissertation (2016)


Gregory Wiedeman. “Describing Web Archives with the Partner Data API,” Archive-It Partner Meeting, 2018
"Ultimately web archiving is about capturing and recapturing aspects of the experience and performance of the live web, and it's up to collectors, users, and subjects to negotiate together what exactly that means."

Describing web archives: a standard with an identity crisis?

Jessica Cebra, cebraj@stanford.edu
Metadata Management Librarian

Thank you!