

7. STANDARDS FOR DATA DOCUMENTATION

7.1. INTRODUCING STANDARDS FOR DATA DOCUMENTATION

Data documentation has several functions. It records where images, text, and other forms of digital data are from, how the data was generated and what form it is in, and describes its location and use, so that users know what information is where and which data it describes. This data is often generated and maintained on an ad hoc basis, since while it is clearly useful information it may not be clear how it will be utilized over the long-term.

It is all too easy to spend many hours of work documenting a panorama only to end up with a mass of data that makes no sense to anyone outside the project. Or, the data may not match up with the system used by the library or archive that will be distributing and preserving the panorama. To avoid this problem, academic and archival communities have developed sets of data documentation standards which are accepted and supported by large segments of these communities. If the commissioner has arranged in advance for the project to be collected and archived by a museum or library, the project staff will probably know to generate or modify documentation according to that institution's accepted standards. If not, it may seem easier to develop internal methods for tasks such as naming files and keeping track of copyright information. Unfortunately, this strategy is likely to backfire in the long run. If the project migrates to an archive five years after it has been created, important information such as the name of the photographer or permissions may have been lost, corrupted, or just never written down. Using widely accepted standards for generating and organizing data means that the data will be in a recognized and supported form.

It is infinitely easier to generate and organize this information correctly when the work is first done rather than patching it together later, so the project commissioner is strongly encouraged to teach everyone working on the project how to document and maintain the project's data.

7.2. DOMAIN SPECIFIC DATA DOCUMENTATION STANDARDS

The Arts and Humanities Data Service's AHDS Visual Arts center in the UK distributes a guide on "Creating Digital Resources for the Visual Arts," written around 2000. It has useful advice on data documentation and domain specific data documentation standards. Domain specific documentation refers to standards that are accepted or required by certain groups (e.g., standards that are used by specific fields of study or types of research). This is outside of the parameters of this guide and is best done by the commissioner.

The AHDS Visual Arts guide suggests looking at three factors when deciding which standards to use: fitness for purpose, reputation, and existing experience. The standards used should be appropriate and relevant to the work and its use, well established and documented, and (ideally) be familiar to the project staff or whoever is handling data documentation.

7.3. INTRODUCTION TO METADATA

Metadata is data that describes data. In the case of photography, it includes the subject of the photo, where and when it was taken, who was the photographer, and what equipment was used. In the case of digital works in a library or archives, metadata is generated and used to control, distribute, and maintain institutional resources. This information is vital to the use and long-term survival of digital panoramas, not only because it allows them to be properly cataloged, but also because it enables the work to be preserved. Properly planned, metadata not only describes the content of the work but also how it was built and how it can be stored and (if necessary) repaired or duplicated. Metadata standards such as METS can also include rights information and contact information for the work's creators.

Depending on the project, you can store metadata in different forms. Text file formats such as HTML and XML can make metadata part of the content. Image files can store metadata in a header file or in a separate file or database. Audio visual file metadata can be stored in a separate file or database.

There are various categories and standards that are commonly used by archives and libraries, such as METS, Dublin Core, MIX, SMIL, and MARC. Note that different standards accomplish different goals and that metadata is often grouped into categories. Broadly speaking, administrative metadata tracks and manages data; descriptive metadata identifies and describes data; and structural metadata records the relationships between the data. If a panorama is going to be collected and distributed by a library or archive, the commissioner or project staff should consult a metadata expert at the library or archive to be sure that the proper information is generated and recorded. Photographers and programmers often keep this information in their heads or in their own records. In order to be sure that it is gathered accurately and in a useful form, the project staff may want to develop forms to be filled in while on-site and during post-production.

7.4. CONTROLLED VOCABULARY

A controlled vocabulary is a set of terms for defining or describing information resources, such as subject headings and index listings. It is similar to a multiple choice list, where the user can choose from a set of pre-determined phrases to describe a particular attribute. Controlled vocabularies are particularly useful for large and complex projects, since they ensure that the information will be formatted and generated consistently.

If the project is going to be collected and archived by a library or archive, the project staff may need to use specific controlled vocabularies for metadata or cataloging purposes.

7.5. RESOURCE DISCOVERY METADATA

When a user looks up a book in a library catalog, he or she is referred to as discovering a resource. Discovery metadata is information that identifies and describes a resource and its location in a physical or digital collection, such as the title, author, publication data, call number, and URL. There are several commonly used standard sets of discovery metadata, although individual collections or groups of collections may have their own discovery metadata. The Library of Congress has a page listing standards for research description, digital libraries, and resource retrieval protocols. Dublin Core, is a well-known and widely used system and probably the easiest to work with. The list of elements that it covers includes:

- Title
- Subject
- Description
- Type
- Source
- Coverage
- Creator
- Publisher
- Contributor
- Rights
- Date
- Format
- Identifier
- Language
- Audience
- Provenance

The Visual Resources Association Core 4.0 is designed for cultural heritage resources. Documentation includes descriptions of elements and examples.

This information is relevant to a panorama if the project is going to be collected and archived by a library or archive. In this case, the project staff may need to follow that institution's guidelines for generating this type of information.

7.6. TERMINOLOGY RESOURCES AND CLASSIFICATION SCHEMES

Vocabulary databases and classification schemes are standard controlled vocabularies that are used by groups of researchers in certain academic fields. The J. Paul Getty Trust Vocabulary Program offers three sets of vocabularies databases: the Art and Architecture Thesaurus, the Union List of Artist Names, and the Getty Thesaurus of Geographic Names. These are commonly used in art, art history, architecture, and so forth.

The Library of Congress subject headings are a commonly used classification scheme and are published in a multi-volume set of its subject headings.

7.7. PHYSICAL GEO-REFERENCING

Geo-referencing matches a project with a real-world physical location. It can be as simple as a map reference for the city or area where a panorama was shot or as detailed as a GPS mash-up with GoogleEarth. The tools and possible applications are constantly changing and a useful discussion is outside the scope of this guide. Web sites such as Google Maps Mania and GISuser.com may be good sources of inspiration and instruction. However, note that any geo-referencing should be documented according to commonly accepted metadata standards.

WEB REFERENCES

Creating Digital Resources for the Visual Arts.

http://vads.ahds.ac.uk/guides/creating_guide/contents.html

Dublin Core. <http://dublincore.org>

J. Paul Getty Trust Vocabulary Program.

http://www.getty.edu/research/conducting_research/vocabularies/

GISuser.com. <http://gisuser.com>

GoogleEarth. <http://earth.google.com>

Google Maps Mania. <http://googlemapsmania.blogspot.com/>

Library of Congress standards. <http://www.loc.gov/standards/>

MARC. <http://www.loc.gov/marc/>

METS. <http://www.loc.gov/standards/mets/>

MIX. <http://www.loc.gov/standards/mix/>

SMIL. <http://www.w3.org/AudioVideo/>

Visual Resources Association Core 4.0.

<http://www.vraweb.org/projects/vracore4/index.html>