Metadata & Taxonomies: Applications and WFS

Beth Moser (09616)

Alexandra Proserpio (09616)

Lisa Wishard (09616)

March 17, 2005

Sandia National Laboratories

What is metadata?

- **Definition : Data about the data**
- **Descriptive information about objects**

Source: WordNet ® 2.0, © 2003 Princeton University meta-data

<data> /me't*-day`t\*/, or combinations of /may'-/ or (Commonwealth) /mee'-/; /-dah`t*/ (Or "meta data") Data about data. In data processing, meta-data is definitional data that provides information about or documentation of other data managed within an application or environment.

For example, meta-data would document data about data elements or attributes, (name, size, data type, etc) and data about records or data structures (length, fields, columns, etc) and data about data (where it is located, how it is associated, ownership, etc.). Meta-data may include descriptive information about the context, quality and condition, or characteristics of the data.

Sandia National Laboratories

Why metadata?

- **Facilitates exchange of information**
- **Facilitates finding information (structured vs. unstructured content)**
- **Facilitates determining if what you've found is what you want**
- **May provide information about an object that isn't in the object (provenance)**
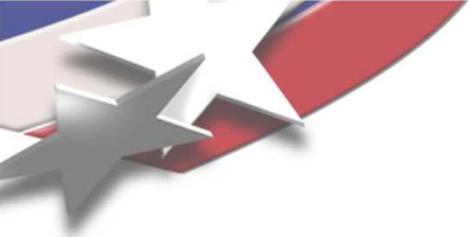
Metadata standards

Standards incorporated in WFS

– Dublin Core

– NWC ICSI Metadata Exchange Guide

Other Standards

– AACR2 / MARC – Library standards

– W3C (e.g. RDF (Resource Description Framework), etc.)

– FIPS – Federal Information Processing Standards (e.g. Digital Signature Standard (DSS), FIPS 186-2, etc.)

– FGDC – Federal Geographic Data Committee (e.g. Content Standard for Digital Geospatial Metadata, FGDC-STD-001-1998 & Spatial Data Transfer Standard (SDTS), FGDC-STD-002, etc.)

Sandia National Laboratories
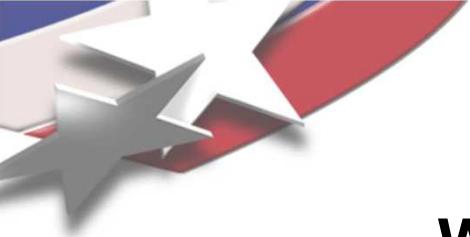
Where does metadata come from?

Example Document:

Guide to Preparing SAND Reports

- Extracting metadata from the document
- Capturing required information for the specified system, e.g. MARC/AACR2 vs. WFS

Report Id

SAND REPORT

SAND2002-2068P
Unlimited Release
Printed September 2002
Supersedes SAND98-0730
dated May 1998

Date

Guide to Preparing SAND Reports and Other Communication Products

Version 1.0, August 2003

Title

Prepared by
Creative Arts Department 12620
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a
Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Adminisration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Originating Or

Sandia National Laboratories

Bib: 183805

Status: cc   Created: 11/11/2002   Updated: 7/1/20(
Owned by: Unowned   Items
Field: Leader [Optional]

000			RecS: c	Type: a	BLvl: m
			ELvl: _	CatF: a	Link: _
008			Date: 021111	DtSt: s	Dates: 2002
			Dates: ____	Ctry: nm_	Ills: ____
			Audn: _	Form: _	Cont: t___
			GPub: _	Conf: _	Fest: _
			Indx: 0	Fict: _	Biog: _
			Lang: eng	MRec: _	Srce: _

027	_	_	¤a SAND2002-2068P
072	_	7	¤2 snlsc ¤a 05
110	2	0	¤a Sandia National Laboratories, ¤c Albuquerque, NM
245	1	0	¤a Guide to preparing SAND reports and other communication products.
250	_	_	¤a Version 01
260	_	_	¤c June 2002.
300	_	_	¤a 92 p.
355	0	_	¤a Unclassified ¤b Unlimited Release
500	_	_	¤a Revised August 2003
700	1	0	¤a Johnson, Debra M.
700	1	0	¤a Brittenham, Phillip W.
780	0	0	¤a Sandia National Laboratories, ¤d May 1998 ¤r SAND98-0730 ¤u SAND98-0730
856	7	_	¤2 http ¤u https://wfsprod01.sandia.gov/groups/srn-uscitizens/documents/document/wfs073820.pd
919	_	_	¤a Suffix
995	_	_	¤a public

Report Id

Originating Org.

Title

Date

Author

Content ID Information Details

Actions: Select an action ▼

Title

Author

Originating Org

Report ID

Date

Content ID: WFS073820 — To send this URL to others in email: IE users:right click, Copy Shortcut, then Edit/Paste Special/Text. Netscape users: right click, then Copy Link Location and paste. Mac users, hold down the mouse button over the link.

Revision: 4

Web Location: https://wfsprod01.sandia.gov/groups/sm-uscitizens/documents/document/wfs073820.pdf

Get Native File: PUBSGUIDE08-11-03.doc (22631 kb) — (To save this file to your PC, IE users: right click/save target as, Netscape users: right click/save link as, Mac users: hold down the mouse button over the link and choose a save option.)

Formats: application/msword, application/pdf

URL/Location:

Submitter: SANDIA\dmjohns

Status: RELEASED

Title: Guide to Preparing SAND Reports and Other Communication Products

Author(s): BRITTENHAM, PHILLIP; JOHNSON, DEBRA M. (Debbie)

Email From:

Email CC:

Originating Org(s): SNL

Report ID: SAND2002-2068P

Release Date: 8/11/2003 3:39 PM

Folder: /Contribution Folders/Infrastructure/IIS-Shared Information/Technical Library/Electronic Resources/

Keywords: Electronic Resources; R&A process; Common Look and Feel requirements; SAND reports; publications; posters; web sites; communication products; printing; trademarks; markings and legends; checklists; report samples

Description/Comments: This guide describes the Review and Approval process, Common Look and Feel requirements, and publishing procedures for communication products at Sandia.

Local:

Information Sensitivity: UNL

Access Control Groups: us-employees, us-contractors

Metadata viewable by all Sandians?: Yes

Group/Checkout Options: Only the submitter can checkout this document

Information Content: Document - Incl. design docs, minutes, presentations, test reports

Dublin Core Categories of Metadata

- **<u>Descriptive</u>:** Metadata that supports the discovery of an object
- **<u>Rights</u>:** Used to provide a link to information about rights held in and over the resource
- **<u>Technical</u>:** Metadata that documents the creation and the digital characteristics of the files
- **<u>Dublin Core Elements</u> (fields)**
  - **Content, e.g. title, subject, description, type, source, relation coverage**
  - **Intellectual, e.g. creator, publisher, contributor, rights**
  - **Instantiation, e.g. date, format, identifier, language**
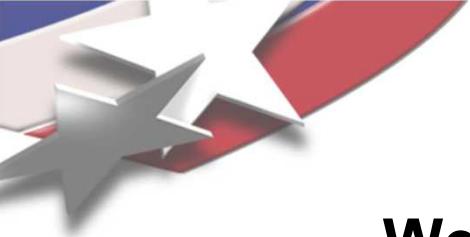
Web FileShare categories of metadata

- **Links and revision information**
- **Citation information**
  - Title, Author(s), **Originating Org(s), Report ID,** Release Date
- **Content description**
- **Classification and Access information**
- **Document type**

Content ID Information Details

Actions: [Select an action ▼]

Technical:
links and revision info

Descriptive:
citation info, e.g. title, author, etc. and content information

Rights:
classification & access

Content ID: WFS073820 — To send this URL to others in email: IE users:right click, Copy Shortcut, then Edit/Paste Special/Text. Netscape users: right click, then Copy Link Location and paste. Mac users, hold down the mouse button over the link.

Revision: 4

Web Location https://wfsprod01.sandia.gov/groups/sm-uscitizens/documents/document/wfs073820.pdf

Get Native File: PUBSGUIDE08-11-03.doc (22631 kb) — (To save this file to your PC, IE users: right click/save target as, Netscape users: right click/save link as, Mac users: hold down the mouse button over the link and choose a save option.)

Formats: application/msword, application/pdf

URL/Location:

Submitter: SANDIA\dmjohns

Status: RELEASED

Title: Guide to Preparing SAND Reports and Other Communication Products

Author(s): BRITTENHAM, PHILLIP; JOHNSON, DEBRA M. (Debbie)

Email From:

Email CC:

Originating Org(s): SNL

Report ID: SAND2002-2068P

Release Date: 8/11/2003 3:39 PM

Folder: /Contribution Folders/Infrastructure/IIS-Shared Information/Technical Library/Electronic Resources/

Keywords: Electronic Resources; R&A process; Common Look and Feel requirements; SAND reports; publications; posters; web sites; communication products; printing; trademarks; markings and legends; checklists; report samples

Description/Comments: This guide describes the Review and Approval process, Common Look and Feel requirements, and publishing procedures for communication products at Sandia.

Local:

Information Sensitivity: UNL

Access Control Groups: us-employees, us-contractors

Metadata viewable by all Sandians?: Yes

Group/Checkout Options: Only the submitter can checkout this document

Information Content: Document - Incl. design docs, minutes, presentations, test reports
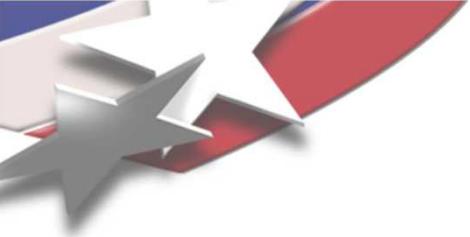
So….where do we go next?

"After all, the Library of Congress wouldn't be of much value if all the books were organized on the floor. The way information is presented and organized becomes as important as the content."

Richard Saul Wurman, Information Architect

Business 2.0, November 2000

Taxonomy

- **The science, laws or principles of classification**
  - **From the Greek: rules of arrangement**

- **A hierarchical arrangement usually expressed parent/child/grandchild relationships that describes a subject domain**

- **A taxonomy is used to populate with content. Populating a taxonomy is called categorization**

Taxonomy: Benefits

- **Presents information in an organized and consistent way**
  - **Reinforces a single voice**

- **Encourages data sharing**
  - **Improves search, navigation, content management and information discovery**
  - **Promotes re-usability and re-purposing of information**
  - **Supports a common language for site architecture**

- **Eliminates duplication of efforts**
  - **Saves resources and expense when information can be retrieved rather than recreated**
  - **Increases collaboration and offers a competitive advantage**

Web FileShare   [   ] Quick Search

Home   ▼ Search   New Check In   My Profile   ▼ Help   Privacy & Security   Add To Favorites   Refresh

▶ My Web FileShare

▼ Browse Content

⊞📁 Library Query Folders
⊞📁 Contribution Folders
  ⊞📁 Capabilities
  ⊞📁 Corporate Operations & Directions
  ⊞📁 Emerging Threats SBU
  ⊞📁 Energy & Critical Infrastructure SBU
  ⊞📁 Homeland Security
  ⊟📂 Infrastructure
    ⊞📁 Business Management & CFO
    ⊞📁 CMM - Capability Maturity Model
    ⊞📁 CMMI - Capability Maturity Model Integration
    ⊞📁 Comparison Activities
    ⊞📁 Computer Security Technology
    ⊞📁 Enterprise Architecture
    ⊞📁 Enterprise Data Initiatives
    ⊞📁 ES&H and Emergency Management
    ⊞📁 Facilities
    ⊞📁 IES Program
    ⊞📁 IIS-Access to Data
    ⊞📁 IIS-CIO Enterprise Activities
    ⊞📁 IIS-Competitive Knowledge
    ⊞📁 IIS-Corporate SRN Access Process
    ⊞📁 IIS-Project Acceleration
    ⊞📁 IIS-Shared Information
    ⊞📁 IIS-Software and Information Life Cycle (SILC)

▶ Search

▶ Content Management

Search

Title ❓ [   ]

Author(s) ❓ [   ]   Examples: moser, moser and stinchcomb, moser or stinchcomb, beth moser

Submitter ❓ [   ]   Examples: ecmoser, ecmoser or sstinch

Content ID ❓ [   ]   Example: WFS000123

Release Date ❓ After [   ] and Before [   ]

Keywords ❓ [   ]   Examples: b61, b61 and safety, w76 or w80

Full-text Search ❓ [   ]   (Separate words or phrases with commas)

Sort Results By: [Title ▼] [Asc. ▼] , Show: [25] per Page

[Search]   [Clear]

Taxonomy: Best Practices

- **Reflects the information needs, behavior, tasks and vocabularies of the users**

- **Aligns with the business strategy/purpose**

- **Not overly complex; it's easy to find needed information**

Taxonomy: Tips on how to start

- **Define Scope**
  - What problems are you trying to solve? Who is the primary audience?
- **User Needs & Behavior**
  - What high value content is used most frequently? What content is it that users need but can't find?
  - Search logs, user interviews—formal and informal
- **Organize Terms**
  - Look for themes: are terms related? are there preferred terms and variant terms? For example, define preferred terms and variant terms, e.g., CSU Special Projects--CSUSP
  - Are there any existing vocabularies or structures that you can use?
- **Create initial taxonomy**
- **Test, test and test**
  - Categorize documents and test

Taxonomy: Things we've learned…

- Designate a dedicated team to develop and implement
- Do not get bogged down in minutia; build a larger layer and fill in the gaps
- Use nouns or noun phrases
- Acronyms should be spelled out
- Anticipate information that is not yet created
- Always provide a feedback mechanism
- Usability--test and test often
- Create a taxonomy change file to track changes made to the taxonomy and why
- Have a plan to maintain your taxonomy or don't even start

Sandia National Laboratories

Resources

- **WFS Help**
  - https://wfsprod01.sandia.gov/Robo/BIN/Robo.dll
- **Field Descriptions and Values for SRN WFS as of Version 6.0**
  - https://wfsprod01.sandia.gov/intradoc-cgi/idc_cgi_isapi.dll?IdcService=GET_SEARCH_RESULTS&QueryText=dDocName=WFS001143
- **SNL Web authors web site**
  - http://www-irn.sandia.gov/webmentor/write-pages/corp-require.htm#pubguidelines
- **Dublin Core**
  - http://dublincore.org/index.shtml
- **Taxonomy**
  - Vincent Duran, Alexandra Proserpio, Lisa Wishard, Sandia National Laboratories, "A case study: levels of order in a chaotic world using Verity Intelligent Classifier to create a corporate taxonomy for an intranet search engine," 2003. http://infoserve.sandia.gov/sand_doc/2003/030854p.pdf
  - Wendi Pohs, IBM Software Group, "Taxonomies, Lexicons and Organizing Knowledge," 2002. http://www.infotoday.com/it2002/presentations/pohs.pps
  - Christian Ricci, Saillant Consulting Group, "Developing and Creatively Leveraging Hierarchical Metadata and Taxonomy," 2004. http://www.boxesandarrows.com/archives/developing_and_creatively_leveraging_hierarchical_metadata_and_taxonomy.php