

Discovering Good Practice: Metadata and the NINCH Guide

Dr Ian Anderson

HATII

University of Glasgow

- Background
- Methodology
- Case Studies and Metadata
 - Public Libraries
 - University Libraries
 - Museums and Galleries
 - Independent Projects
 - University Projects
- Challenges
- Trends

- Humanities Advanced Technology and Information Institute (HATII), University of Glasgow
 - Established 1998
 - Research
 - Teaching
 - Support
- National Initiative for a Networked Cultural Heritage (NINCH)
 - Founded 1996
 - 100+ organisations

- 1998 NINCH Best Practices Working Group
- Getty Grant Program
- 1999 HATII contracted
- In collaboration with NINCH Working Group
 - Methodology
 - Research instruments
 - Case Studies
 - Themed Reports
 - Guide
- Oct 2002 *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*

- NINCH Working Group Principles
 1. *Optimize interoperability of materials*
 2. *Enable broadest use*
 3. *Address the need for the preservation of original materials*
 4. *Indicate strategy for life-cycle management of digital resources*
 5. *Investigate and declare intellectual property rights and ownership*
 6. *Articulate intent and declare methodology*

- Literature Review
 - Print
 - Online
 - Grey
- Site Selection
 - Not Scientific Sample
 - Diversity
 - Themes
 - Good Practice
- Good Practice not Best Practice

- Research Instruments
 - Web site surveys
 - Site visits
 - Telephone interviews
 - Correspondence
- Site Visits
 - Digitization Data Collection Instrument for Site Visit Interviews
 - 6 sections, 31 parts, 202 questions
 - 4 researchers, 20 sites, 36 projects, 68 interviewees
 - 90 minutes to 3 hours

- The Sites
 1. The Bancroft Library
 2. Brown University, Women Writers Project
 3. The Online Archive of California
 4. The Art Institute of Chicago, Imaging Department
 5. University of Chicago, The Oriental Institute
 6. University of Chicago Library
 7. Colorado Digitization Project
 8. Library of Congress, Geography and Map Division
 9. Library of Congress, Preservation Reformatting Division
 10. Library of Congress, Prints and Photographs Division
 11. Library of Congress / National Digital Library Program (NDLP)
 12. Cornell University
 13. University of Illinois, Digital Imaging and Media Technology Initiative (DIMTI)

14. Images of England project (National Monuments Record, English Heritage)
15. Göttingen State and University Library, Center for Retrospective Digitization
16. Harvard University, Visual Information Access (VIA)
17. Harvard University Fine Art Museums / Collection Management System
18. Harvard University Library / Library Digital Initiative Program (LDI)
19. Indiana University, Cook Music Library
20. The Church of Jesus Christ of Latter-Day Saints, Genealogical Society of Utah (GSU)
21. University of Michigan
22. Museums and the Online Archive of California
23. New York Public Library (NYPL)
24. National Library of Norway

25. Survivors of the SHOAH Visual History Foundation
26. Indiana University, Thesaurus Musicarum Latinarum (TML)
27. Tufts University (Perseus Project)
28. Library of Virginia (LVA)
29. University of Virginia, Center for Digital History (Alderman Memorial Library)
30. University of Virginia, Electronic Text Center
31. University of Virginia, Institute for Advanced Technology in the Humanities (IATH)
32. University of Virginia, Robertson Media Center (Clemons Library)
33. University of Virginia, Special Collections (Alderman Memorial Library)
34. University of Virginia, William Blake Archive
35. Walker Art Center (WAC) and Minneapolis Institute of Art (MIA)
36. WGBH Media Archives and Preservation Center & Digital Media Production

- Thematic Research
 - text encoding
 - digital preservation
 - asset management
 - rights management
 - quality assurance
- Metadata integrated into Guide
- Appendix B expands

- Public Libraries
 - National Library of Norway, NYPL, Library of Congress, Library of Virginia
- MARC & DC predominate
- EAD & TEI
- Database, SGML, XML implementations
- LC subject, LC thesaurus, CDWA, AAT, in-house name authority files/controlled vocabulary

- Public Libraries
- New York Public Libraries (NYPL)
- Examined OA models
- Selected OAIS
- METS ingest format
- XML object
- Transcription to MARC

- University Libraries
 - Bancroft (UCL), Chicago, Cornell, Harvard (VIA & DIP), Göttingen, Michigan, Virginia (Alderman & Clemons)
- MARC, DC, EAD, TEI
- MOA2, RDF, MAB-2
- Database, XML, SGML implementations
- LC, CDWA, AAT, VRA, TGM, TGN, FGDC, AACR2, CIMI, in-house name authority files/controlled vocabulary

- University Libraries
 - Göttingen State and University Library Centre for Retrospective Digitization (Göttinger DigitalisierungsZentrum GDZ)
 - DAMS (AGORA), RDB driven
 - Satz-Rechen-Zentrum (SRZ) Berlin
 - XML/RDF interface (input/output)
 - MAB-2

- Museums and Galleries
 - Art Institute of Chicago, Walker Art Center & Minneapolis Institute of Art, University of Chicago Oriental Institute, Harvard Fine Arts Museums, Berkley Art Museum and Pacific Film Archive (MOAC)
- Mostly DC, CIMI, MARC, EAD
- Also TEI, MOA2, NISO
- Database, XML, SGML implementations
- LC, CDWA, AMICO, AAT, VRA, in-house name authority files/controlled vocabulary

- Independent Projects
 - Images of England, Genealogical Society of Utah, SHOAH, WGBH
- All bespoke/in-house standards – modified DC, NISO Z39.19
- LC, Royal Commission for Historic Monuments England, in-house vocabularies

- University Projects
 - Brown Women Writers Project, Indiana TML, Tufts Perseus Project, Virginia – IATH, Center for Digital History, E-Text Center, William Blake Archive
- TEI, MARC, DC EAD, ASCII, CIMI
- Own DTDs
- RDB, XML, SGML implementation
- TGN, LC, AAT, TLG
- Own controlled vocabulary, specialist referene works

- What is Good Practice?
- Hybrid Metadata Approach
 - Descriptive
 - Administrative
 - Structural
- Use Standards, but how to negotiate between the ideal and the practical?
 - Adopt
 - Adapt
 - Implement
- Use information professionals
- QC

- Role
 - Both data and service providers
 - Most likely to develop as data and service providers?
- Technical
 - Profusion of metadata schemes
 - Multiple implementations
 - Use of XML DTDs not schemas

- Metadata Content
 - Time consuming
 - Expensive
 - Libraries best placed?
 - Inadequate thesauri, subject headings/keywords
- TheGlasgowStory
 - Unesco, LC, Dewey
 - Football = soccer
 - Pubs = wine bars
 - Trams = street rail
 - Penny Geggies = ????

- Users – Primarily History Project
- UK Academic Historians
- 105 surveys 25 interviews
- What historians want:
 1. Greater detail, especially item level
 2. Catalogues online (including retroconversion)
 3. Indexing & cross referencing
 4. Digitised & electronic sources
 5. Sceptical about accuracy
- Need to provide mediation and context online

- Semantic interoperability - multi-faceted vocabulary
- UK Archival Thesaurus (PRO & University of London)
- Representation in EAD/Qualified DC?
- More Mapping Tools
- Focus on lower levels of Metadata:
 - Item/Content
 - Collection
- Integrating Metadata (hidden metadata)
 - RDF (an alternative to OAI-PMH?)
 - METS (Metadata Encoding and Transmission Standard)
 - SMIL (Synchronized Multimedia Integration Language)

- HATII: <http://www.hatii.arts.gla.ac.uk>
- NINCH Guide: <http://www.ninch.org>

- AACR – Anglo American Cataloguing Rules
- AAT – Art and Architecture Thesaurus
- AMICO – Art Museum Image Consortium
- CDWA – Categories for the Description of Works of Art
- CIMI – Consortium Museum Intelligence
- DAMS – Digital Asset Management System
- DC – Dublin Core
- EAD – Encoded Archival Description
- FGDC- Federal Geographic Data Committee
- LC – Library of Congress (subject headings (LCSH) and thesaurus)

- MARC – Machine Readable Catalogue
- MOA2 – Making of America 2
- NISO – National Information Standards Org
- SGML – Standard Generalised Markup Language
- TEI – Text Encoding Initiative
- TGM – Thesaurus of Graphic Material (LC)
- TGN – Thesaurus of Geographic Names
- TLG – Thesaurus Linguae Graecae
- RDF – Resource Discovery Framework
- ULAN – Union List of Artists Names
- VRA – Visual Resources Association
- XML – Extensible Markup Language