The following MS PowerPoint slide show, entitled

The Metadata Blues, or Metadata in Search of an Audience

was presented
at the

20th Annual
Minnesota Government
Information Technology Symposium

on

December 13, 2001

at

RiverCentre
Saint Paul, Minnesota

by

Bruce Abbott
Data Administrator
MIS Bureau
Minnesota Department of Natural Resources

who can be reached via

Telephone: 651-296-4798
E-mail: bruce.abbott@state.mn.us

The Metadata Blues

or

Metadata in Search of an Audience

Bruce Abbott
Data Administrator
Minnesota Department of Natural Resources

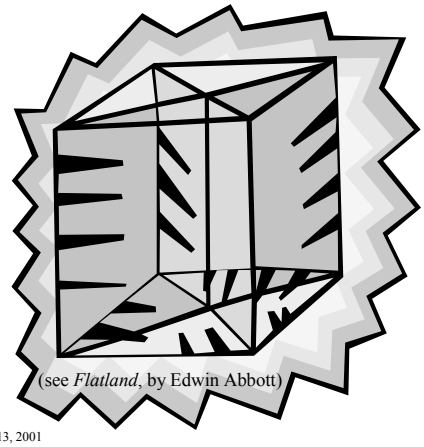
"Metadata Blues" lyrics copyright 2001, Bruce Abbott

Problem:

I got them mean-old, low-down, meta-data blues.

I got them mean-old, low-down, nasty meta-data blues.

"Data 'bout data" has got me so confused.



Dimensions of Metadata

- *Metadata communities*
- Metadata standards
- Metadata domains
- Metadata elements
- Metadata audiences

. December 13, 2001 3

Metadata Community	*Data Element*	*Web*	*GIS*	*Record-Keeping*
Metadata Standards				
Metadata Domain				
Metadata Elements				
Metadata Audiences				

. December 13, 2001 4

Metadata Community	***Data Element***	*Web*	*GIS*	*Record-Keeping*
Metadata Standards				
Metadata Domain				
Metadata Elements				
Metadata Audiences				

. December 13, 2001 5

Database / Data Element Metadata

Those meta-data blues comin' after me.

Those meta-data blues they comin' after me.

They gonna denormalize me, make a
`flat file` outa me.

December 13, 2001 6

Database / Data Element Metadata

They gonna lock my table, throw away the foreign key.

Gonna lock up my table, and throw away the foreign key.

Lord, I'm so tired of messin' with cardinality.

December 13, 2001 7

Data Element Metadata

- Standards
  - ISO/IEC 11179: Specification and Standardization of Data Elements (Parts 1-6, 1994-2000) (http://www.nist.gov/L8/11179)
  - ISO/IEC 11179: Metadata Registries (2001 draft revisions)
    - Part 3: Registry Metamodel
      - Go to http://www.jtc1sc32.org/sc32/jtc1sc32.nsf/DocView750?OpenView&Count=125
      - Choose Document #643 – Part 3 FCD
      - Download 32N0643T.PDF

December 13, 2001 8

Data Element Metadata

- Standards (continued)
  - ISO/IEC *11179: Metadata Registries* (2001 draft revisions)
    - Part 5: Naming and Identification Principles
      - Go to http://www.jtc1sc32.org/sc32/jtc1sc32.nsf/DocView750?OpenView&Count=125
      - Choose Document #657 – Part 5 WD
      - Download 32N0657.PDF
    - Parts 1,2,4,6
      - Go to http://xw2k.sdct.itl.nist.gov/L8/document-library/projects/11179-revision/
      - Look for latest versions in sub-directories
      - IGNORE Part 3 and Part 5 sub-directories

December 13, 2001 9

Data Element Metadata

- Standards (continued)
  - Object Management Group's *Common Warehouse Metamodel* (2000) (http://www.omg.org)
  - Metadata Coalition's *Open Information Model* (1999) – merged into OMG 9/25/2000

December 13, 2001 10

Data Element Metadata

- Domain
  - *11179: Metadata Registries* – Definition, specification, and content of data element dictionaries or metadata registries
  - *Common Warehouse Metamodel* – Easy interchange of warehouse metadata between warehouse tools, warehouse platforms, and warehouse metadata repositories

December 13, 2001 11

Data Element Metadata (11179)

- Elements
  - Data element concepts
  - Data elements
  - Conceptual domains
  - Value domains

December 13, 2001 12

Data Element Metadata (11179)

- Audiences
  - Users (internal and external) of an enterprise's data
  - Systems analysts and data stewards
  - Application software developers
  - Developers of data management software

December 13, 2001 13

Documentation, Data Dictionary, or Metadata Registry/Repository?

- Terms have evolved over time
- 11179 Part 1: Framework (2001-06-29 draft):
  - Data element dictionary: An information resource that lists and defines all relevant **data elements**. See also **register**.
  - Data element registry: An information resource kept by a **registration authority** that describes the meaning and representational form of **data elements**, including registration **identifiers**, **definitions**, **names**, **value domains**, **metadata** and administrative **attributes**, etc. See also **register**.

December 13, 2001 14

Metadata Community	*Data Element*	***Web***	*GIS*	*Record-Keeping*
Metadata Standards				
Metadata Domain				
Metadata Elements				
Metadata Audiences				

December 13, 2001 15

Web Metadata

Forgot what it was that I was lookin' for.

Forgot what it was that I thought I was lookin' for.

Web searchin's got me shook up, down to my Dublin Core.

December 13, 2001 16

Web Page Metadata

- Standards
  - Dublin Core Metadata Element Set, version 1.1 (7/2/1999) [approved 7/2001 as National Information Standards Organization (NISO) Z39.85] (http://dublincore.org)

December 13, 2001 17

Web Page Metadata (Dublin Core)

- Domain
  - Resource discovery for networked (World Wide Web) resources

December 13, 2001 18

Web Page Metadata (Dublin Core)

- Elements
  - Content: title, subject, description, type, source, relation, coverage
  - Intellectual Property: creator, publisher, contributor, rights
  - Instantiation: date, format, identifier, language

December 13, 2001 19

Web Page Metadata (Dublin Core)

- Audiences
  - Web users
  - Web page sponsors
  - Web page builders
  - Search engine developers

December 13, 2001 20

Controlled Vocabulary

I wanna tell you somethin', let the words just tumble free,

Say I wanna tell you somethin', let the words just tumble free,

But the Man has got control over my vocab'lary.

December 13, 2001 21

Controlled Vocabulary

- Associated primarily with Web metadata
- Also useful in other communities
  - 11179 Metadata Registry – classification scheme, as well as value domains, represent controlled vocabularies
  - MN Geographic Metadata Guidelines have a field for reference to a formally registered thesaurus

December 13, 2001 22

Metadata Community	*Data Element*	*Web*	***GIS***	*Record-Keeping*
Metadata Standards				
Metadata Domain				
Metadata Elements				
Metadata Audiences				

December 13, 2001 23

GIS Metadata

Spatial metadata – I ain't gonna take this rap.

Spatial metadata – I just won't take this rap.

"Datum?" "Resolution?" – Man, just give me a map.

December 13, 2001 24

Geographic Information System (GIS) Metadata

- Standards
  - ISO 19115: *Geographic Information – Metadata,* Draft International Standard (8/2001, out for vote 9/2001 – 2/2002)
    - Document: http://xw2k.sdct.itl.nist.gov/L8/Document-library/Other/ISO_DIS_19115.zip
    - ISO / Technical Committee 211 scope: http://www.isotc211.org/scope.htm#scope
    - NCITS / Technical Committee L1 scope: http://www.ncits.org/tc_home/l1.htm
  - Federal Geographic Data Committee's *Content Standard for Digital Geospatial Metadata*, version 2.0 (1998) (http://www.fgdc.gov/metadata/contstan.html) [Harmonization with ISO 19115: http://www.fgdc.gov/metadata/whatsnew/fgdciso.html]
  - MN Governor's Council on Geographic Information's *Minnesota Geographic Metadata Guidelines,* version 1.2 (1998) (http://www.gis.state.mn.us/stds/metadata.htm)

. December 13, 2001 25

GIS Metadata
(MN Geographic Metadata Guidelines)

- Domain
  - Geospatial datasets, mostly at the set level, with some reference to attribute value domains within the set

. December 13, 2001 26

GIS Metadata
(MN Geographic Metadata Guidelines)

- Elements
  - Identification Information
  - Data Quality Info
  - Spatial Data Organization Info
  - Spatial Reference Info
  - Entity and Attribute Info
  - Distribution Info
  - Metadata Reference Info

. December 13, 2001 27

GIS Metadata
(MN Geographic Metadata Guidelines)

- Identification Information
  - Citation
    - Originator
    - Title
    - Unique ID
  - Description
    - Abstract
    - Purpose
  - Spatial Domain
    - Spatial Extent of Data
  - Keywords
  - Access and Use Constraints

. December 13, 2001 28

GIS Metadata
(MN Geographic Metadata Guidelines)

- Audiences
  - GIS dataset users (including policymakers and analysts)
  - GIS dataset developers

. December 13, 2001 29

Metadata Community	*Data Element*	*Web*	*GIS*	***Record-Keeping***
Metadata Standards				
Metadata Domain				
Metadata Elements				
Metadata Audiences				

December 13, 2001 30

Recordkeeping Metadata

"Mandate, rights, retention" is more than I can take.

"Mandate?" "Rights?" "Retention?" – it's more than I can take.

Recordkeeping metadata's too many rules to break.

December 13, 2001 31

Recordkeeping Metadata

- Standards
  - National Archives of Australia's R*ecordkeeping Metadata Standard for Commonwealth Agencies*, version 1.0 (1999) (http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm)
  - Minnesota Recordkeeping Metadata Development Committee's *Minnesota Recordkeeping Metadata Standard*, draft version 1 (2001) (http://www.mnhs.org/preserve/records/metadev.html)

December 13, 2001 32

Recordkeeping Metadata (Minnesota Standard)

- Domain
  - A variety of information systems, file types, and record types, such as: word processing documents, e-mail, databases, data warehouses, web pages, spatial data, graphical files, microforms, videotapes and audio tapes, correspondence, maps

December 13, 2001 33

Recordkeeping Metadata (Minnesota Standard)

- Elements
  - Content: title, subject, description, type, relation, coverage
  - Intellectual Property: agent, rights management
  - Instantiation: date, record identifier, language, format
  - Records management: function, aggregation level, management history, location, disposal, mandate, use history, preservation history, fitness for use/quality

December 13, 2001 34

Recordkeeping Metadata (Minnesota Standard)

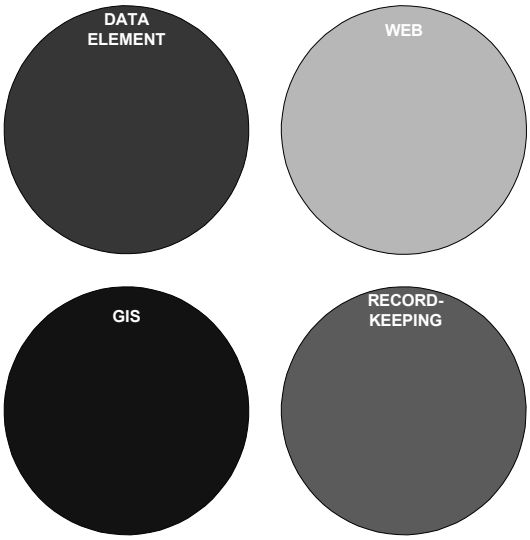
- Audiences
  - Agency recordkeeping staff
  - Agency IT staff
  - Agency information users, developers

December 13, 2001 35

Metadata Community	*Data Element*	*Web*	*GIS*	*Record-Keeping*
Metadata Standards	11179	Dublin Core	MGMD Guidelines	AU / MN Standards
Metadata Domain	Data element metadata registries	Networked (Web) resources	Geospatial datasets	Electronic record sets
Metadata Elements	Data element concepts Data elements Conceptual domains Value domains Classification schemes	Content: title, subject, description, type, source, relation, coverage Intellectual Property: creator, publisher, contributor, rights Instantiation: date, format, identifier, language	Identification Information Data Quality Info Spatial Data Organization Info Spatial Reference Info Entity and Attribute Info Distribution Info Metadata Reference Info	Content: title, subject, description, type, relation, coverage Intellectual Property: agent, rights management Instantiation: date, record identifier, language, format Records management: function, aggregation level, management history, location, disposal, mandate, use history, preservation history, fitness for use/quality
Metadata Audiences	Users (internal and external) of an enterprise's data Systems analysts and data stewards Application software developers Developers of data management software	Web users Web page sponsors Web page builders Search engine developers	GIS dataset users (including policymakers and analysts) GIS dataset developers	Agency recordkeeping staff Agency IT staff Agency information users, developers

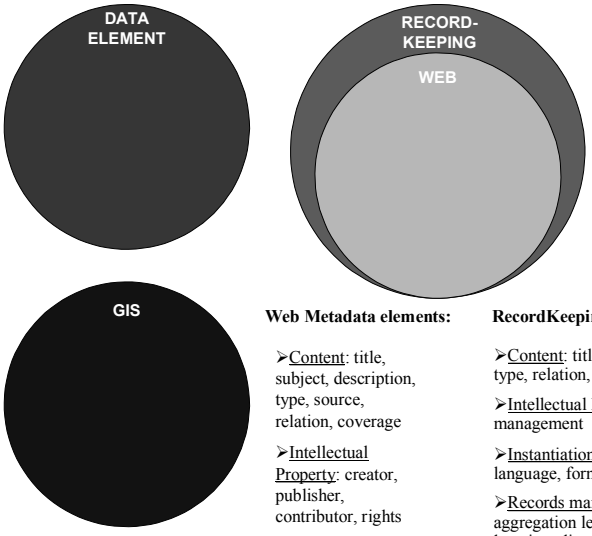
. December 13, 2001 36

Overlapping Domains and Elements



. December 13, 2001 37

Overlapping Domains and Elements



Web Metadata elements:

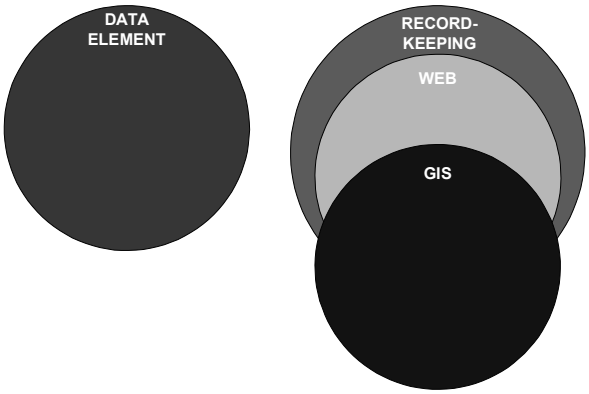
- Content: title, subject, description, type, source, relation, coverage
- Intellectual Property: creator, publisher, contributor, rights
- Instantiation: date, format, identifier, language

RecordKeeping Metadata elements:

- Content: title, subject, description, type, relation, coverage
- Intellectual Property: agent, rights management
- Instantiation: date, record identifier, language, format
- Records management: function, aggregation level, management history, location, disposal, mandate, use history, preservation history, fitness for use/quality

. December 13, 2001 38

Overlapping Domains and Elements



Overlap:

Web, RecordKeeping, and GIS metadata deal generally with collections of data-- information resources such as Web pages, books, thematic GIS data sets.

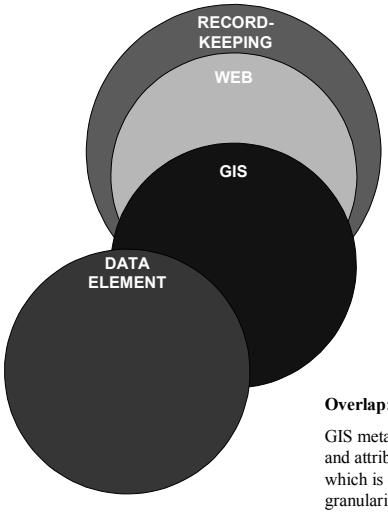
GIS Metadata unique features:

Data quality info

Spatial reference info

. December 13, 2001 39

Overlapping Domains and Elements



Overlap:

GIS metadata contains entity and attribute information, which is comparable in granularity to Data Element metadata.

. December 13, 2001 40

Metadata Production Difficulties

Creatin' metadata is worse than poundin' rocks.

I say creatin' metadata is worse than poundin' rocks.

It musta been invented by a man who sorts his socks.

December 13, 2001 41

Poundin' Rocks

- Difficult to motivate metadata producers
  - Not a new problem – Documentation has always suffered
- Metadata creation tools
  - MetaPro (*Data Element* Metadata Registry Builder - US EPA – http://www.epa.gov/metapro/index.htm)
  - TagGen (*Web* metatag generator, MN Bridges Project, http://www.bridges.state.mn.us/taggen.html)
  - DataLogr (MN *Geographic* Metadata Guidelines, http://www.lmic.state.mn.us/chouse/datalogr.html)

December 13, 2001 42

Metadata Audience Uncertainty

Sweatin' over metadata – who's gonna read this stuff?

I say I'm sweatin' over metadata – who's gonna read this stuff?

Oh, Lord, I try, but the Man says it just ain't good enough.

December 13, 2001 43

Common Audiences

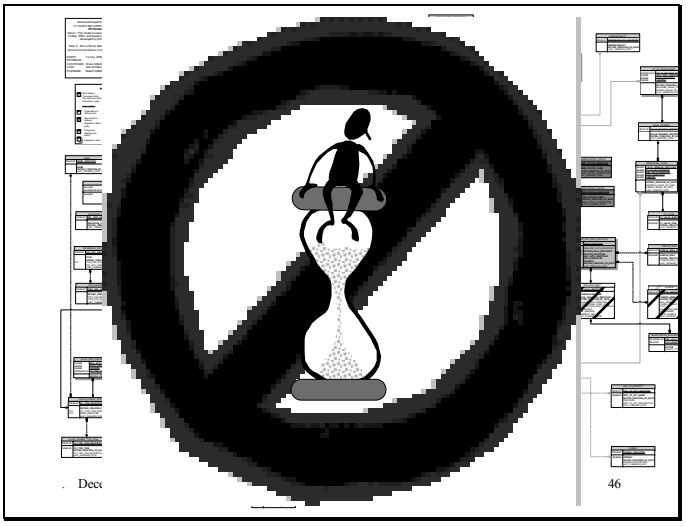
- Across all four metadata communities, audiences can be grouped into three types:
  - **Producers**
  - **Consumers**
  - **Funders**
- Producers are often also consumers (of other data sets' metadata)
- Funders may also be consumers
  - At the very least, they must see metadata as being useful to their constituents

December 13, 2001 44

Marketing to these audiences

- Need to market conceptually as well as technically to producers
- Market conceptually and with examples to consumers, funders
- Describe how metadata effort will meet business needs of particular audience

December 13, 2001 45

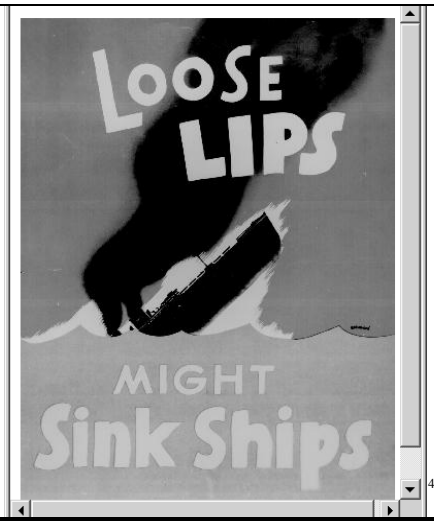


Dec 46

Start Simply:

Develop Catchy Messages

Market to a Broad Audience



December 13, 2001 47

Loose Data Ain't So Greata

Add "Meta"

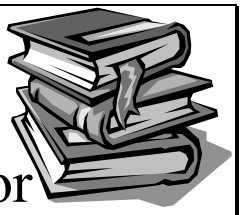
To Your Data

To Make Your Data Greater

- More readable
- Better understood
- Less subject to misinterpretation or misuse
- Less dependent on your memory for understanding
- More valuable

. December 13, 2001 48

Introduce Metadata – to anyone – via Metaphor



Imagine, if you will, a university town that's home to a world-class research library. Centuries worth of literature, and science, and philosophy are stored inside those ivy-covered walls, and have helped many a graduate student to absorb, and digest, and understand information about their field of study. Until one day, a brilliant but sociopathic associate professor of Library Science named L. C. Dewey is denied tenure – and vows to exact a terrible revenge. Late that night, she sneaks into the reference room and randomizes the contents of the card catalog. Then she goes over to the reference shelves with an Exacto knife and, very carefully, removes the tables of contents and the indexes of the three most-used reference sources in the library: the Physician's Desk Reference, the Minnesota Statutes, and Bartlett's (formerly) Familiar Quotations.

My name is Bruce Abbott. I'm the Data Administrator for the Minnesota Department of Natural Resources. And I'll be your host for the next half-hour as we explore the chaos left behind by the diabolical Dr. Dewey, after she destroyed, not the information resources of the research library, but the road maps or pointers or METADATA – data about data – that allowed people to discover, understand, and analyze those information resources.

. December 13, 2001 49

Intentions (What are the business purposes, goals, objectives for this project? What business results are expected? Why is this effort worth doing?)

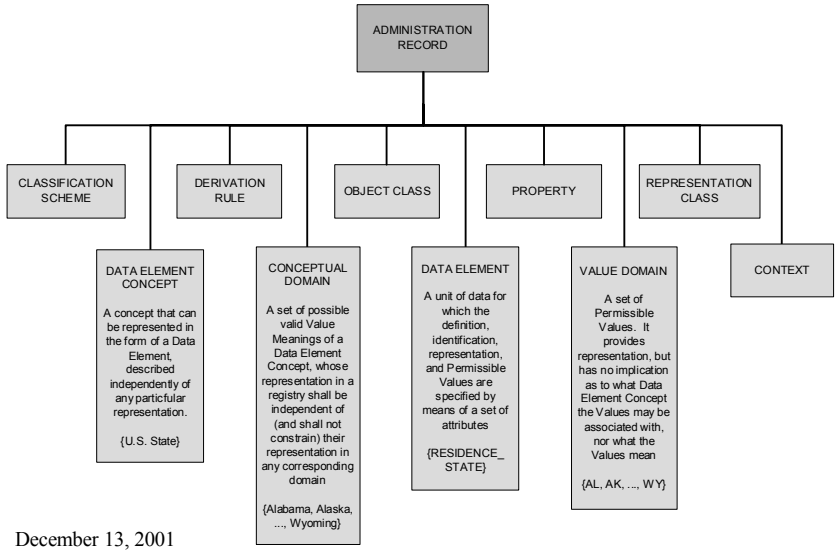
Expected Business Results	Effort Contribution of This Project to Support these Results
1. Improve the quality of and accessibility to state data	• Build a metadata registry with strong quality control standards with an emphasis on user access and accessibility
2. Enable professionals (including the legislature and executive branch) to make better decisions	• Facilitate cross agency comparisons, analysis, and sharing of data. • Help ensure data is accurately interpreted and appropriately used
3. Increase the ability to more quickly respond to public and other (legal, press,…) misrepresentations of data	• Enable the context of data to be known so a response can be quickly made
4. Increase self-reliance for users of state agency data	• Help enable users to create their own reports and appropriately interpret the data
5. Enhance the ability to maintain and extend contractor-built work products after contract programmers leave	• Use metadata registry to hold contractors to both data element and metadata standards

Provide BUSINESS – not technical – reasons for undertaking metadata projects

. December 13, 2001 50

Even for a technical audience, approach metamodel detail and complexity incrementally

(adapted from ISO/IEC FCD 11179-3)



. December 13, 2001 51

Restatement of Problem:

December 13, 2001 52

Opportunity:

I got them I-can-do-it, you-can-do-it, meta-data blues.

Got them he-can-dig-it, she-can-dig-it, meta-data blues.

"Data 'bout data?" – Come on, let's spread the news.

December 13, 2001 53

Special Thanks To:

- Eileen Quam, that blues-moanin' conga-boppin' Web metadata expert, and the rest of the Twin Cities' neo-beat hipsters, *Opiate of the Masses*, for breathing life into the *Metadata Blues* (e-mail*: opiate@MNinter.net*)
- Steve Ring, MN Dept. of Health Data Administrator and a fellow music lover, for proposing the idea to write the *Metadata Blues* (and for improving my 12-bar blues phrasing)
- Data and metadata thinkers and meta-thinkers in Minnesota government, MN DIG-IT (Data Issues Group – Information Technology), and throughout the world-wide 11179 community

December 13, 2001 54