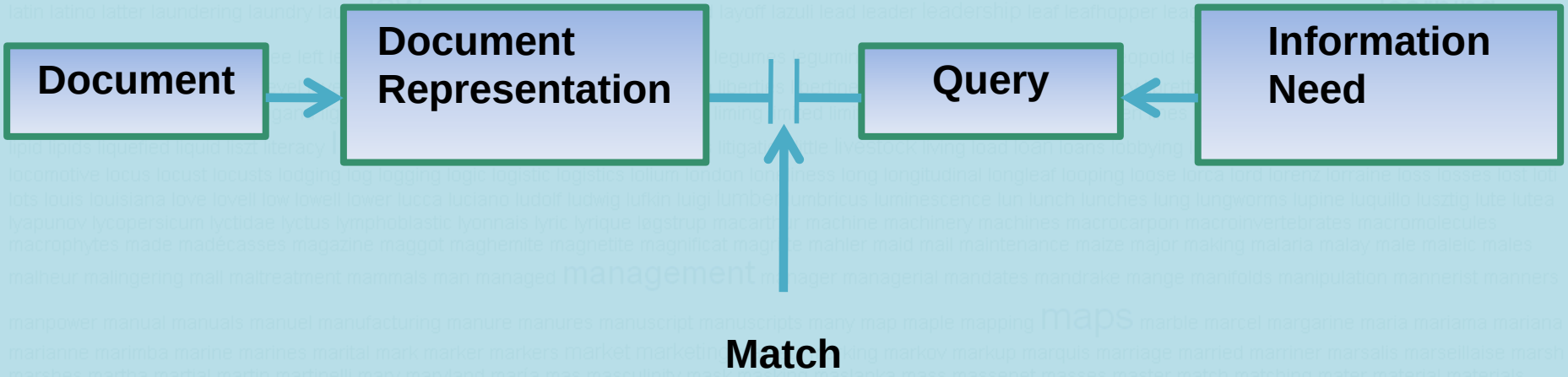
Metadata Quality Assurance: The University of North Texas Libraries Experience

Daniel Gelaw Alemneh & Hannah Tarver

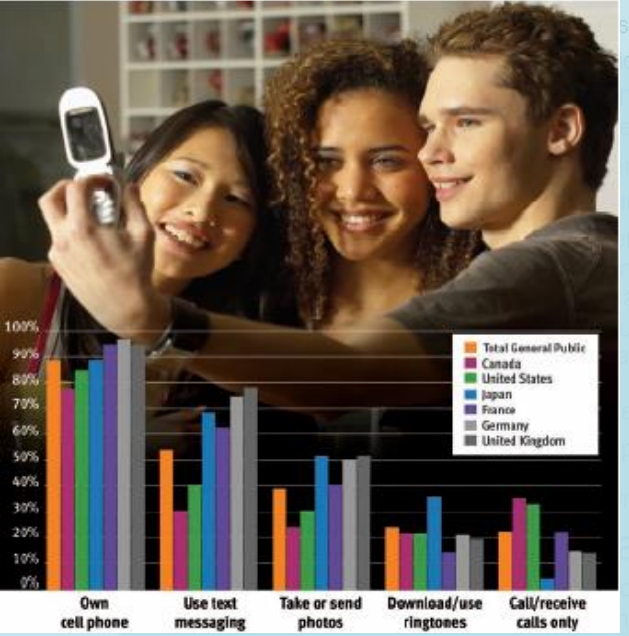
3rd annual Texas Conference on Digital Libraries (TCDL)

May 27-28, 2009

Information Retrieval



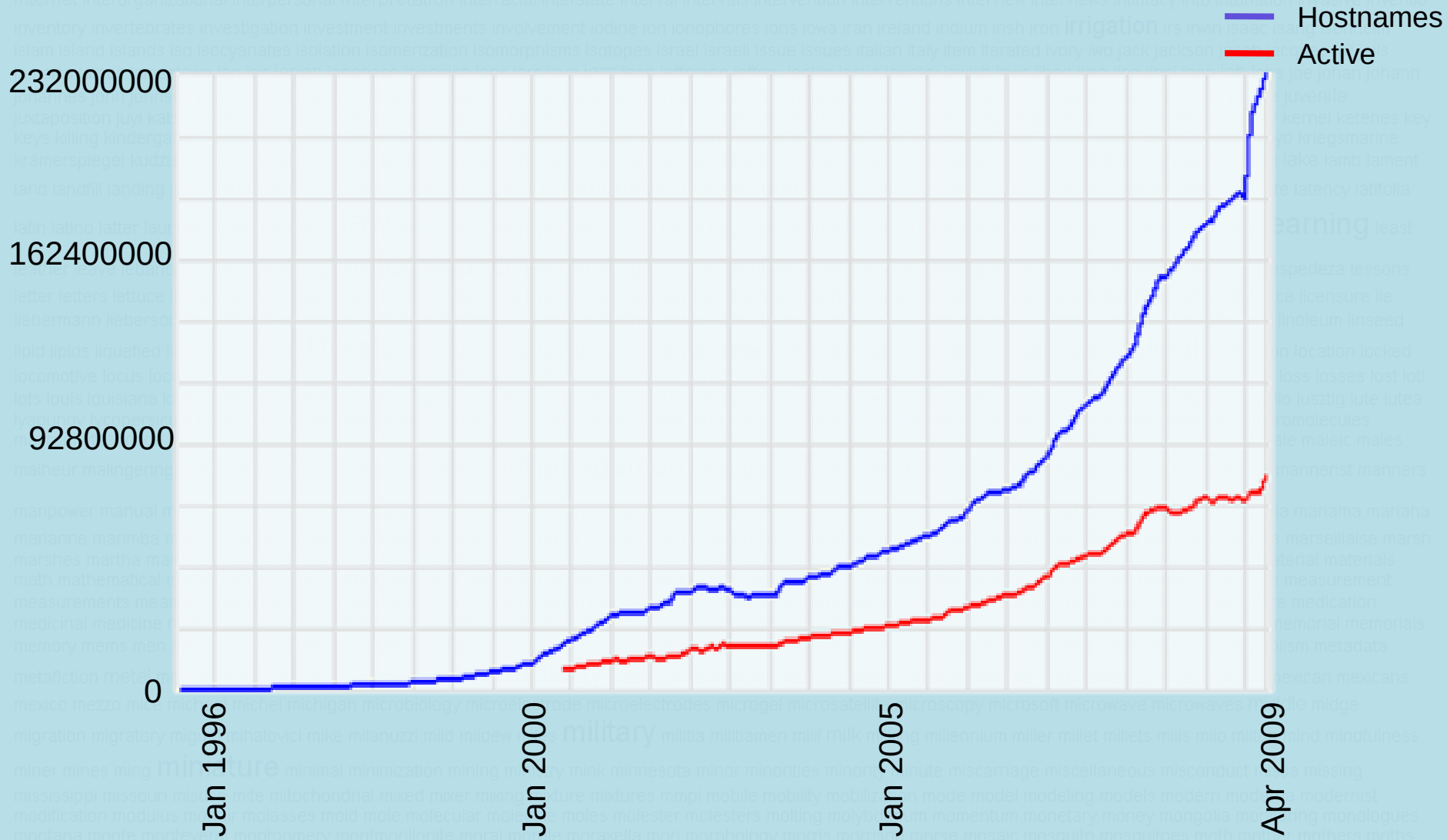
Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13*(5), 407-424.



Trends

- Information creation, organization, retrieval, use, and preservation is becoming more complex
- User as creator, annotator, indexer, searcher, and eventual user of his/her content
- Visualization of the information space instead of a ranked list of search results

Total Sites Across All Domains August 1995-April 2009



UNIVERSITY OF NORTH TEXAS
LIBRARIES
DISCOVER THE POWER OF IDEAS

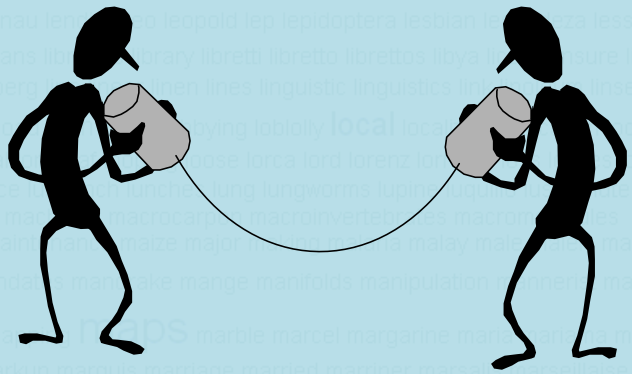
Digital Projects

- **UNT Digital Collections**
- **Portal to Texas History**
  - **100 + Collaborators**
- **Congressional Research Service Archives**
- **Other Statewide and National Projects**

Factors Influencing Metadata Quality

Local Requirements:

- **Objects**
- **Granularity**
- **Functionality**



Collaborative Requirements:

- **Diversity of Users**
- **Interoperability**
- **Digital Rights Issues**

Poor Metadata Quality

Ambiguities

Poor recall

Poor precision

Inconsistency of search results

Common Errors

The data is:

- **Incorrect**
- **Missing**
- **Ambiguous**

Metadata Quality Assurance Mechanisms & Tools at UNT

Pre-Ingest

Training

Creation Tools

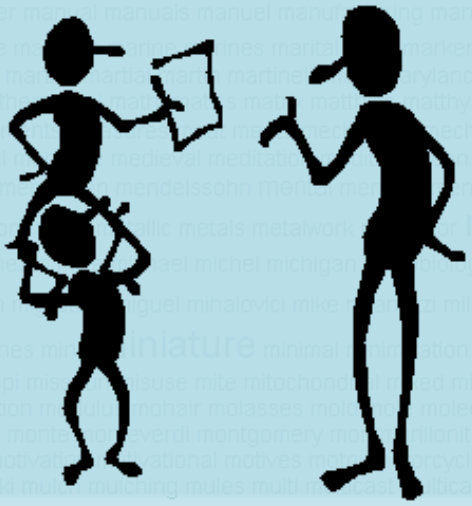
Proofing & Editing Tools

Post-Ingest

Analysis Tools

Training

- Face-to-Face Instruction
- Metadata Schema & Documentation
- Internal Project Wikis
- Staff Support



Metadata Creation Template

Title
Select a value Add a row
The official name given to the resource

Identifier
Select a value Add a row
A unique identifier or 'permanent name' for a resource

Note
This element will serve as a catch-all note field

Institution
Select a value Add a row
A consistent reference to the institution or administrating unit that owns the digital resource for which metadata was created

Collection
Select a value Add a row
Collection title

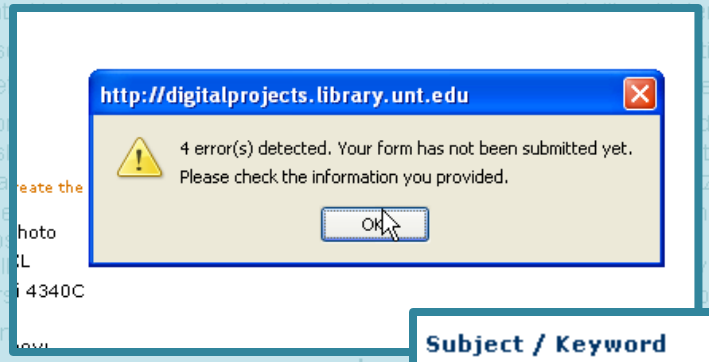
Subject / Keyword
Select a value Add a row
The subject or topic of the resource that succinctly describes the content of the resource

Creator
Creator Name:
Creator Type:
Select a value
Creator Role:
Select a value
Creator Info:
Add a row

Primary Source
Is the document considered a primary source?

Content Description
A textual description of the content or purpose of the resource

Template Validation



Subject / Keyword

submarines

Select a value

This field is required.

Add a row

The subject or topic of the resource that succinctly describes the content of the resource

Creator

Creator Name:

Creator Type:

Select a value

Creator Role:

Select a value

Creator Info:

Add a row

Primary Source

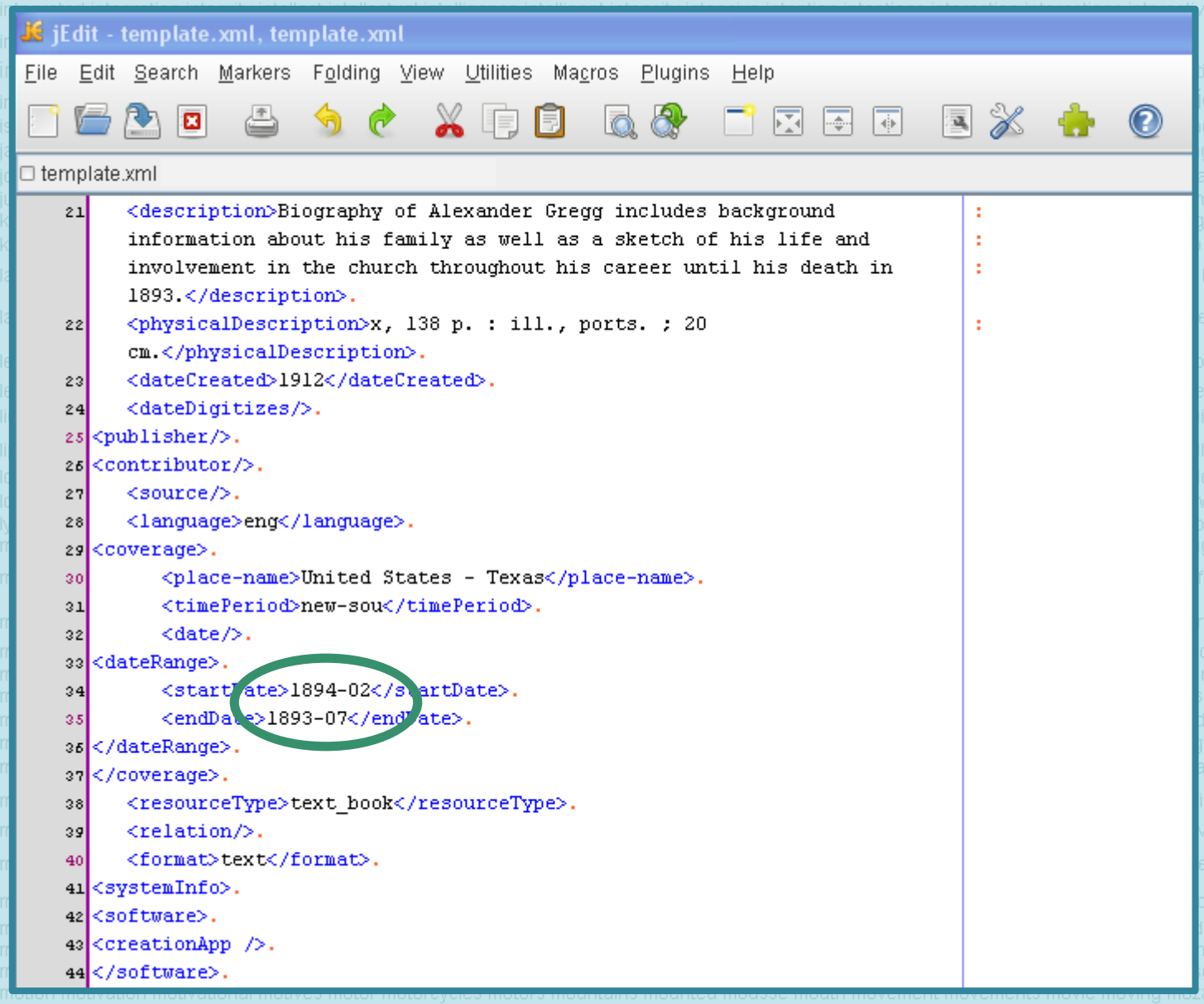
Is the document considered a primary source?

Content Description

This field is required.

Template Reader

Title(officialtitle):	Alexander Gregg: First Bishop of Texas
Identifier(OCLC):	2629011
Note:	
Institution:	UNT
Collection:	THC
Creator	Name: Gregg, Wilson Role: aut Type: per
Subject(LCSH):	Bishops -- Texas.
Subject(LCSH):	Gregg, Alexander, 1819-1893.
Subject(UNTL-BS):	People - Individuals
Subject(UNTL-BS):	Religion - Denominations - Episcopal
Subject(KWD):	biographies
Description:	Biography of Alexander Gregg includes background information about his family as well as a sketch of his life and involvement in the church throughout his career until his death in 1893.
Physical Description:	x, 138 p. : ill., ports. ; 20 cm.
Date Created:	1912
Date Digitized:	
Publisher	
Contributor:	
Source:	
Language:	eng
Coverage:	Place: United States - Texas Era: new-sou Start Date: 1894-02 End Date: 1893-07 error! End-date should be later than start-date
Resource Type:	text_book
Relation:	
Format:	text



jEdit Text Editor

UNT Metadata Analysis Tools: Post-ingest

Reports:

- NULL Values
- List All Values
  - List Values by Institution
  - List Values by Collection
- Browse All Values
  - Browse Values by Institution
  - Browse Values by Collection
- List All Records
  - List Records by Institution
  - List Records by Collection
- Authorities Values

Graphs:

- Records Added Over Time
- Records Added Per Month
- Files Added Over Time
- Map of Texas
  - by Collection
  - by Institution

Other Stuff:

- Search PTH System
  - Search a Collection
  - Search a Institution
- Word Clouds
  - by Collection
  - by Institution
- Word List
  - by Collection
  - by Institution

Values that are more than 90% similar to the previous value will be highlighted. off/on

ignore qualifiers | use qualifiers

all qualifiers

value	count
aut: Hare, Maud Cuney	1
aut: Hartley, Oliver Cromwell, 1823-1859	5
aut: Hartley, Oliver Cromwell, 1823-1859.	13
aut: Hartley, R. K.	12
aut: Hartnett, Minerva	1
aut: Hawker, H. W.	2
aut: Hawker, H. W. (Herman William)	12
aut: Hayes, F. A.	1
aut: Hays, Hugh	1
aut: Hazzard, J. L.	1
aut: Hearn, W. Edward (Williamson Edward)	3
aut: Henderson, J. P.	1
aut: Hendrickson, B. H. (Bertram Higbie)	4
aut: Hendrix, R.	1

Enhanced by
Highlighter – On/Off

ignore qualifiers | use qualifiers

UNTL-BS - (64394)

value
UNTL-BS: Business, Economics and Finance - Medicine - Dentist
UNTL-BS: Business, Economics and Finance - Medicine - Dentists
UNTL-BS: Business, Economics and Finance - Medicine - Doctors
UNTL-BS: Business, Economics and Finance - Medicine - Hospitals
UNTL-BS: Business, Economics and Finance - Medicine - Nurses
UNTL-BS: Business, Economics and Finance - Medicine - Nursing Homes
UNTL-BS: Business, Economics and Finance - Medicine - Red Cross

Enhanced by
Qualifier – Use/Ignore

Metadata Analysis: NULL

DC Element	Number that are NULL
DC Title	22
DC Creator	26390
DC Subject	1516
DC Description	0
DC Contributor	39967
DC Publisher	41885
DC Identifier	4942
DC Source	43613
DC Relation	45914
DC Language	271
DC Coverage	1241
DC Date	2125
DC Resource Type	287
DC Format	194
DC Rights	16171
UNTL Degree	46826

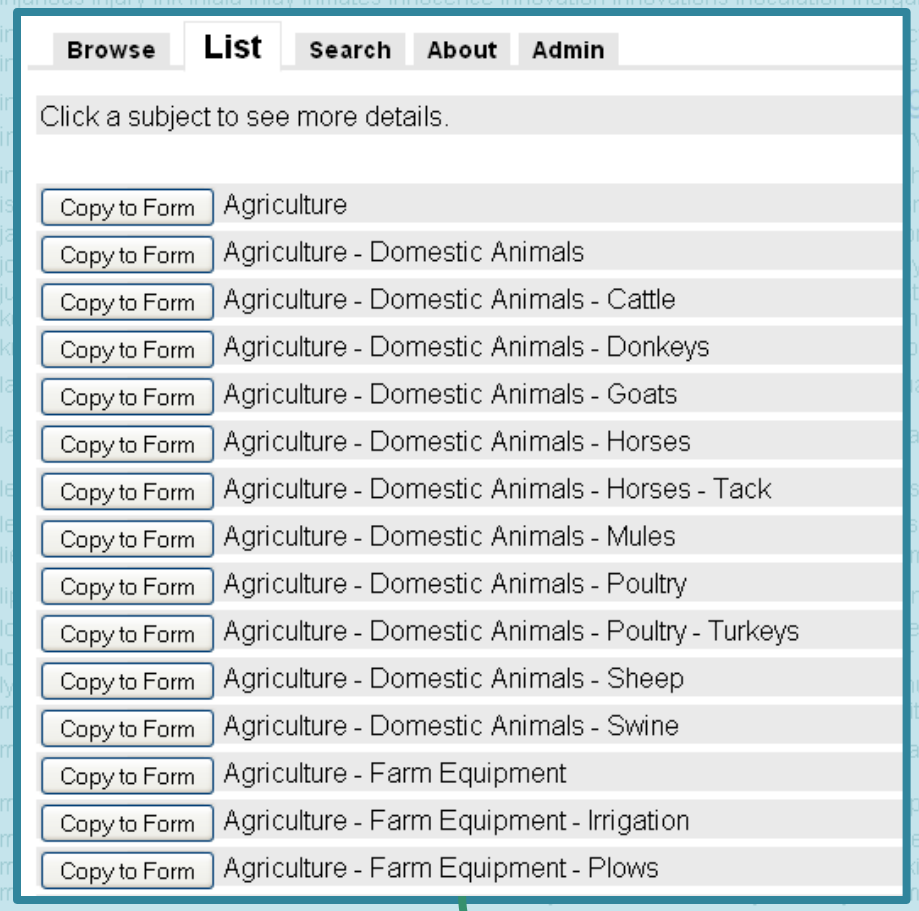
46828 Total Records

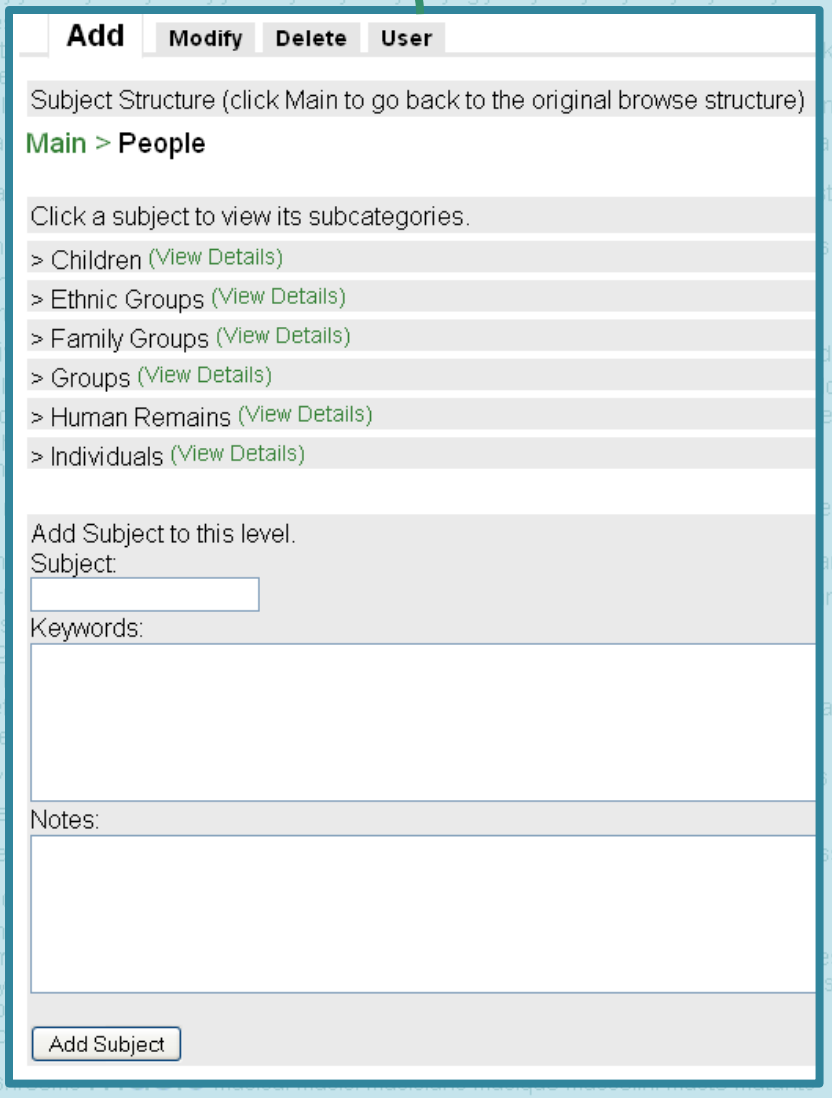
Metadata Analysis: NULL dc_title

details | thumbnails

Value:	Count:	
1	meta-pth-6111	edit
2	meta-pth-6577	edit
3	meta-pth-6580	edit
4	meta-pth-6581	edit
5	meta-pth-6582	edit
6	meta-pth-9528	edit
7	meta-pth-9546	edit
8	meta-pth-9547	edit
9	meta-pth-9617	edit
10	meta-pth-14408	edit
11	meta-pth-14409	edit
12	meta-pth-14410	edit
13	meta-pth-14411	edit
14	meta-pth-14412	edit
15	meta-pth-14413	edit
16	meta-pth-14414	edit
17	meta-pth-14415	edit
18	meta-pth-15485	edit
19	meta-pth-39258	edit
20	meta-pth-39259	edit
21	meta-pth-39260	edit
22	meta-pth-39261	edit

Null Value Analysis Tools





Controlled Vocabularies (UNTL-BS)

Explore by Type

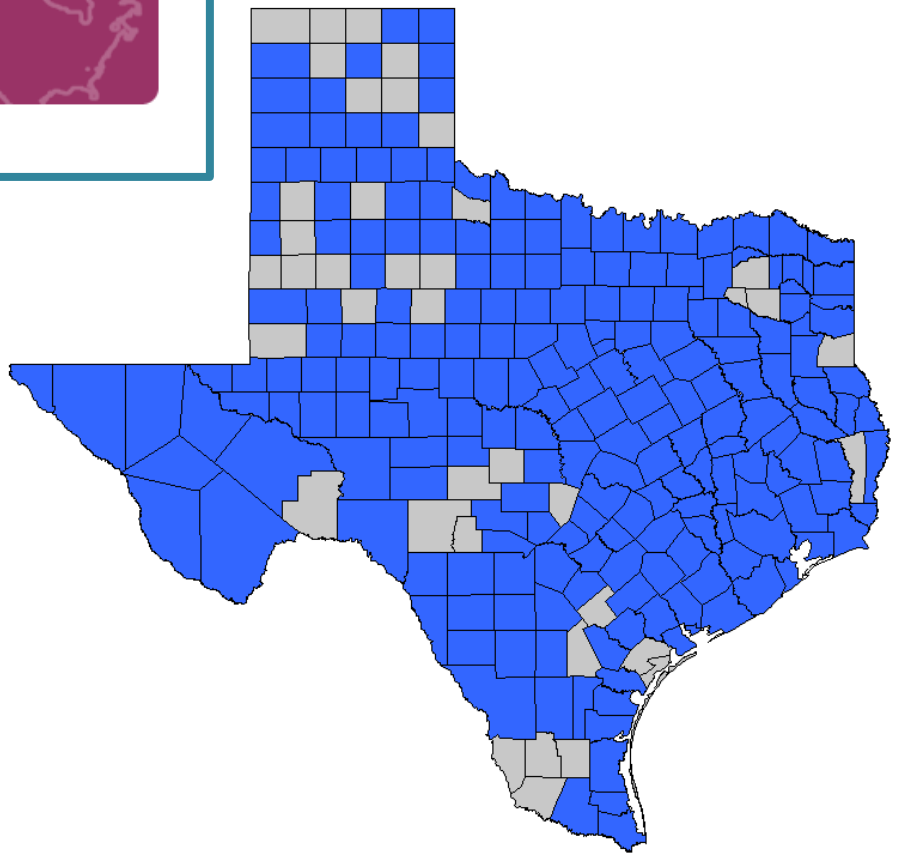
Artwork

Books

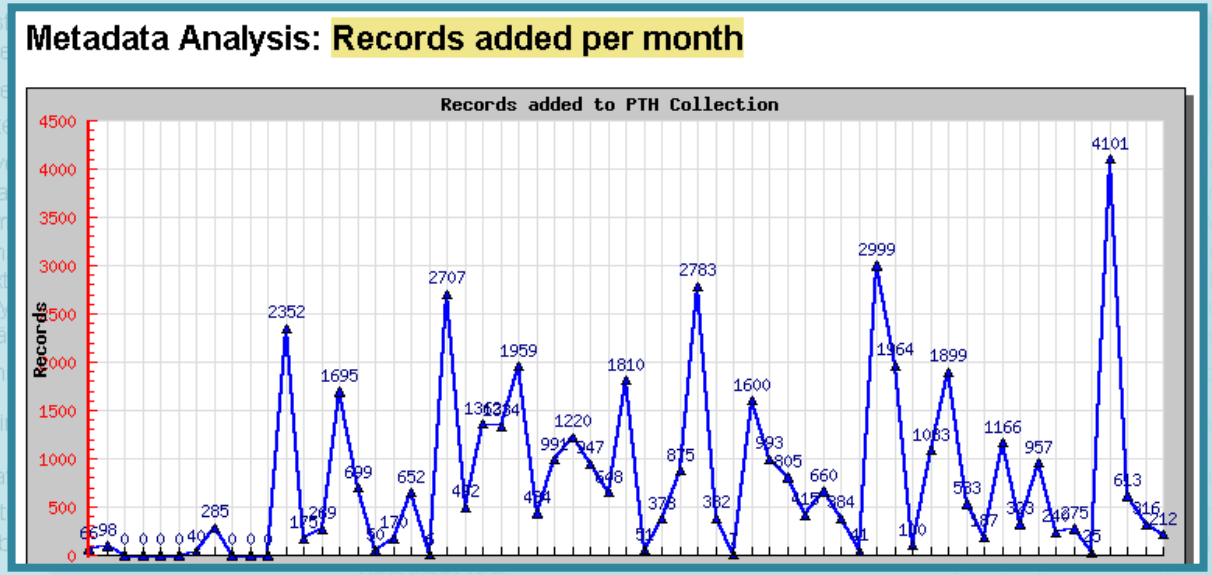
Letters

Maps

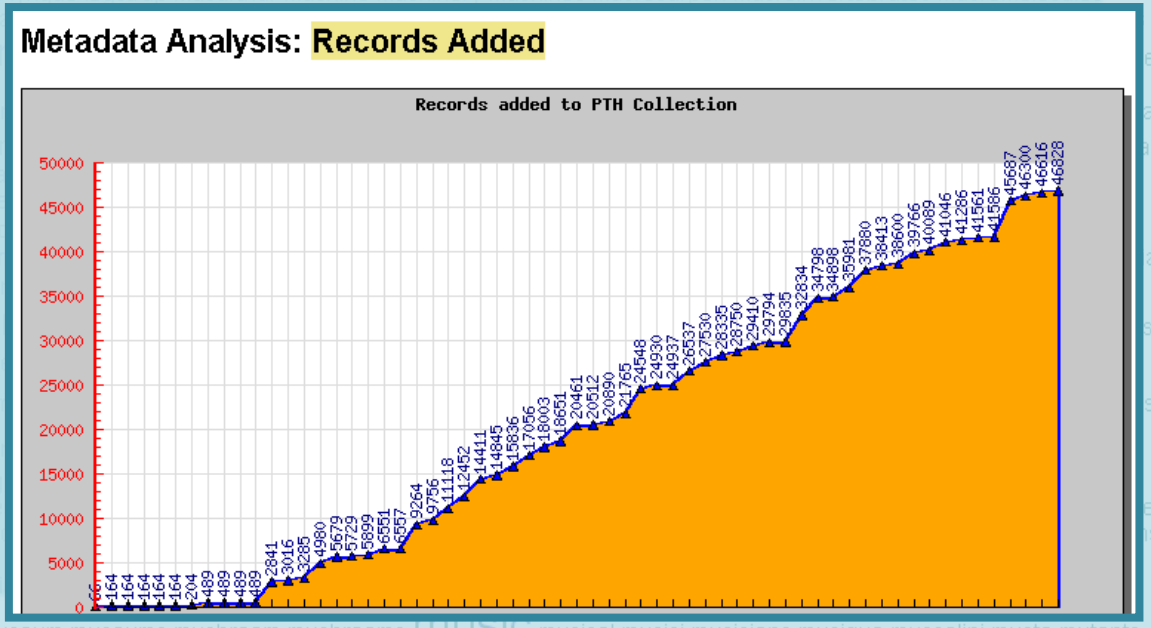
Better Metadata



More Functionality



Better Metadata

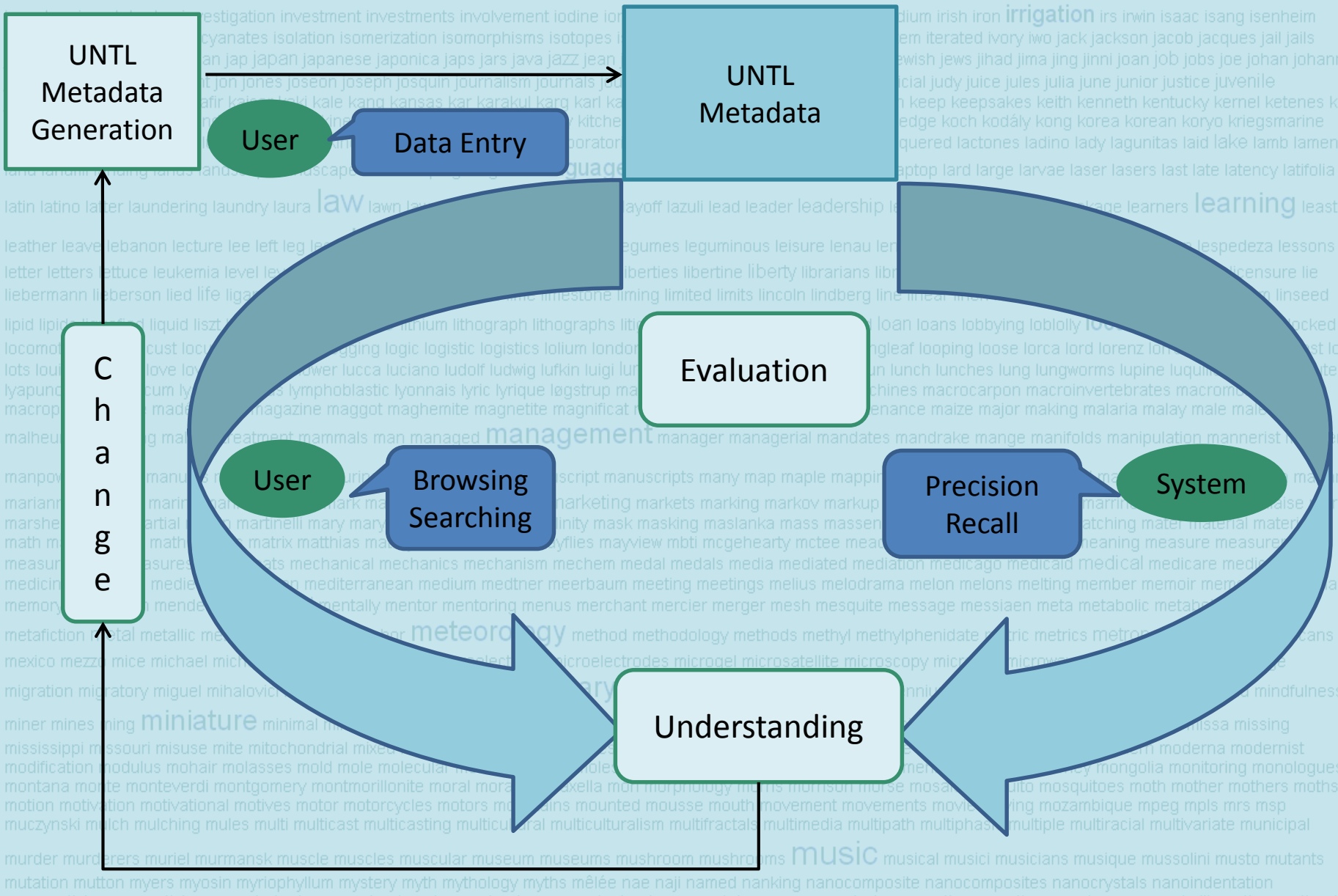


More Functionality

Summary

- Determine level of quality required
- Determine nature of gap and how to close it
- Machine verses human error handling
- Compromise
- Prioritize
- Test the workflow

Measure Quality and Usefulness of UNT Metadata



References & Web Sites Consulted

- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13(5), 407-424.
- Netcraft (2009). April 2009 Web Server Survey. Retrieved May 19th, 2009 from http://news.netcraft.com/archives/web_server_survey.html
- OCLC (2007). Sharing, privacy and Trust in our Networked World. Retrieved May 19th, 2009 from http://www.oclc.org/reports/pdfs/sharing.pdf
- TechSmith, Co. (2008). "UX 2.0: Any User, Any Time, Any Channel." Retrieved May 19th, 2009 from: http://download.techsmith.com/morae/docs/UserExperience2_0.pdf
- UNT Libraries Metadata Initiative page. Retrieved May 19th, 2009 from: http://www.library.unt.edu/digitalprojects/metadata

Questions?