

# **Metadata Quality Assurance: The University of North Texas Libraries Experience**

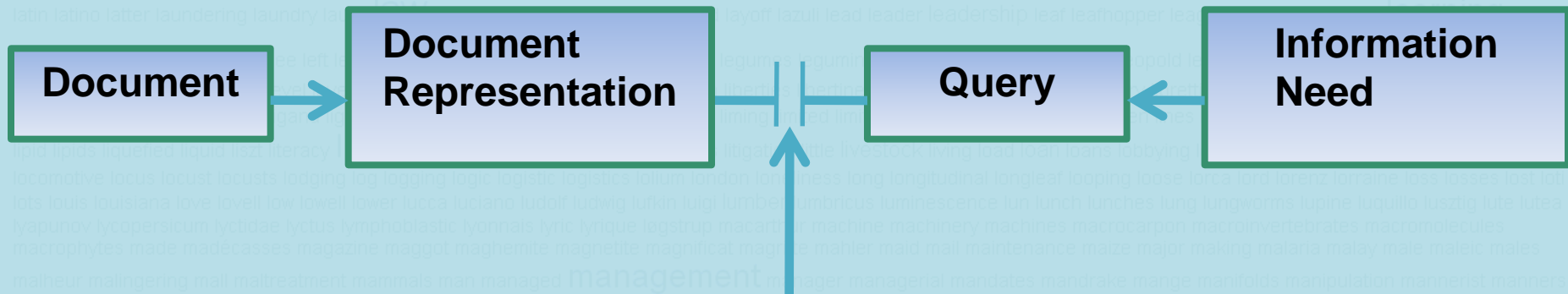
---

**Daniel Gelaw Alemneh & Hannah Tarver**

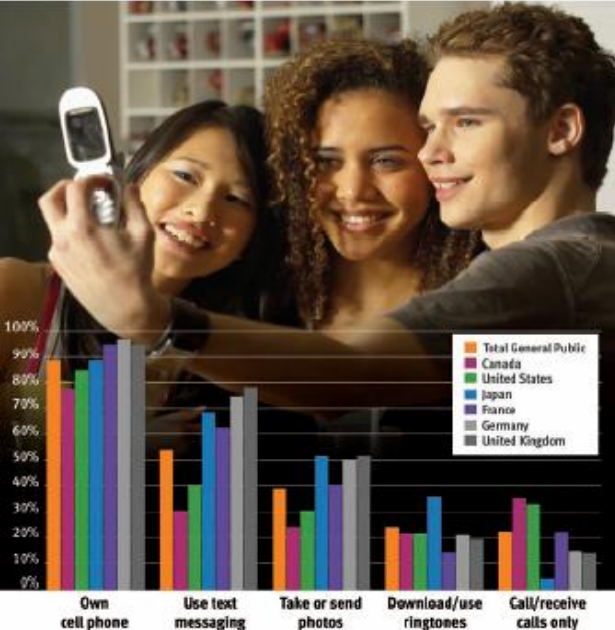
3rd annual Texas Conference on  
Digital Libraries (TCDL)

May 27-28, 2009

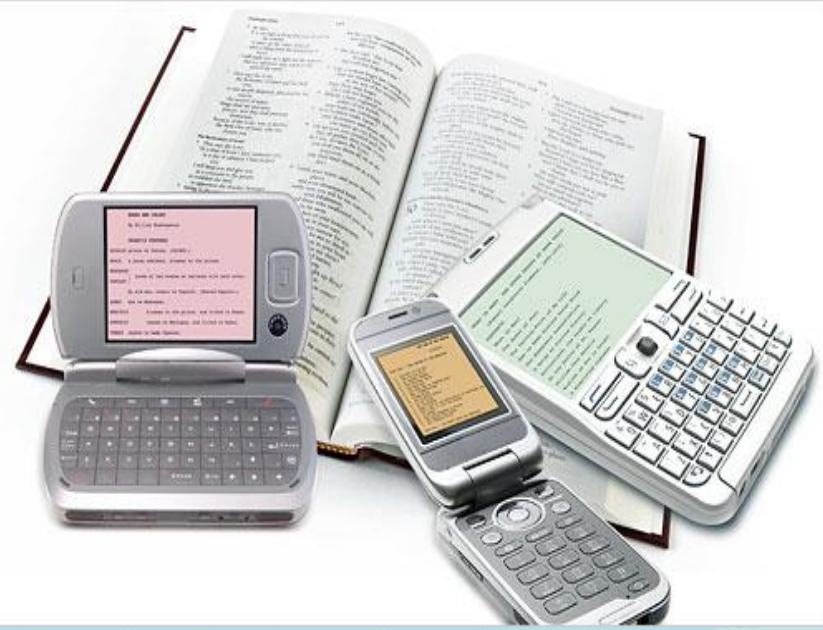
# Information Retrieval



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

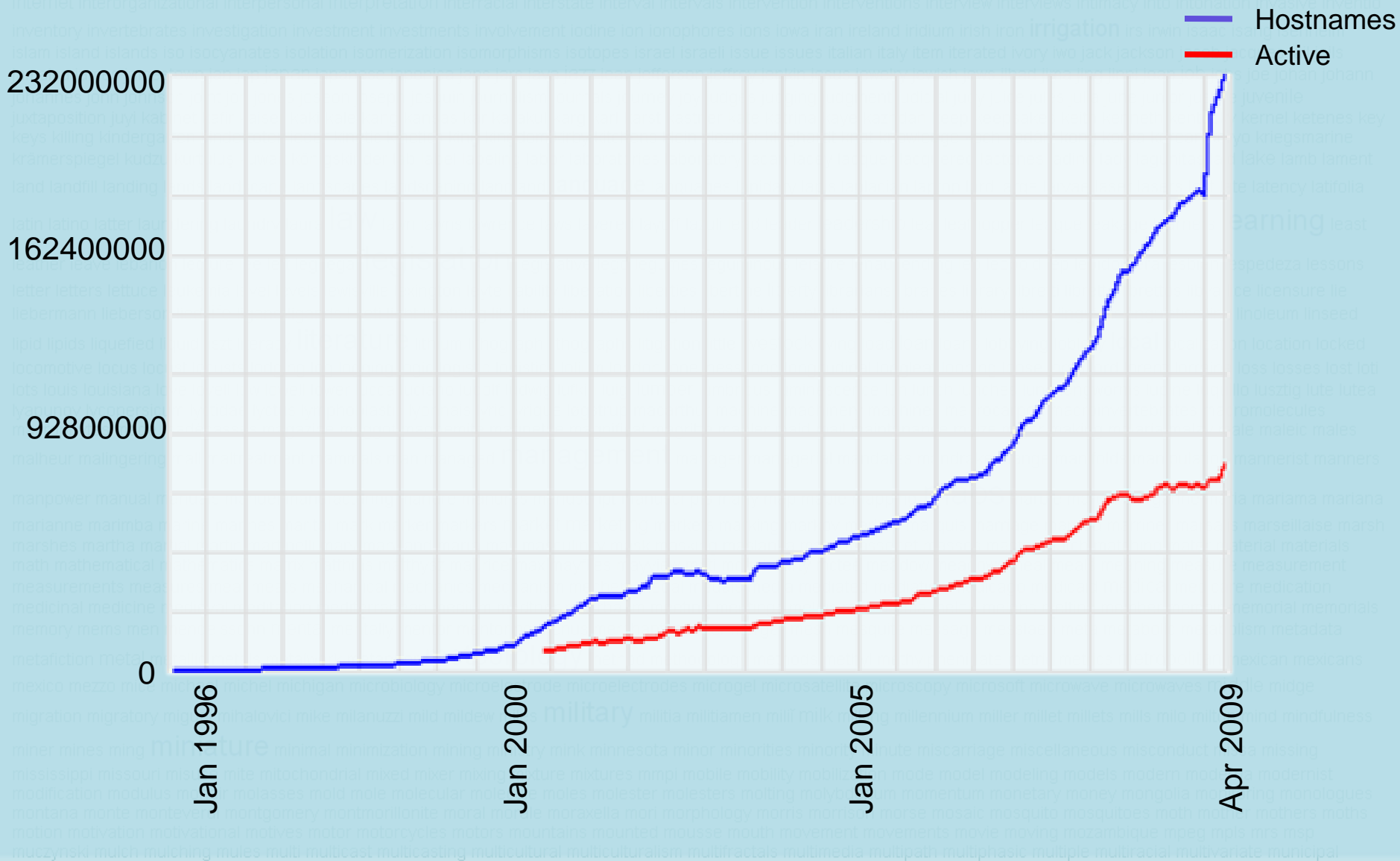


# Trends



- 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





# Digital Projects

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

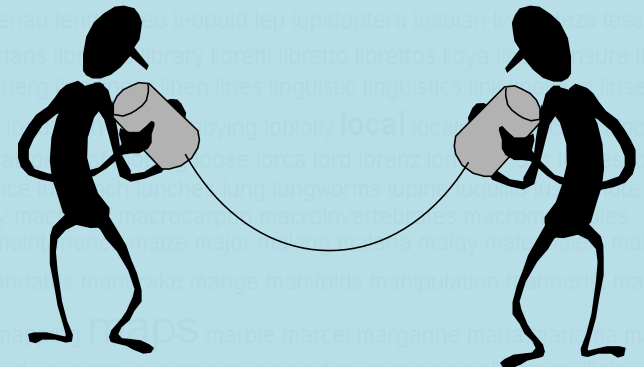
# Factors Influencing Metadata Quality

## 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

Post-Ingest

Training

Creation Tools

Proofing & Editing Tools

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

Creator

# Template Validation

http://digitalprojects.library.unt.edu



4 error(s) detected. Your form has not been submitted yet. Please check the information you provided.

OK

## 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 - template.xml, template.xml
File Edit Search Markers Folding View Utilities Macros Plugins Help
[Icons]
template.xml
21 <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.</description>.
22 <physicalDescription>x, 138 p. : ill., ports. ; 20
    cm.</physicalDescription>.
23 <dateCreated>1912</dateCreated>.
24 <dateDigitizes/>.
25 <publisher/>.
26 <contributor/>.
27 <source/>.
28 <language>eng</language>.
29 <coverage>.
30 <place-name>United States - Texas</place-name>.
31 <timePeriod>new-sou</timePeriod>.
32 <date/>.
33 <dateRange>.
34 <startDate>1894-02</startDate>.
35 <endDate>1893-07</endDate>.
36 </dateRange>.
37 </coverage>.
38 <resourceType>text_book</resourceType>.
39 <relation/>.
40 <format>text</format>.
41 <systemInfo>.
42 <software>.
43 <creationApp />.
44 </software>.
```

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
<a href="#">aut: Hare, Maud Cuney</a>	1
<a href="#">aut: Hartley, Oliver Cromwell, 1823-1859</a>	5
<a href="#">aut: Hartley, Oliver Cromwell, 1823-1859</a>	13
<a href="#">aut: Hartley, R. K.</a>	12
<a href="#">aut: Hartnett, Minerva</a>	1
<a href="#">aut: Hawker, H. W.</a>	2
<a href="#">aut: Hawker, H. W. (Herman William)</a>	12
<a href="#">aut: Hayes, F. A.</a>	1
<a href="#">aut: Hays, Hugh</a>	1
<a href="#">aut: Hazzard, J. L.</a>	1
<a href="#">aut: Hearn, W. Edward (Williamson Edward)</a>	3
<a href="#">aut: Henderson, J. P.</a>	1
<a href="#">aut: Hendrickson, B. H. (Bertram Higbie)</a>	4
<a href="#">aut: Hendrix, R.</a>	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
<a href="#">DC Title</a>	22
<a href="#">DC Creator</a>	26390
<a href="#">DC Subject</a>	1516
<a href="#">DC Description</a>	0
<a href="#">DC Contributor</a>	39967
<a href="#">DC Publisher</a>	41885
<a href="#">DC Identifier</a>	4942
<a href="#">DC Source</a>	43613
<a href="#">DC Relation</a>	45914
<a href="#">DC Language</a>	271
<a href="#">DC Coverage</a>	1241
<a href="#">DC Date</a>	2125
<a href="#">DC Resource Type</a>	287
<a href="#">DC Format</a>	194
<a href="#">DC Rights</a>	16171
<a href="#">UNTL Degree</a>	46826

46828 Total Records

# Null Value Analysis Tools

## Metadata Analysis: NULL dc\_title

details | [thumbnails](#)

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

[Browse](#)

**List**

[Search](#)

[About](#)

[Admin](#)

Click a subject to see more details.

- [Copy to Form](#) Agriculture
- [Copy to Form](#) Agriculture - Domestic Animals
- [Copy to Form](#) Agriculture - Domestic Animals - Cattle
- [Copy to Form](#) Agriculture - Domestic Animals - Donkeys
- [Copy to Form](#) Agriculture - Domestic Animals - Goats
- [Copy to Form](#) Agriculture - Domestic Animals - Horses
- [Copy to Form](#) Agriculture - Domestic Animals - Horses - Tack
- [Copy to Form](#) Agriculture - Domestic Animals - Mules
- [Copy to Form](#) Agriculture - Domestic Animals - Poultry
- [Copy to Form](#) Agriculture - Domestic Animals - Poultry - Turkeys
- [Copy to Form](#) Agriculture - Domestic Animals - Sheep
- [Copy to Form](#) Agriculture - Domestic Animals - Swine
- [Copy to Form](#) Agriculture - Farm Equipment
- [Copy to Form](#) Agriculture - Farm Equipment - Irrigation
- [Copy to Form](#) Agriculture - Farm Equipment - Plows

**Add**

[Modify](#)

[Delete](#)

[User](#)

Subject Structure (click Main to go back to the original browse structure)

**Main > People**

Click a subject to view its subcategories.

> Children ([View Details](#))

> Ethnic Groups ([View Details](#))

> Family Groups ([View Details](#))

> Groups ([View Details](#))

> Human Remains ([View Details](#))

> Individuals ([View Details](#))

Add Subject to this level.

Subject:

Keywords:

Notes:

[Add Subject](#)

**Controlled Vocabularies  
(UNTL-BS)**

## Explore by Type

Artwork



Books



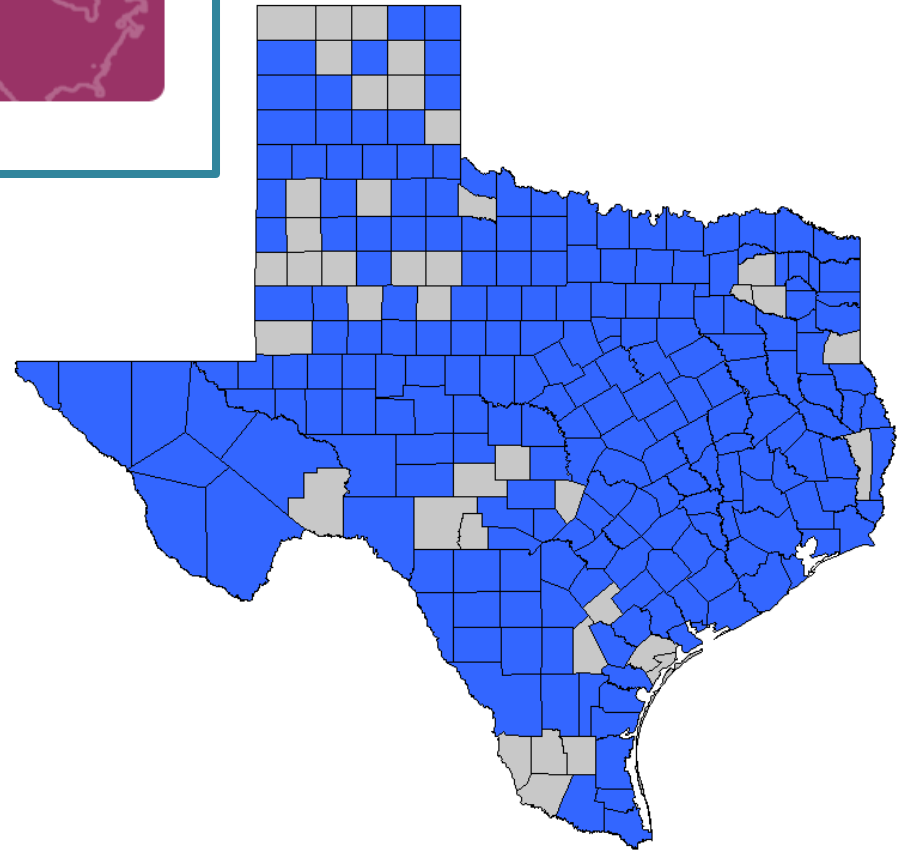
Letters



Maps



# Better Metadata



# More Functionality

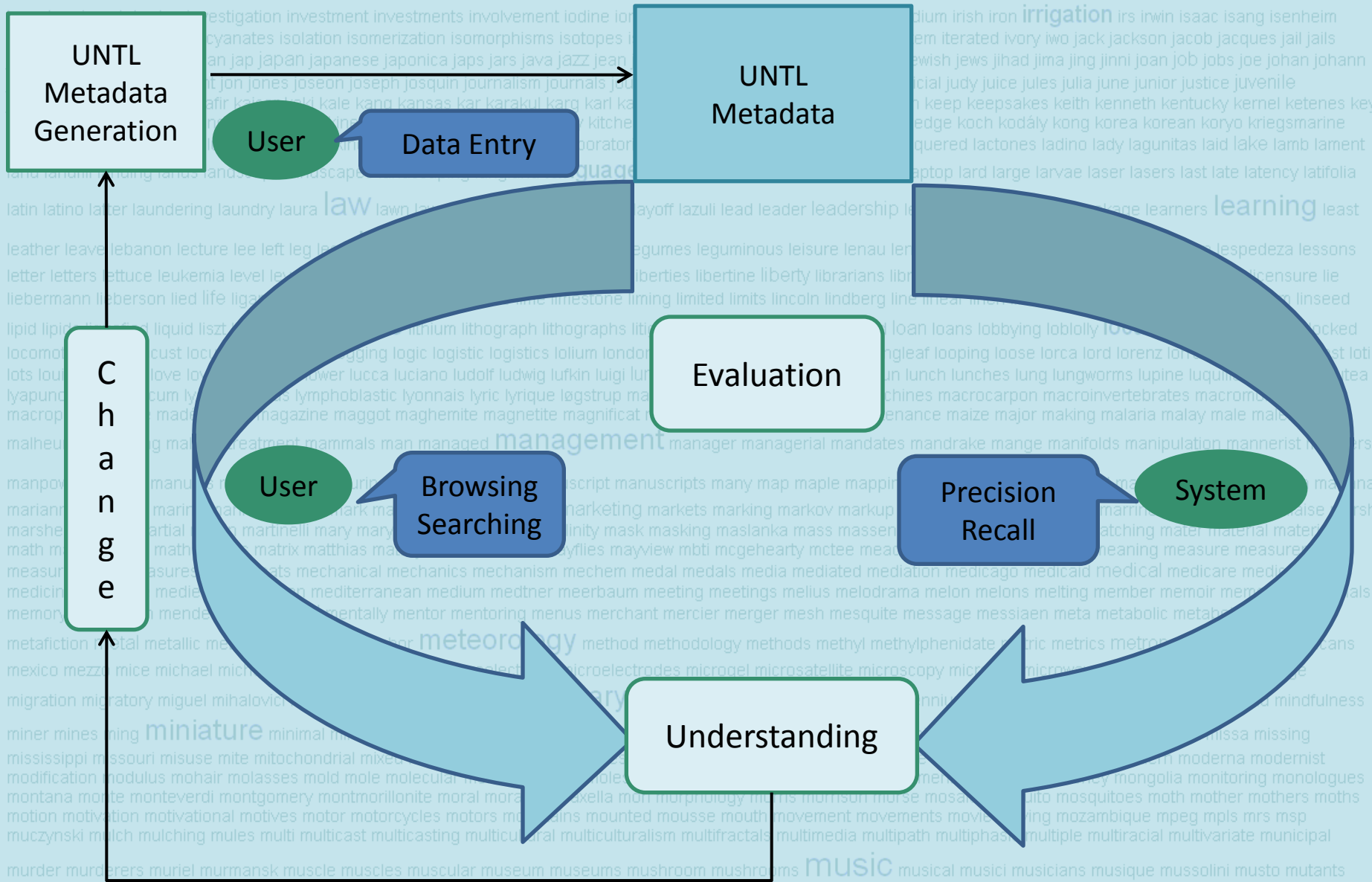


# Summary

- **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 19<sup>th</sup>, 2009 from [http://news.netcraft.com/archives/web\\_server\\_survey.html](http://news.netcraft.com/archives/web_server_survey.html)
- OCLC (2007). Sharing, privacy and Trust in our Networked World. Retrieved May 19<sup>th</sup>, 2009 from <http://www.oclc.org/reports/pdfs/sharing.pdf>
- TechSmith, Co. (2008). "UX 2.0: Any User, Any Time, Any Channel." Retrieved May 19<sup>th</sup>, 2009 from: [http://download.techsmith.com/morae/docs/UserExperience2\\_0.pdf](http://download.techsmith.com/morae/docs/UserExperience2_0.pdf)
- UNT Libraries Metadata Initiative page. Retrieved May 19<sup>th</sup>, 2009 from: <http://www.library.unt.edu/digitalprojects/metadata>

