# Integrating Controlled Vocabularies into Cultural Heritage Digital Collections: The Portal to Texas History Experience

**Daniel Gelaw Alemneh, Mark Edward Phillips, and Dreanna Belden**

UNIVERSITY OF NORTH TEXAS LIBRARIES, DIGITAL PROJECTS UNIT, DENTON, TEXAS

## Introduction

The Portal to Texas History provides a metadata framework that fosters a collaborative environment for more than 60 participating institutions. The Portal collaborators share a number of goals, including ensuring long-term and easy access to a wide variety of cultural heritage collections that embraces all geographic areas of Texas and covers prehistory through the twentieth century.

## The UNT Endeavor and Challenges

"Relevance" is subjective and situational. Although collaborating institutions have much in common, they may have conflicting requirements and expectations.

Participation in the wider community that can impose access restrictions and (formal or informal) requirements for interoperability: [Structure, semantics, and syntax.]

Maintaining usable & sustainable digital collections requires a complex set of actions. The UNT support various metadata-driven organization, navigation, and searching systems/mechanisms: -searching fields, -natural language vs. controlled vocabulary, -Integrate search and browse by dividing contents into well-defined groups/types: [alphabetical, chronological, geographical, institutional, topical, audience oriented]

All these enhanced and value added functionalities require high quality metadata. The vocabularies and taxonomies used to describe heterogeneous digital resources within metadata records does not rely solely on information contained within the resource itself.

### Diverse Collaborators

- Libraries
- Museums
- Historical Societies
- Individuals

### Heterogeneous Collections

- Scholarly Documents
- Museum Objects
- Archives
- Historical Documents
- Photographs

### Diverse Users

- Students
- Researchers
- Historians
- Geneologists

## The UNTL Controlled Vocabularies



As a controlled vocabulary, the UNT Libraries Browse-Subjects (UNTL-BS) draw different terms and concepts into one single word or phrase to enhance search and navigation. These vocabularies enable data enterers to easily select appropriate values and place them in metadata records. In addition to facilitating categorization for browsing, selecting a value from a controlled vocabulary ensures metadata consistency. Consequently, precision across all digital resources will be maintained. Considering the diversity of participating institutions and heterogeneity of the collections, all possible digital resources may not be described adequately using pre-determined or controlled terms. To overcome the limitations and balance the issues, the UNT Libraries are implementing a hybrid system that uses both controlled terms and free keywords in order to describe all possible resources adequately. This flexible approach of pre-defined and custom-generated vocabularies provides maximum flexibility to capture complete and high-quality metadata for all types of digital resources.

The use of both controlled vocabulary and natural language allows flexibility to meet the changing demands of sources as well as users.



## Summary

An integrated approach of using both structured and unstructured data provides enhanced browsing and searching capabilities and offers users better search results and a better experience. Although the debate of Controlled vocabularies versus free-text keywords continues, the UNT's Portal to Texas History experiences strengthen the assertion that (despite the rise of folksonomy) it is still ideal to use both systems in conjunction.