Terminologies Services Strawman

Background

This document was drafted for discussion for a meeting at the Metropolitan Museum of Art on September 12, 2007. ***This document was not intended to represent a work agenda, only to provoke discussion to help RLG Programs and OCLC Research determine the most promising and productive ways forward.*** The document was distributed to participants in advance of the meeting with instruction to share and discuss it with three peers in order to rank various proposed services in order of importance. Meeting participants reported on their conversations during the event, and, informed by the feedback of their peers, ranked the components of a terminologies service outlined in the strawman.

The goal of the meeting: to determine which components of a Terminologies Service the participants (representing a wide range of different communities and roles) find the most compelling. Desired meeting outcomes included a list of the most important features and use-cases listed below to prioritize efforts of RLG Programs and OCLC Research in this area of work.

To contextualize the focus of the meeting, it may be helpful to place our effort within the larger context of Terminologies Services related issues. Some of these issues are foundational to any of the work under discussion in the strawman; some of this work may be derivative of the work on particular service components. The particular conversations we acknowledged, but ruled out-of-scope for our particular meeting:

- Standards and best practices for technologies which allow terminologies providers to raise their terminologies to the network level.
- Support of terminologies services for cataloguing in local tools for cataloguing, discovery and access.
- Research on how terminologies services can be exploited most efficiently to meet our user's expectations.
- Business models for sustaining the terminologies themselves and the terminologies services.

For outcomes from the meeting, please consult the summary minutes report available at http://www.oclc.org/programs/events/2007-09-12.htm.

Terminologies Strawman

Prepared by:
Merrilee Proffitt, Günter Waibel (RLG Programs)
Diane Vizine-Goetz, Andrew Houghton (OCLC Research)

For discussion at:
Terminologies Meeting
Metropolitan Museum of Art, NYC
September 12th 2007

Note: The first use of a term defined in the glossary is marked by italics.

High-level characterization

Different communities and descriptive strategies share a common need to unambiguously identify a place or a person and to provide access points through subject terms or keywords. Activities surrounding the use of *terminology* resources could be raised to the *network level* through a series of services that support a range of activities including metadata creation, search formulation and optimization, and management of terminology resources as "local authorities" on the network. These *network services* can leverage the combined expertise and investment in description across libraries, archives, museums and visual resources to produce more authoritative records in a less expensive fashion. Terminologies services become the powerful building-blocks for entire records, especially if they provide access to *authority records*. When describing rare and unique materials, terminologies services could provide a measure of copy-cataloging economy by providing ready access to authoritative chunks of records.

Facets of Terminologies Services:

- Network services that support metadata creation activities, e.g., selection of terms for describing a resource (see 1. below)
- Network services that support search optimization for resource discovery, e.g. query expansion. (see 2. below)
- Network services that support terminologies management and sharing, e.g. creating and sharing a *local terminology* (see 3. below)
- Network services that support terminologies social interactions, e.g. *user* supplied variant terms (see 4. below)

Once different terminologies sit in the same network space, the creation of relationships between them could further enhance the power of the services offered, or spawn entirely new services (see 5. below).

For Prioritization – Components of Terminologies Services

1. Services that support metadata creation

    a. Browsing, searching and retrieving terms
       Use case: A cataloger is looking for a genre term associated with a resource. He types a term into the search box of a cataloging tool and retrieves a list of terms from one or more terminologies. He selects the appropriate term and moves on to another task.

    b. Displaying part of an authority record
       Use Case: A museum registrar chooses a term for a descriptive record. She has the option to dynamically display or embed the additional information contained in the authority record in the local collections management system. For example, a personal name could be accompanied by supplementary information such as nationality, gender, birth and death dates, place of birth, etc.

2. Services that support search optimization

    a. Browsing, searching, retrieving and displaying the authority record
       Use case: In an end user environment, the term itself as well as its contextual information from the authority record can be displayed and searched. For example, information about an artist can be retrieved by artist name, or the user can browsed for an artist after querying approximate life dates from the name authority.

    b. Leveraging terminology for search optimization
       Use case: A user formulates a query for the place name "Augsburg". The query is expanded to include equivalent terms, e.g., "Augusta Vindelicorum" (the original name of the Roman settlement).
       Use case: A user formulates a query for the place name "Bavaria" (state). The query is expanded to include terms in the same hierarchy, e.g., "Schwaben" (a district within Bavaria).

3. Services that support terminologies management and sharing

    a. Managing local terminologies
       Use case: An institution uses a set of local genre terms for assignment to resources. The genre terms are managed in a shared Excel spreadsheet. The institution elects to migrate from the Excel spreadsheet to the use of a terminologies service to manage genre terms and consolidate future terminologies.

    b. Sharing local terminologies
      Use case: A historical society creates a list of local place names not found in published sources and shares them with other institutions.

    c. Combining local, shared or published terminologies
      Use case: A museum wants to use a set of local personal names in conjunction with names drawn from a *published terminology*, such as ULAN. The institution creates a combined terminology composed of the locally generated terms and the published terminology.

    d. Establishing a project-specific subset of terms
      Use case: In a library, a professional selects a list of descriptive terms from published and local terminologies for use by paraprofessionals and interns to describe resources for a specific project.

    e. Contributing to a published terminology
      Use case: An institution has created a local set of terms that are extensions to a published terminology. The institution submits the terms for review and authorization so the terms can be incorporated into the published terminology.

4. Services that support terminologies social interactions

    a. Joint editing and annotation of local terminologies by *experts*
      Use case: A digital library has established a list of local place names that a museum would like to use for a project cataloging local artifacts. In the course of the project, the museum discovers that it needs local place names not represented on the existing list. In addition, some of the local place name metadata is missing or not sufficiently detailed. The library and museum agree to collaboratively update and annotate the list of local place names.

    b. Capturing locally contributed end-user terminology
      Use case: A researcher uses a finding aid at an archive to locate a collection of materials. While reading through the documents in the collection the researcher realizes that the finding aid does not list persons and places associated with the documents. The researcher updates and annotates the finding aid to include the missing information.

5. Value-added intelligence: creating relationships between terminologies

    a. Generating terminology mappings using data mining techniques
      Use case: The system generates co-occurrence mappings between published terminologies.
      Use case: The system generates co-occurrence mappings between end-user terminology and published terminologies.

b. Generating terminology mappings using human intelligence
Use case: A professional creates mappings between published terminologies.
Use case: A professional creates mappings between end-user terminology and published terminologies.

For Prioritization – Categories of Terminologies

- Form (Genre/Class)
- Object types
- Organizational names
- Personal names
- Place names
- Publisher names
- Relator terms
- Time periods
- Topical subjects

Glossary

- *Authority record* – a collection of information about a name, subject, or title, used as an access point, including the established form, references for associated names or terms, and other information as needed.
- *Expert* – a professional creator of metadata, usually employed by a library, archive, museum or visual resources center.
- *Local terminology* – a terminology originally prepared for local distribution or access.
- *Network level* - the space that is exposed to all participants accessing the World Wide Web.
- *Network service* - any system that provides one or more functions that are accessible over the World Wide Web.
- *Published terminology* – a terminology prepared for public distribution or access (usually includes mechanisms for contribution as well as editorial oversight).
- *Terminology* – a list or vocabulary of terms (words or phrases) or notation used to describe, navigate, and search content.
- *User* - someone who uses a system to access materials such as a researcher, teacher or student.