Creating an OLAC data provider at the Alaska Native Language Center

Gary Holton
ANLC

LSA Symposium:
The Open Language Archives Community
4 January 2002

ANLC Archives

- 10,000+ paper documents
- 5000+ recordings
- Nearly everything written in or about Alaska Native and languages
- Existing catalog uses proprietary metadata format

Data Provider responds to metadata harvest requests

Harvest request is a URL

- A request to the data provider has the form of a URL which specifies a script and one of six request "verbs"
  http://archives.anlc.uaf.edu/anlc.php?verb=VERB
  
  base URL
  script name
  verb

Harvest request verbs:

- Identify
- ListMetadataFormats
- ListSets
- ListIdentifiers
- ListRecords
- GetRecord

Response to "Identify" request

http://archives.anlc.uaf.edu/olac/anlc.php?verb=Identify
Response to "GetRecord" request


Implementation

- Use scripting language (php, perl, etc.) to describe response to request verbs
- Identify, ListMetadataFormats, ListSets have standard (fixed) responses
- ListIdentifiers requires simple query of catalogue identifiers in database
- GetRecord requires mapping from local metadata to OLAC metadata format

Original ANLC catalog record

Krauss, Michael S.
Notes on Gwich’in orthography. 2 pp.
1975a
Ms. notes on particular points of possible revision and questions on Gwich’in orthography.

ANLC metadata format

Catalog: KU962K1975a
Author: Krauss, Michael
Title: Notes on Gwich’in orthography
Description: Ms. notes on particular points of possible revision and questions on Gwich’in orthography.
Comments: 2 pp.
Keywords: orthography
URL: http://archives.anlc.uaf.edu/docs/KU962K1975a.pdf

Mapping ANLC to OLAC

- one-to-one mappings
- many-to-one mappings
  - using "refine" attribute
  - collapsing elements
- one-to-many mappings

One-to-one mappings

- some examples
  - Catalog number → <identifier>
  - Author → <creator>
  - Language → <subject.language>
  - Keywords → <type.linguistic>
Many-to-one mappings using refinements

- Interviewer → <creator refine="interviewer"/>
- Editor → <creator refine="editor"/>

Title → <title>Tsaan'suug</title>
English title → <title refine="alternate" lang="en">Ruffed Grouse</title>

Many-to-one mappings without using refinements

- some elements collapse to a single OLAC element

One-to-many mappings

- ANLC catalog number may also encode language and date information
- Although this violates database design principles, we can provide legacy support
- Example: KU962K1975a
  <subject language code="X-SIL-KUC"/>
  <date code="1975"/>

Communicating with the Service Provider

- OLAC protocols provide the means for the Service Provider to communicate with the Data Provider
- Service Provider harvests periodically
- <datestamp> allows harvester to determine which records need to be harvested
- Service Provider then stores information in any desired format

Advantages for Data Provider

- can maintain legacy databases
- no need to provide search facilities and other value-added products
- light load on web-server, since the server only responds to periodic harvest requests

...the Service Provider