Getting Involved in OLAC

Steven Bird
University of Pennsylvania

LSA Symposium:
The Open Language Archives Community
4 January 2002

How can I get involved?

1. As a resource user:
   - Locate useful data, tools and advice

2. As a resource creator:
   - Contribute metadata so your resources can be found, used, cited

3. As a developer of standards and best practices:
   - Help to refine the OLAC infrastructure

1. As a resource user:

   Use the OLAC search engine:
   - http://www.linguistlist.org/olac/

   Join OLAC-General for news updates:
   - Moderated, ~1 message per month
   - http://www.language-archives.org/

2. As a resource creator:

   Three ways to contribute metadata:
   A. Conventional Data Providers
   B. Vida – the Virtual Data Provider
   C. ORE – the OLAC Repository Editor

A. Conventional Data Providers

Your website

Existing database SQL

HTTP: getRecord

OLAC data provider

OLAC harvester

XML document

Combined database

Credits…

The development of core OLAC infrastructure is funded by three NSF grants:
- ISLE: International Standards in Language Engineering (9910603)
- TalkBank: A Multimodal Database of Communicative Interaction (9978056, 9980009)
- E-MELD: Electronic Metastructure for Endangered Languages Data (0094934)

We gratefully acknowledge the support of the Open Archives Initiative (www.openarchives.org)
A. Conventional Data Providers

What you need:
- An existing catalog in a database
- Permission to install scripts on a web server
- Access to a programmer

But its not too difficult...
- Open source implementations exist
- Written in several programming languages

B. Vida – the Virtual Data Provider

What you need:
- An existing catalog in a database
- Permission to install scripts on a web server
- Access to a programmer

But its not too difficult...
- Open source implementations exist
- Written in several programming languages

B. Vida – the Virtual Data Provider

What you need:
- An XML editor
- if you have no pre-existing catalog
- OR: a programmer
- who can convert your existing data into XML
- Access to a web site
- simply to upload the single XML file

C. ORE – OLAC Repository Editor

What you need:
- A web browser

Demonstration:
- http://wave.ldc.upenn.edu/language-archives/ORE

3. As a developer of standards and best practices:

- The OLAC Process
  - A document which describes how OLAC is organized, and how it operates
- OLAC Documents
  - 3 types: Standard, Recommendation, Note
  - 6 status levels: Draft, Proposed, Candidate, Adopted, Retired, Withdrawn
- OLAC Working Groups
  - open, self-organizing, develop OLAC documents
Document Types

- Standard
  - procedures that participating archives and services must follow
- Recommendation
  - OLAC consensus on best current practice for some aspect of language-resource archiving
- Note
  - Implementation details

Document Status Levels

- Draft: under development by a working group
- Proposed: posted for open peer review
- Candidate: approved, undergoing implementation and testing before full adoption
- Adopted: approved following an adequate period of experience with implementation
- Retired: community decides the document is no longer relevant
- Withdrawn: removed from the process before reaching adopted status

Working Groups

- The primary source of documents that enter the OLAC document process
- Any member of the community can create or participate in a working group
- Working group members represent at least three different institutions
- First working group: language codes

Open Language Archives Community

An international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by:

- developing consensus on best current practice for the digital archiving of language resources
- developing a network of interoperating repositories and services for housing and accessing such resources

OLAC Works...

- Built on proven standards from digital libraries
- Already has 13 participating archives
  - US, UK, France, Germany, Netherlands
  - Total of 18,000 metadata records
- Cross-archive search on LINGUIST site
- Low barrier for new archives
- Three methods: Conventional, Vida, ORE

More archives plan to join (12)

- Aboriginal Studies Electronic Data Archive
- Academia Sinica
- Archive of the Indigenous Languages of Latin America
- Child Language Data Exchange System
- Deutsches Spracharchiv
- Institute National de la Langue Française
- Language and Culture Atlas of Ashkenazic Jewry
- Max Planck Institute
- National Anthropological Archives
- Oriental Institute
- Rosetta Project
- Tibetan and Himalayan Digital Library
OLAC Phases

1. Development phase (2001)
   - Built the infrastructure (software, standards)
   - 13 alpha testers had a moving target
   - Freeze the standards to encourage adoption
   - Review and refine standards (late 2002)
3. Operational phase (2003-)
   - Best practices for digital content

OLAC: An Unprecedented Opportunity

Language documentation and description
- Creation of digital resources is skyrocketing
- Web will be the main dissemination method
- People want to discover reusable resources

Two possible futures:
- Unparalleled frustration and confusion
- Unparalleled access to information

Act in community to define best practice...