Getting Involved in OLAC

Steven Bird
University of Pennsylvania

LREC Symposium:
The Open Language Archives Community
29 May 2002

Credits…

Core OLAC infrastructure is funded by NSF grants:
- **ISLE**: International Standards in Language Engineering
- **TalkBank**: A Multimodal Database of Communicative Interaction
- **E-MELD**: Electronic Metastructure for Endangered Languages Data

Software developers at the LDC: Eva Banik & Alan Lee
We gratefully acknowledge the support of the Open Archives Initiative

OLAC Launch, LREC-02

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

OLAC Launch, LREC-02

---

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
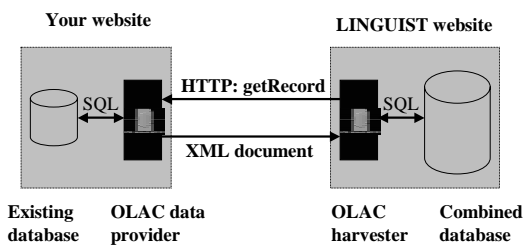- http://www.language-archives.org/

OLAC Launch, LREC-02

---

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

OLAC Launch, LREC-02

A. Conventional Data Providers

OLAC Launch, LREC-02
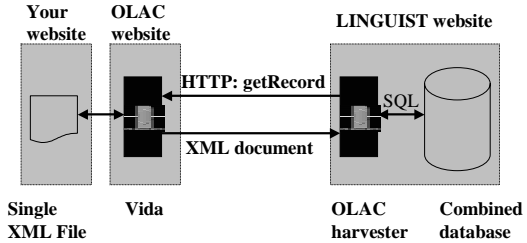
A. Conventional Data Providers

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

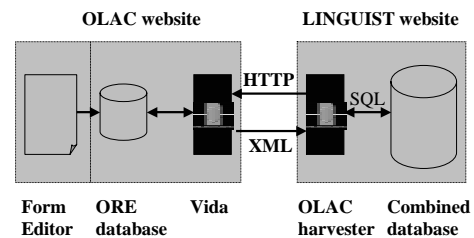| Your website	OLAC website		LINGUIST website
		HTTP: getRecord	SQL
		XML document	
Single XML File	Vida	OLAC harvester	Combined database

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

OLAC website			LINGUIST website
		HTTP	SQL
		XML	
Form Editor	ORE database	Vida	OLAC harvester / Combined database

---

2. As a resource creator - summary

A. Conventional Data Providers
- database, programmer
- web *server* (CGI processing)

B. Vida – the Virtual Data Provider
- dump database to XML, or use XML editor
- web *site* (XML file hosting)

C. ORE – the OLAC Repository Editor
- fill in forms
- web *browser* (access to online service)

---

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

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*

OLAC Phases

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

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
  - Dublin Core; Open Archives Initiative
- Already has 20 participating archives
  - France, Germany, Netherlands, UK, US
  - ~30,000 metadata records
  - Many more archives plan to join
- Cross-archive search on LINGUIST site
  - Anyone can set up a harvester and a service
- Low barrier for new archives
  - Three methods: Conventional, Vida, ORE

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…