The Seven Pillars of Open Language Archiving: Introducing the OLAC Vision

Gary Simons
SIL International

LREC Symposium:
The Open Language Archives Community
29 May 2002

Unprecedented opportunity

Digital archiving of language documentation and description on the World-Wide Web offers:

- Minimal cost multimedia publishing
- Maximal access by the citizens of the world

Two possible futures

Digital archiving could hold:

- the promise of unparalleled access to information,
  or,
- the specter of unparalleled frustra-tion and confusion

The outcome will depend on whether we:

- act in community to define and follow best common practice,
  or,
- act in isolation to proliferate idio-syncratic practices

A community to bridge the gap

- What users want—the ideal
- What users actually get —the gap
- What it would take to bridge the gap—a community that provides the infrastructure for acting in concert

---

A building metaphor

The infrastructure is erected on seven pillars:

- Data
- Tools
- Advice
- Gateway
- Metadata
- Review
- Standards

What users want

The individuals who use and create language documentation and description are looking for three things:

- Data
- Tools
- Advice

1. Data

- Information that documents or describes a language of interest
- A wide variety of formats: print publications, computer data files, sound recordings, hand-written index cards, and so on
- A wide variety of content: word lists, paradigms, texts, annotations, lexicons, grammar descriptions, and so on

2. Tools

- Computational resources that facilitate creating, viewing, querying, or otherwise using language data
- These include: application programs, components, fonts, style sheets, document type definitions, and so on
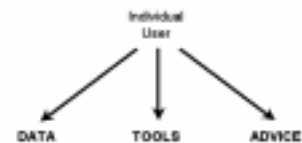
3. Advice

- Information that users would typically solicit when they need help
- For instance,
  - What data sources should I rely on?
  - What software tools should I use?
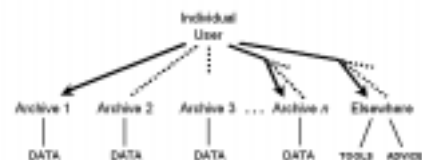 - What practices should I follow when creating data? When using data?

The ideal situation

What users actually get

- The data are archived at hundreds of sites
  - Some are on Web and user does find them
  - Some are on Web but user can't find them
  - Some are not even on Web
- The tools and advice are at hundreds of other sites

The gap

It's even worse

- The user may not find all existing data about the language of interest because different sites have called it by different names.
- The user may not be able to use an accessible data file for lack of being able to match it with the right tools.
- The user may locate advice that seems relevant but then has no way to judge how good it is.

OLAC Launch, LREC-02

What a community could provide

In order to bridge the gap, the individuals who use and create language documentation and description need a community that provides four things:

- A single gateway
- Uniform metadata
- A review process
- Standards

OLAC Launch, LREC-02

4. Gateway

- A single portal through which users gain access to all available data, tools, and advice
- The actual data, tools, and advice are located on hundreds of sites all over the Internet—the gateway stores links to them.
- By accessing the single gateway site, the user gains access to all available data, tools, and advice.

OLAC Launch, LREC-02

5. Metadata

- Uniform descriptions of all available data, tools, and advice
- Not the data itself, but data about the data; thus it works for digital and non-digital holdings.
- Uses specialized metadata elements to meet requirements specific to language archives:
  - uniformly identifying languages
  - matching data formats to the appropriate tools

OLAC Launch, LREC-02

6. Review

- Peer review is an important function of any academic community
- Peer evaluation of available data, tools, and advice
- Review by individuals (with responses)
- Review by community to establish some advice as recommended best practice

OLAC Launch, LREC-02

7. Standards

Provide the framework that allows the core infrastructure to function:

- Gateway—governed by a protocol for harvesting metadata from participating archives
- Metadata—governed by an XML schema that ensures uniformity across all archives
- Review—governed by a process that allows community to establish best practice

OLAC Launch, LREC-02

The community infrastructure

Open Language Archives Community

See Handout

- Mission statement
- Twenty participating archives and growing
- Three governing standards based on proven standards from the digital library community
- Three ways to get involved