An Overview of OLAC:
The Open Language Archives Community

Gary Simons and Steven Bird

Workshop on The Digitization of Language Data: The Need for Standards
21-24 June 2001

The vision: a gap analysis

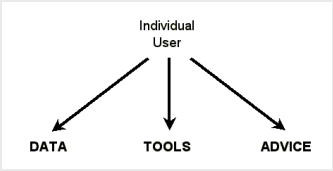
- What users want—the ideal
- What users actually get —the gap
- What it would take to bridge the gap—a community infrastructure

What users want

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

- Data
- Tools
- Advice

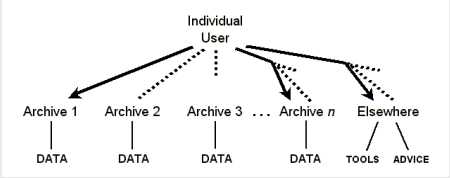
The ideal situation



What users actually get

- The data are archived at hundreds of sites
  - Some are on Web and user finds them
  - Some are on Web but user can't find them
  - Some are not even on Web
- Other problems
  - Sites use different names for same language
  - Tools and advice are on different sites than data
  - Don't know which tools go with which data

The gap

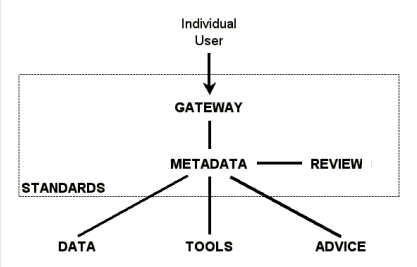


What a community could provide

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

Proposed community infrastructure



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

Participating Archives

- Alaska Native Language Center
- Comparative Bantu Online Dictionary
- Deutsche Forschungszentrum für Künstliche Intelligenz
- European Language Resources Association
- Langues et Civilisations à Tradition Orale
- Linguistic Data Consortium
- Perseus Project
- SIL International

Prospective Participants

- Aboriginal Studies Electronic Data Archive
- Academia Sinica
- American Indian Studies Research Institute
- American Philosophical Society American Indian Manuscript Collections
- Institut National de la Langue Française
- Language and Culture Atlas of Ashkenazic Jewry, EYDES
- The LINGUIST List
- Max Planck Institute
- National Anthropological Archives
- Oxford Text Archive
- Rosetta Project
- Tibetan and Himalayan Digital Library
- Typological Research Center

Going deeper

- Visit the Web site
  - http://www.language-archives.org
- The foundational standards
  - Open Archives Initiative metadata harvesting protocol
  - OLAC Metadata Set
  - OLAC Process
- Join the community by self-subscribing
  - http://lists.linguistlist.org/archives/olac-general.html