

User Services

Live Archives offer three core classes of user services: upload, management and access.

Upload

Digital archiving is an ongoing process. Depositors and users can upload additional resources or update existing ones. Data management systems will provide facilities for upload, while tracking versions and identifiers so that important information and references will not be lost.

Management

LRA's data management systems allow users to define access policies and rights, to carry out checks, to do conversions and downloads, and to view statistics about their data and the rate of access.

Access Services

LRA's will offer a number of access services:

- WWW catalogue browsing and searching via metadata or via geographic interfaces, to support resource discovery
- WWW access to resources (where permissions allow) via standard web browsers
- individual resources as well as sub-corpora, together with relevant metadata, should be downloadable to a variety of devices
- specialized applications to allow complex resources to be accessed via the WWW
- open-standard metadata (OAI-PMH, XML-Schema) provided for harvest by other bodies to expand resource discovery potential



Digital Language Resource Archives

Supporting Institutions

The intention is that as many institutions as possible that hold language resources will support the *Live* Archives principles and directions, as described in this leaflet. If your institution may be interested in supporting it, or finding out more information, please fill in the form on the web-site:

www.mpi.nl/dam-lr/lra-flyer

Currently, the following institutions have declared support for *Live* Archives (see the website for the latest information):

- MPI for Psycholinguistics, Nijmegen*
- Centre for Languages and Literature, University of Lund*
- TST-INL, Leiden*
- SOAS, University of London*
- Radboud University Nijmegen
- University Sheffield
- Charles University Prague
- Bulgarian Academy of Sciences, Sofia
- Berlin-Brandenburgische Academy of Sciences
- University of Bergen
- Istituto di Linguistica Computazionale, CNR, Pisa
- University of Antwerp
- Centre de Ressources sur la Description de l'Oral, CNRS, Paris
- University of Helsinki
- Institute for Language and Speech Processing, Athens
- University of Zagreb

Other institutes that declared their support for this document are listed on the web-site.

*The *Live* Archives statement was initiated by the DAM-LR partners (www.mpi.nl/dam-lr).



Digital Language Resource Archives

LIVE ARCHIVES

A checklist of principles and tasks

March 2006



Digital Language Resource Archives

Principles

Live Archives adhere to a number of principles. They will:

- implement strategies for long-term preservation in collaboration with other institutions
- adopt and promote open standards where possible
- use resource identifiers that can be universally and perpetually used to uniquely identify materials
- federate their catalogues so that authorized users can seamlessly browse and access their combined holdings

- distinguish between frequently-changing physical storage infrastructure and comparatively stable language resource organization
- provide data management systems that rigorously define procedures, policies, rights, and delegation mechanisms
- offer appropriate forms of resources and access according to the needs of recognized user groups
- support the repurposing and mobilization of resources in new and existing genres, for example for language revitalization
- implement systems that allow depositors to upload new or updated resources
- provide capability for users to add layers of interpretation, annotation and commentary

The OAIS reference model is an excellent description of the general architecture and tasks of digital archives, but the principles of Live Archives go beyond this model. LRAs adhering to the Live Archives principles can create Federations that maximize their capabilities for long-term preservation, the value of their resources, and their potential to serve users. Such Federations will be practical partnerships that share common aims, ethical frameworks, and technology for data exchange and access management.

Tasks

Preservation

LRAs undertake long-term preservation of resources, many of which are unique and irreplaceable, such as recordings of endangered languages, experimental data, and historical documents. Digital technologies provide the capability to repeatedly copy information without loss, so LRAs focus on preserving only information content, not physical carriers such as paper, tapes or disks. However, due to the complexity and changing nature of digital technologies, LRAs have to be actively involved with the forefront of these technologies.

Access

While traditional archives restrict access to their physical holdings for preservation reasons, digital archives can provide high levels of access to resources, such as via the WWW. On the other hand, many LRA resources are of a sensitive or restricted nature, and access needs to be controlled; here digital technologies offer appropriate means of authorizing users and ensuring that access protocols are protected.

their holdings, presenting them seamlessly to users. In addition, users anywhere are potentially able to add or update resources, notes and comments, and to collaborate with others in performing such enrichments.

Enrichment

The WWW not only offers access to resources from most parts of the earth, but also the possibility for federated LRAs to interconnect

Legal and ethical framework

LRAs hold valuable intellectual property which needs to be protected according to the wishes of depositors, those recorded, communities, and according to relevant laws on privacy and copyright. Depositors should be confident that LRAs will uphold the highest legal and ethical standards.

Resources

Today's Language Resource Archives (LRAs) hold a wide range of digital resources from various areas

in linguistics including corpus linguistics, language documentation, historical linguistics, sign languages research, computational linguistics, phonetics, speech engineering, and many others.

These digital resources are of many types:

text, such as:

- text corpora
- digitized books and manuscripts
- web-sites
- treebanks

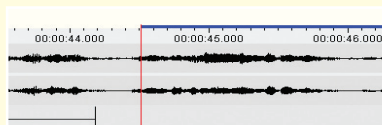
media, such as:

- sound and video recordings
- spoken corpora
- multimodal materials

knowledge sources, such as:

- lexica
- grammars
- metadata and ontologies
- geographic information

Each resource type has distinct storage requirements. Texts are compact in size but may include complex links and relationships across many files. Sound and video demand high storage capacity and do not readily offer means of search or browse. Today's LRAs expect to store up to 50 terabytes with even greater demands in the future.



File	Tier	Before	Annotation	After	Begin Time	End Time
1 NCT_AH_A60	Translation	A dog was	A dog was	Here, that hole.	00:00:00:00	00:00:10
2 NCT_AH_A60	Translation	Water was stre	The dog	Suddenly he was	00:00:27:00	00:00:31
3 NCT_AH_A60	Translation	The dog, will pr	Suddenly he	The dog barked	00:00:31:00	00:00:35
4 NCT_AH_A60	Translation	Suddenly he was	The dog	The dog ran fast	00:00:35:00	00:00:39
5 NCT_AH_A60	Translation	The dog barked	The dog	That other hole	00:00:39:00	00:00:43
6 NCT_AH_A60	Translation	If only I had that	The dog	He growled "I'm	00:00:43:00	00:00:48

