

## Goals and Federation

DAM-LR is integrating the language resource archives (LRAs) of the partner institutions so that they appear to users as one single large repository. The DAM-LR partners support the „Live Archives“ principles for Digital Language Resource Archives.

### Goals

The goals of the DAM-LR project are to create an integrated and unified domain of:

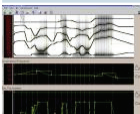
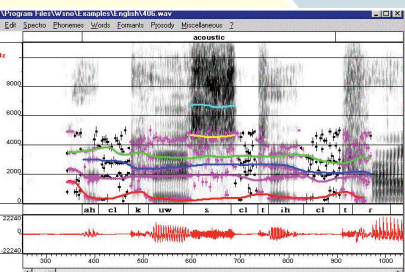
- trusted servers and services
- deep metadata for research purposes
- stable and unique resource identifiers
- user management and authentication
- exchange of user credentials for access authorization
- longer-term potential for exchanging resources to strengthen preservation purposes

### Federation Platform

DAM-LR's integrated and unified domain can be called a Federation of LRAs, if the partners agree on issues including:

- a shared mission to provide integrated services
- mutual trust that the partners follow mutually agreed rules
- a common ethical and legal ground for all joint activities
- a number of practical guidelines such as which user credentials are to be exchanged
- the originating institute for a resource retains control of rights and access to it
- each LRA retains independence of operation

DAM-LR will establish a formal federating agreement in 2006.



## DAM-LR



### Partner Institutions

The DAM-LR project was started as a small scale project under the 6th Framework Program of the European Commission (DG Research) to introduce Grid technology and to virtually connect the language resource archives that are housed by the partner institutions listed below.

#### MPI for Psycholinguistics, Nijmegen (coordinator)

The MPI stores several types of linguistic resources (corpora, lexica etc.) collected by a variety of projects ranging from child acquisition, second learner acquisition, national spoken corpora, and sign language to endangered languages (DOBES). Currently, the archive holds more than 150.000 objects occupying over 15 terabytes. It is organized using the open IMDI metadata infrastructure.

#### Centre for Languages and Literature, University of Lund

The Lund centre houses a broad range of language departments and a high-end laboratory tailored to the study of cognition and language behaviour online. The centre relies on IMDI for the organization of its various language corpora - from child language development, dialects, field research in South East Asia - as well as its rapidly growing body of data from eye-tracking research (mostly reading), finger-tracking research tailored to the study of tactile reading, motion-tracking research (language and gestures), phonetic investigations, studies of online writing, and electrophysiological measures of reading, writing, speaking and listening activities.

#### SOAS, University of London

ELAR, The Endangered Languages Archive at SOAS, is a digital archive for worldwide endangered languages resources. The collection consists of linguistic and multimedia documentation materials deposited by funded researchers and others.

#### Institute for Dutch Lexicology, Leiden

The INL collects Dutch words. These words are stored in a database: The Language Database. The INL's TSTcentrale (language and speech technology centre) maintains and distributes state-of-the-art digital (Dutch) language resources such as the Dutch Spoken Corpus (CGN).

[www.mpi.nl/dam-lr](http://www.mpi.nl/dam-lr)

A 6th Framework Program Project DG Research

# DAM-LR

## Distributed Access Management for Language Resources

A Grid Project

March 2006

## Phases

### First Phase (2005)

The DAM-LR project started in January 2005. Its main objectives in the first phase were:

- to establish stable and comprehensive archives at the partner institutions (partners have invested considerable resources to accomplish this)
- to discuss and establish the basics for a federation of archives
- to discuss the architecture and components of the integrated archive domain and start installing, developing and testing

## Architecture

The DAM-LR architecture is **fully distributed** and based on components drawn from the Grid, Digital Library and Language Resource communities. It supports normal web browsers as well as more advanced operations through the use of specialized web applications.

### Network and Trust

DAM-LR relies on continuously improving network infrastructures such as GEANT which will soon support the fast transmission of high-quality video. To implement a trusted domain of servers and services, the TACAR list of mutually agreed certificates was created, based on the principles of EUGridPMA. National bodies declare that they will accept certificates from each other, with a Public Key Infrastructure used to sign certificates. DAM-LR will make use of these agreements, i.e., each partner will become a certification or registration authority to be able to issue/request accepted certificates.

### Second Phase (2006)

Now that the basic design and architecture have been settled, the principles of a federation discussed, and components are being installed, tested and deployed, the DAM-LR project starts its second phase. In this phase we will carry out (1) the necessary integration work and (2) start preparing for dissemination and training. This education and training aspect will be offered to other interested institutions. Seminars will be conducted, mainly in Lund, for interested parties to train them in the relevant technologies.

### Metadata

The unified metadata domain provides detailed metadata descriptions to support research queries. The existing, well-known IMDI metadata infrastructure is used ([www.mpi.nl/IMDI](http://www.mpi.nl/IMDI)), although each partner is free to use its own metadata set. For more general resource discovery, OLAC records will also be created.

### Unique Resource Identifiers

Resources will have stable and unique identifiers, known as URIDs. The Handle System was selected to manage these identifiers. Since access conditions will be the same for all instances of a resource, this information is also coupled with the URIDs.

### User Management

User administration is undertaken locally by each partner institution, using LDAP. The DAM-LR partners agreed on the set of user credentials to be collected and exchanged. DAM-LR partners can provide accounts for external users that are not staff/students of their institutions.

### Third Phase (2007- ...)

The third phase will start in 2007 when we intend to (1) optimize the integration technology and (2) **scale up** the federation within the 7th Framework Program of the EC. The DAM-LR federation can be seen as the basis for a broader research infrastructure for the domain of Language Resources and Technology.

### Access Authorization

To determine whether a particular user has permission to access a resource, we use Shibboleth to securely exchange user credentials between a user's home institution and the institution granting access. Shibboleth has already proven its robustness and has become widely used in the Digital Library field.

### Future Components

- We are still working on some components:
- a Resource Manager to compare the credentials provided by a home institution with access rights data held with the URIDs
  - an Access Manager to allow managers to define permissions and policies required for a distributed scenario
  - some partners are adapting their local architecture to meet DAM-LR requirements

