



UMR 6057 CNRS

Laboratoire Parole & Langage

Université de Provence Aix-en-Provence, France

SLDR - Speech and Language Data Repository - www.sldr.org

SLDR (*Speech and Language Data Repository*) is offering labs and scholars a free-of-charge service for sharing and archiving oral/linguistic resources with the help of procedures compliant with the OAIS model for long-term preservation. It originated (as CRDO-Aix) from a call for projects by *Direction de l'Information Scientifique* and the Department of social sciences and humanities at CNRS (France) in 2006. After the completion of a pilot project coordinated by TGE-Adonis in 2008-2010, SLDR is fully operational as an interoperable open system involving two major computing centres: *Centre informatique de l'enseignement supérieur* (CINES) for long-term preservation and *Centre de calcul de l'Institut national de physique nucléaire et de physique des particules* (CC-IN2P3) for the dissemination of shared resources.

The aim of SLDR is to bring together resources associated with oral communication. This is necessary for avoiding the waste of time and money creating again resources that are already available. Further, resource sharing in speech research is a strategic response to the challenge of capitalizing knowledge from low-interconnected scientific domains such as descriptive, formal and computational linguistics, sociolinguistics, literature, language translation, neuroscience, psycholinguistics etc.

Data shared with the scientific community (and the public at large) include all signals associated with oral production, documents created or collected during an experiment or a field enquiry, material derived from primary data (transcription, translation, annotation, analysis...) with their associated resources

(lexica, grammars, frequency tables, knowledge bases...) and tools designed for data processing in the domain. A corpus may contain audio/video recordings as well as measurements of physiological activity: endoscopy, palatography, evoked potential etc.

Speech & Language Data Repository

Speech & Language Data Repository (SLDR) <http://sldr.org>

Logos: tge, DONIS, CLARIN, Open archives (OAI-PMH)

Navigation: [Sign up] / [Login] / 中文 / English / español / français

Flash Player plugin is required to read previews. If the preview cell is empty or the player is not active, you may download Adobe Flash Player.

Display corpora in: Arabic, Standard (العربية الفصحى) [arb] (1) [ok]

1 item(s)

Id	Name	Description	Downloading	Preview
[ARK] 101002745 => wiki	Standard ARABIC - statements (Amina CHENTIR)	language_description/narrative (Publications) Laboratoire parole et langage (LPL, Aix-en-Provence FR) Affirmative sentences in standard Arabic recorded in a sound-proof studio by 5 North-African speakers (2 men and 3 women). These sentences contain all the phonemes of standard Arabic. جمل مفيدة باللغة العربية الفصحى مسجلة داخل غرفة صامتة، من طرف خمسة أشخاص من شمال إفريقيا (رجال و نساء). تحتوي على جميع فونيمات العربية الفصحى (language_documentation)	You are not logged in. Downloaded 3 time(s) (see members)	Preview icons

SLDR architecture is based on the *Open Archival Information System* (OAIS) and it allows access to documents via unique and persistent identifiers. A procedure has been implemented to reconstruct items from datastreams stored at the dissemination site (CC-IN2P3), thereby making it possible to release storage space on the submission site. In response to the most diverse needs, our team opted for a generic submission system minimizing technical constraints on the structure and size of resources: large items are automatically segmented to smaller information packages and there is no restriction on file names. The work environment is also set for medium-term archiving and it offers a flexible and accurate management of access rights at different phases of a project.

Downloading documents may be restricted to certain categories of users. Access restrictions pertaining to documents stored as public archives are compliant with derogations to the principle of open access listed in the French *Code du patrimoine* (art. L213-2, act of 15 July 2008).

SLDR website keeps a trace of transactions on some resources so that their users may be bound by a non-commercial licence to mentioning their usage on any related scientific production. This follow-up is beneficial to collaborative work beyond geographic and institutional barriers.

The site is accessed in four languages: English, Chinese, Spanish and French. It takes advantage of the Web 2.0 approach for users' communities and it provides links to social networks.

Contact : <http://sldr.org/contact>

