

The OAIS model in Digital Humanities: feedback from archival practice in linguistics

<http://sldr.org/doc/show/roma-sept2015/OAIS-SLDR-en.pdf>

Bernard Bel <contact@slidr.org>
Laboratoire Parole et Langage (LPL)
CNRS – Aix-Marseille University

Tavola rotonda: Speech audio archives: preservation, restoration, annotation, aimed at supporting the linguistic analysis. Centro Linceo Interdisciplinare “Beniamino Segre”, Roma (Italy), 14 sept. 2015



ORTOLANG is a beneficiary of French State support via its
“Investissements d'avenir” program (ANR-11-EQPX-0032).



Abstract

- OAIS is a conceptual framework for an **Open Archival Information System** (ISO 14721:2003) initially designed for constructing archives of quantitative data. In 2006, the National Center for Scientific Research in France ([CNRS](#)) launched a pilot project setting-up services for the sharing and long-term preservation of speech data and related material. Several institutional partners were involved, among which the [CINES](#) (National Computing Center for Higher Education).
- The **Speech & Language Data Repository** (SLDR, www.sldr.org) is an outcome of this project. Success of the venture depends on fostering an effective collaboration between the major actors: research scholars, curators, archivists and computer scientists. We will expose (positive and negative) examples of digital curation to identify good practice in terms of deposit, descriptive metadata, integrated access-rights management and the packaging of archival items for their medium and long-term storage.

Who are we?

The **Speech & Language Data Repository** ([SLDR](http://www.sldr.org) – www.sldr.org) is a digital repository for the **sharing** and long-term **preservation** of linguistic data created by *Laboratoire Parole et Langage* ([LPL](http://lpl.cnrs.fr), UMR 7309 CNRS-AMU).

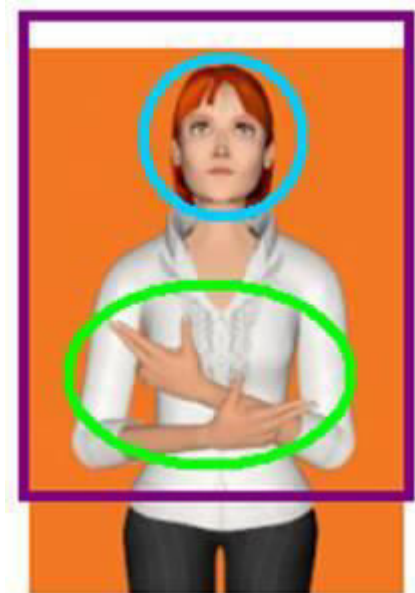
The aim of SLDR is to bring together resources associated with **oral communication**. This is necessary for avoiding the waste of time and money creating again resources that are already available. Further, resource sharing in speech research is a strategic response to the challenge of **capitalizing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

SLDR is open to research projects regardless of geographical/national locations. The site is accessed in four languages: English, Chinese, Spanish and French.

SLDR is in the process of migrating its data and services to the new **ORTOLANG** platform (*Open Resources and TOols for LANGuage*) in which new tools for the processing of oral/written linguistic data will be made available. [ORTOLANG](http://ortolang.org) is a service centered on language studies completing the general offer of [Huma-Num](http://huma-num.org) (a Very Large Research Infrastructure), aiming at **interoperability** with European networks [CLARIN](http://clarin.eu) (linguistics) and [DARIAH](http://dariah.eu) (*Digital Humanities*).

SLDR architecture is based on the **Open Archival Information System** (OAIS) and it guarantees access to documents via **persistent identifiers**. The dissemination site is hosted by [INIST](http://inist.fr). Submissions for long-term preservation are sent to [CINES](http://cines.fr) following a procedure engineered by Huma-Num.

A broad range of disciplines...



from experimental linguistics...



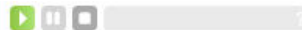
... to field
linguistics



... including didactics, processing written forms, understanding of speech information, the assessment and retraining of voice, speech and language dysfunctioning etc.

[ARK] **Primary data (corpus)** Derrida - State Of The Lie, Lie of the State (Bernard BEL)
 Centre de sciences humaines (CSH, New Delhi IN)
 "State of the Lie, Lie of the State": A conference by Jacques Derrida at Delhi School of Economics, 24 January 1997. [More]
 (philosophy_of_language)
English

embargo -> 2047



[ARK] **Tool** Anonymise sound files (Daniel HIRST)

Laboratoire parole et langage (LPL, Aix-en-Provence FR)

PRAAT script.

Purpose: replace portions of a long sound which are labelled with a key word on the accompanying TextGrid with a hum sound with the same prosodic characteristics as the original sound

Original long sound can be mono or stereo, anonymised sound will be same. [More]

(applied_linguistics, cognitive_science, language_documentation, speech_prosody, computational_linguistics)

>> Collection LPL tools crdo000763



Anonymise sound files by Daniel Hirst is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

public - Creative Commons licence

Items of SLDR include speech/singing corpora, their annotations, lexica and other

[ARK] **Resource** Valjouffrey - work (Audrey THOMAS, Clément GIRARD, Médéric GASQUET-CYRUS)

Laboratoire parole et langage (LPL, Aix-en-Provence FR)

Département de linguistique et phonétique générales, Université de Provence (Aix-en-Provence FR)

1) Translations, transcriptions and annotations of corpora in Valjouffrey patois (dialect)

2) Sound/video files, annotations and subtitling files in public access for streaming examples [More]

(language_documentation, lexicography, sociolinguistics, anthropological_linguistics,

text_and_corpus_linguistics, phonology, phonetics)

Occitan (post 1500); Occitan (post 1500) -> Valjouffrey dialect

>> Collection Valjouffrey crdo000007

non-commercial licence



knowledge bases as well as tools

[ARK] **Primary data (corpus)** PSH/DISPE - Subaquatic and hyperbaric speech (Alain MARCHAL)

Laboratoire parole et langage (LPL, Aix-en-Provence FR)

Issu d'une collaboration entre le Laboratoire Parole et Langage (LPL) et l'Institut de Plongée Professionnelle de Marseille (INPP) en 1991, le corpus PSH/DISPE répond à la demande d'une base de sons pour le développement de nouveaux procédés de « décodage » de la parole hyperbare, et d'un outil pour l'évaluation des systèmes de communication vocale.

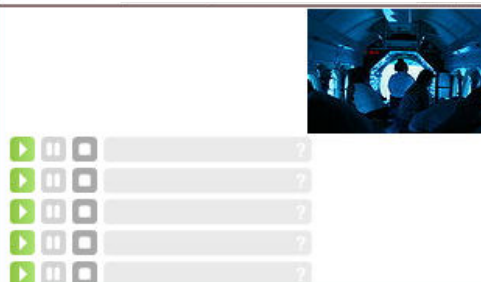
Les fichiers d'annotations sont conformes au format standard SAM Europec du projet CEE-ESPRIT n°2589.

[More]

(phonetics, phonology)

English; French

non-commercial licence



associated with data processing. Corpora

[ARK] **Primary data (corpus)** Mờ Piu (Geneviève CAELEN-HAUMONT)

Multimedia, Information, Communication and Applications (MICA, Hanoi VN)

Data collected from the Mờ Piu community during an initial field trip in Nậm Tu Thợng between 8 and 12 June 2009. Recordings of speech, songs and music are indexed on day, theme, speaker, question, (song or musical piece). Each question/answer pair is assigned a WAV file. Recordings are in stereo format. On the first track the day, theme, speaker and question number are indicated, followed by a question in Vietnamese (asked by a Vietnamese speaker). On the second track the same question is asked in Mờ Piu and followed by its answer.

4 male speakers, 3 female speakers, 7 female singers et 2 male singers have been recorded. In the whole, 7 hours of speech and 1 hour of singing. [More]

(anthropological_linguistics)

Vietnamese -> Mờ Piu

in progress

Rapport.pdf

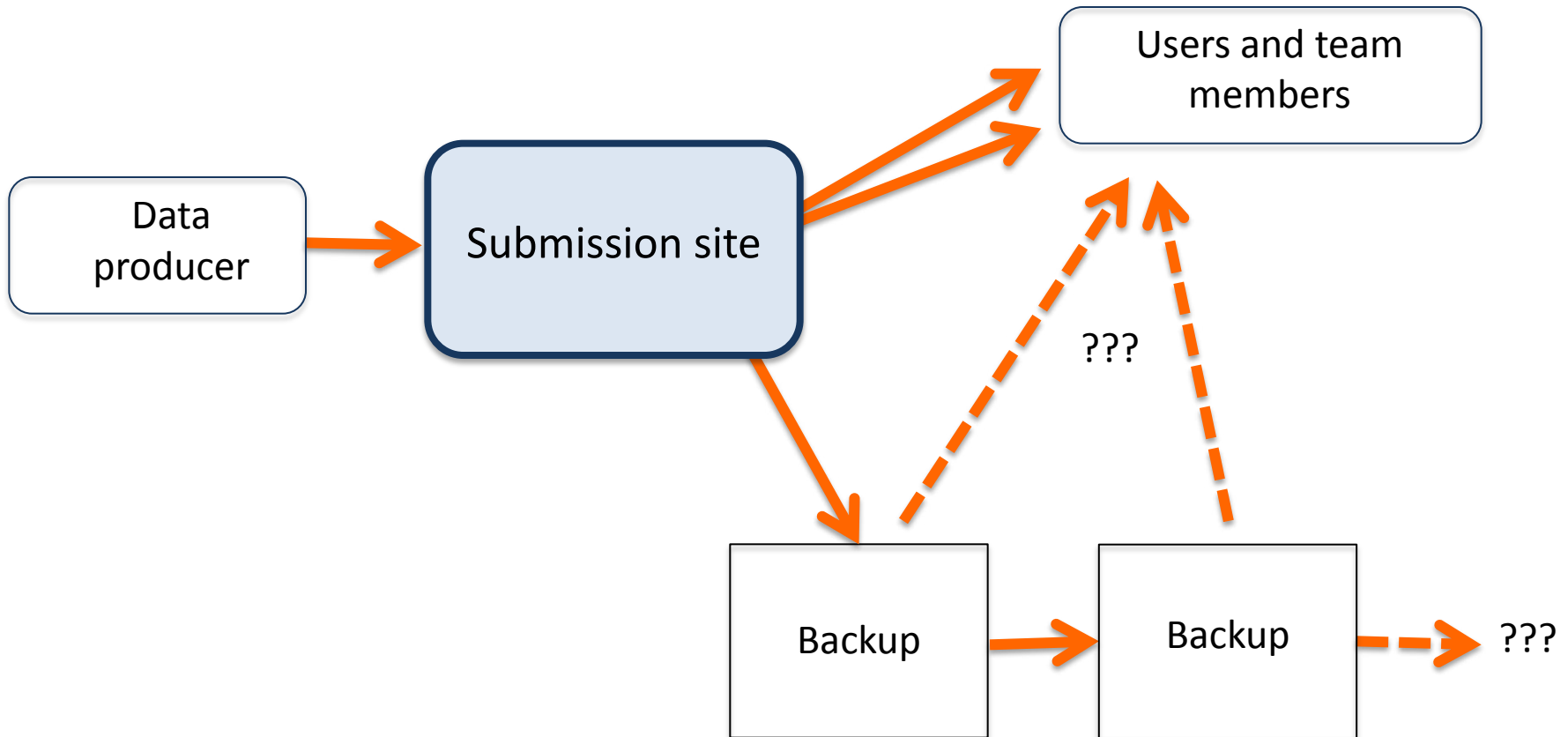


may comprise audio/video recordings and measurements of physiological activity.

Why and how we use the OAIS model



A classical set-up: safe storage and dissemination



A classical set-up: safe storage and dissemination (ctd.)

Problems commonly associated with safe storage

1. Access to data is **centralised**: the dissemination site is also the submission site. => **Weak argument**
2. Financial **support** to backups is granted for ... years. What will happen beyond this period? Who will negotiate new arrangements? On what legal/ethical grounds? => **Strong argument**
3. Format **migration** is under the care of data producers. What will happen in the long term? => **Very strong argument**

The OAIS solution

- Preservation + migration of data is handed over to an **institutional archive** (or, at least, a Trusted Data Repository).
- This solution generates **technical** and **legal** constraints.

Why and how we use the OAIS model...

- The **Open Archival Information System (OAIS)** is ISO 14721 standard. It was imposed by institutions heading our pilot project for the sake of formal agreements between the [French National Archive](#), [CINES](#) and [CNRS](#).
- We drew experience from the practice of long-term data preservation in astrophysics.
- The actual implementation takes into account **specific features of oral/linguistic** resources, notably:
 - The **diversity of file formats**: sound/video, all signals associated with speech/singing, pictures, texts, tables etc.
 - Secondary data (annotations etc.) and metadata are **mutable and extensible**.
 - A **multilingual approach** of descriptive metadata, extra-European scripts, transliteration/annotation standards (IPA etc.)

Why and how we use the OAIS model...

- Aim 1: **preserving** data
- Aim 2: making it accessible and **eligible for reuse** after an *unspecified period of time*. This is the challenge of **long-term preservation**.
- These issues deal with the **very long term**, practically meaning **more than 30 years**.
- Nonetheless, **medium-term** preservation is also based on **the same OAIS framework**. => **This is cost-effective!**

Implications of long-term preservation

- Preserve the **intelligibility of a document**
 - The archive site accepts a restricted number of **preservation formats** whose specifications are freely accessible.
=> <http://sldr.org/wiki/Formats>
 - WAV, AIFF, AAC, FLAC, MP4, OGG, MKV, PDF, PDF/A, PNG, TIFF, SVG, GIF, XML, HTML, TXT, ODT
 - The archive site ([CINES](#)) is committed to **migrating** formats once they have become obsolete. *This is the job of archive curators, not data producers!*
- Preserve the **signification of its content**
 1. **Descriptive** metadata (public + confidential);
 2. **Archival** metadata;
 3. A formal description of **conventions** used in the archive.
PPDI = *Project Preservation Description Information*.
<http://sldr.org/ppdi>

The “pilot project” OALS framework (2008-2010)



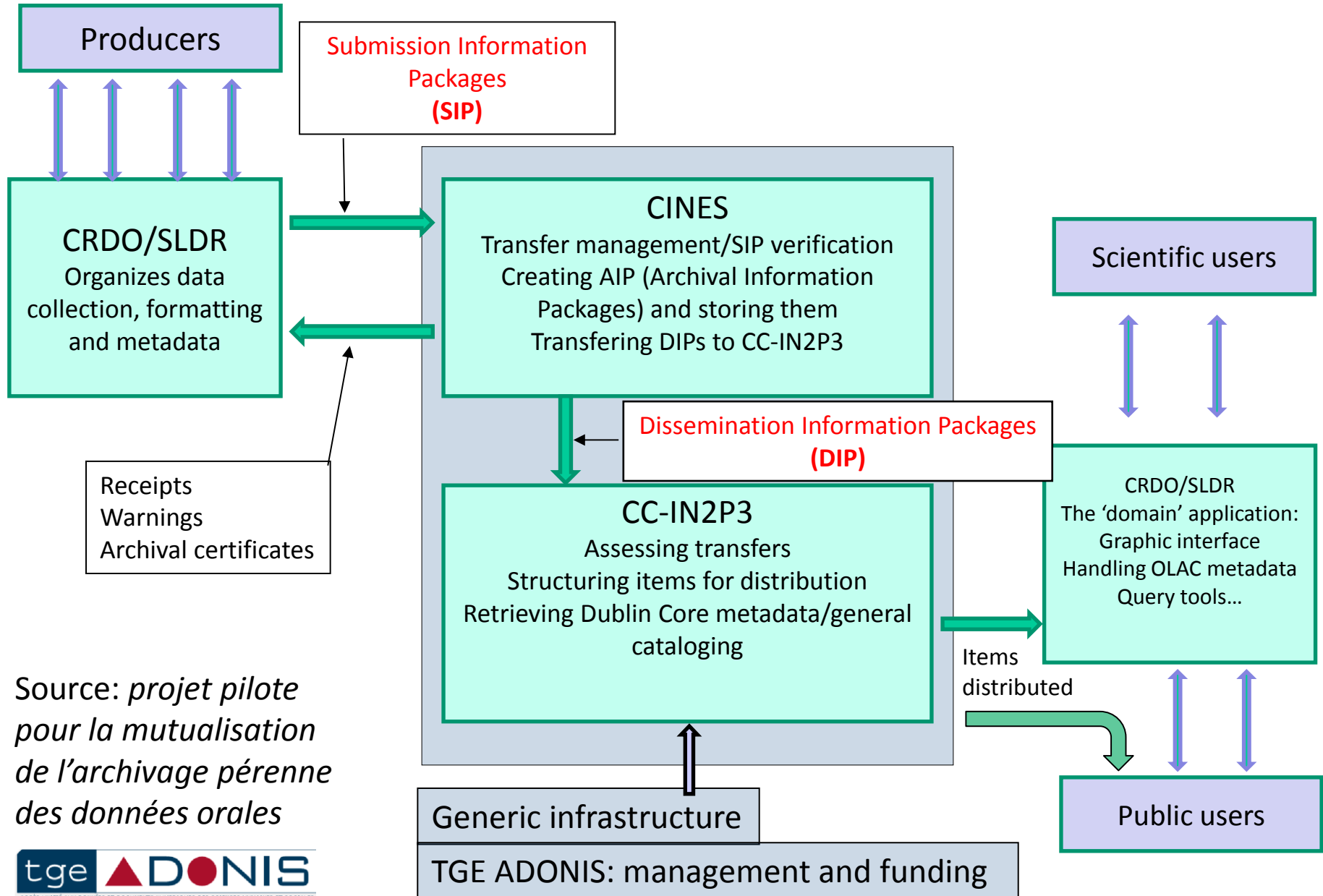
The “pilot project” OAIS framework

The pilot project involved 3 actors:

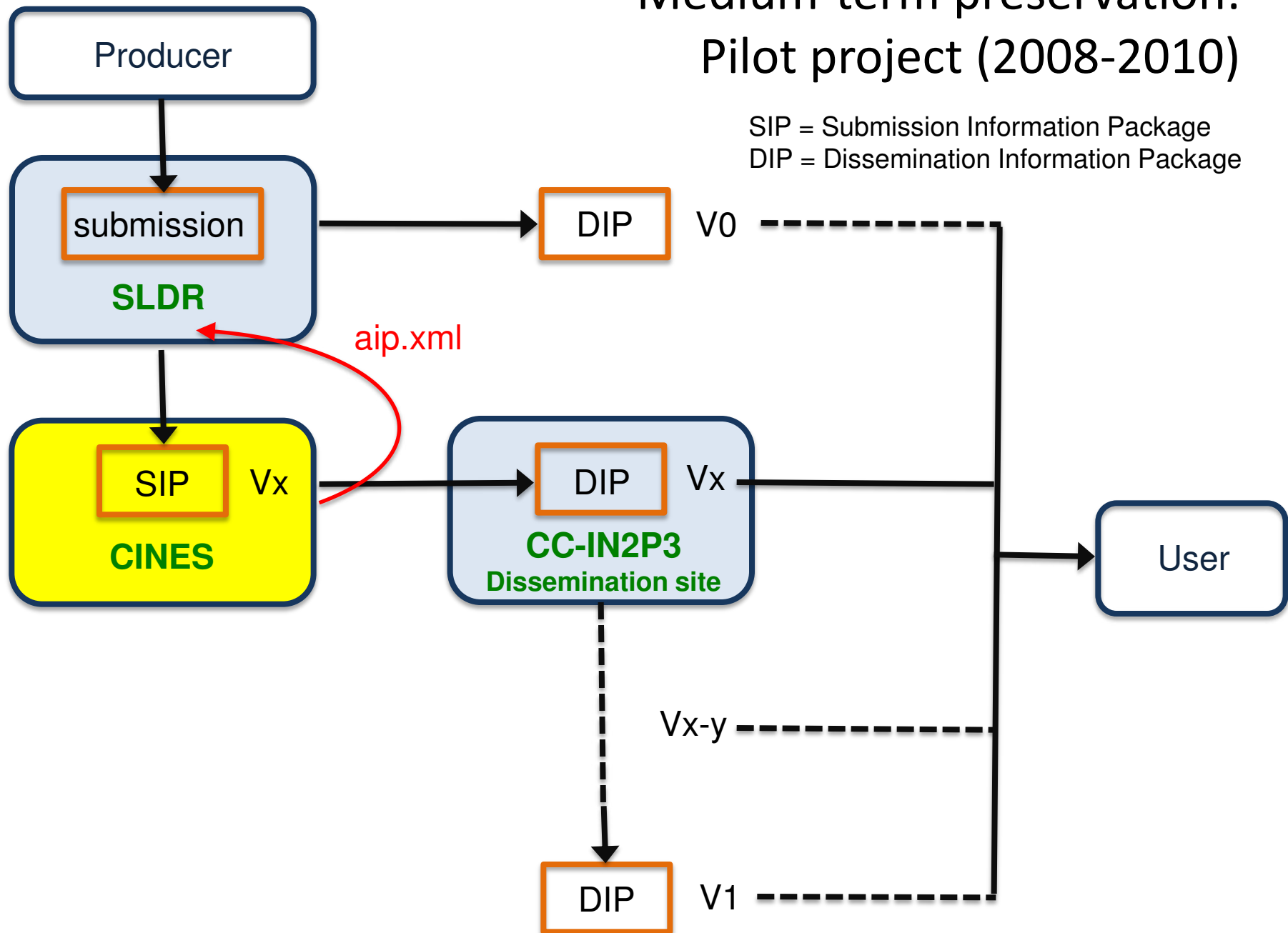
- 1) **Submission sites:** CRDO-Aix (SLDR) and CRDO-Paris
- 2) **An archive site:** *Centre informatique national de l'Enseignement supérieur* ([CINES](#), Montpellier)
- 3) **A dissemination site:** the TGE-Adonis grid at *Centre de calcul de l'Institut national de physique nucléaire et de physique des particules* ([CC-IN2P3](#), Lyon)



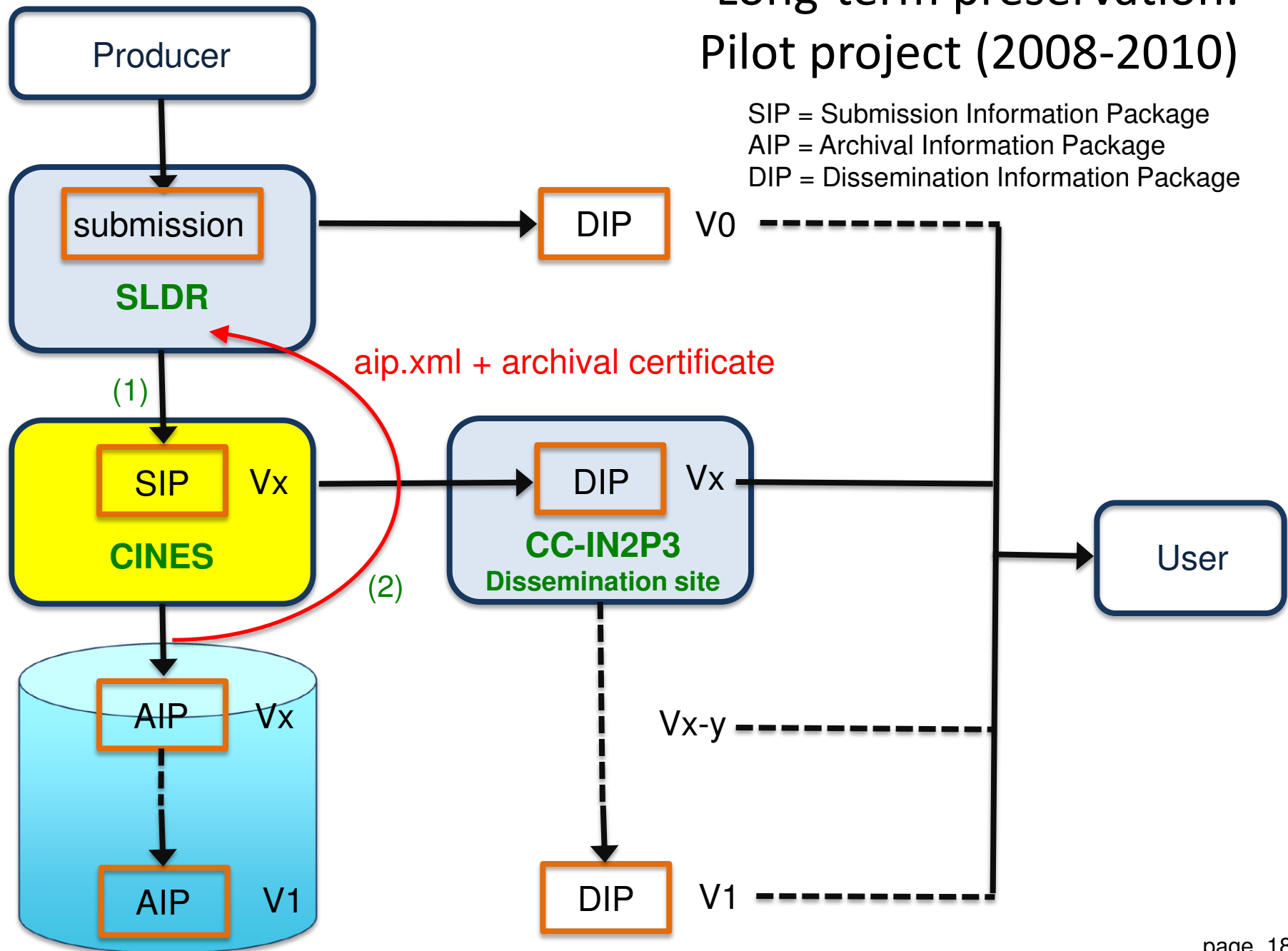
A functional diagram of the OAIS setup



Medium-term preservation: Pilot project (2008-2010)



Long-term preservation: Pilot project (2008-2010)



Two problems solved
by the Pilot project...



How to deal with mutable/extendible data?

Problem

- In a given item there are bits of information that remain persistent (e.g. primary data of sound/video resources) whereas others (such as descriptive metadata) may need to be **updated more frequently**.
- An institutional archive preserves **all versions** of information packages that have been submitted.
- It would be detrimental to submit again the entire package (including large amounts of primary data) every time a **small amount** of descriptive data has been modified.

Solution worked out with CINES

- Each *Submission Information Package* (SIP) contains two folders: one which is dedicated to **persistent data** and the other one to **metadata and documentary files**.
- It is possible to submit a new version either for the entire package or for its documentary files. The latter is named '**metadata updating**'.

What to do with 'uncompliant' files?

Problem

- A few file formats (e.g. MP3, FLV) are **uncompliant with long-term preservation** though they are needed for data dissemination (e.g. streaming).
- In the OAIS model all data is **first sent to the archive site**, checked, stored in the archive and thereafter forwarded to the dissemination site.

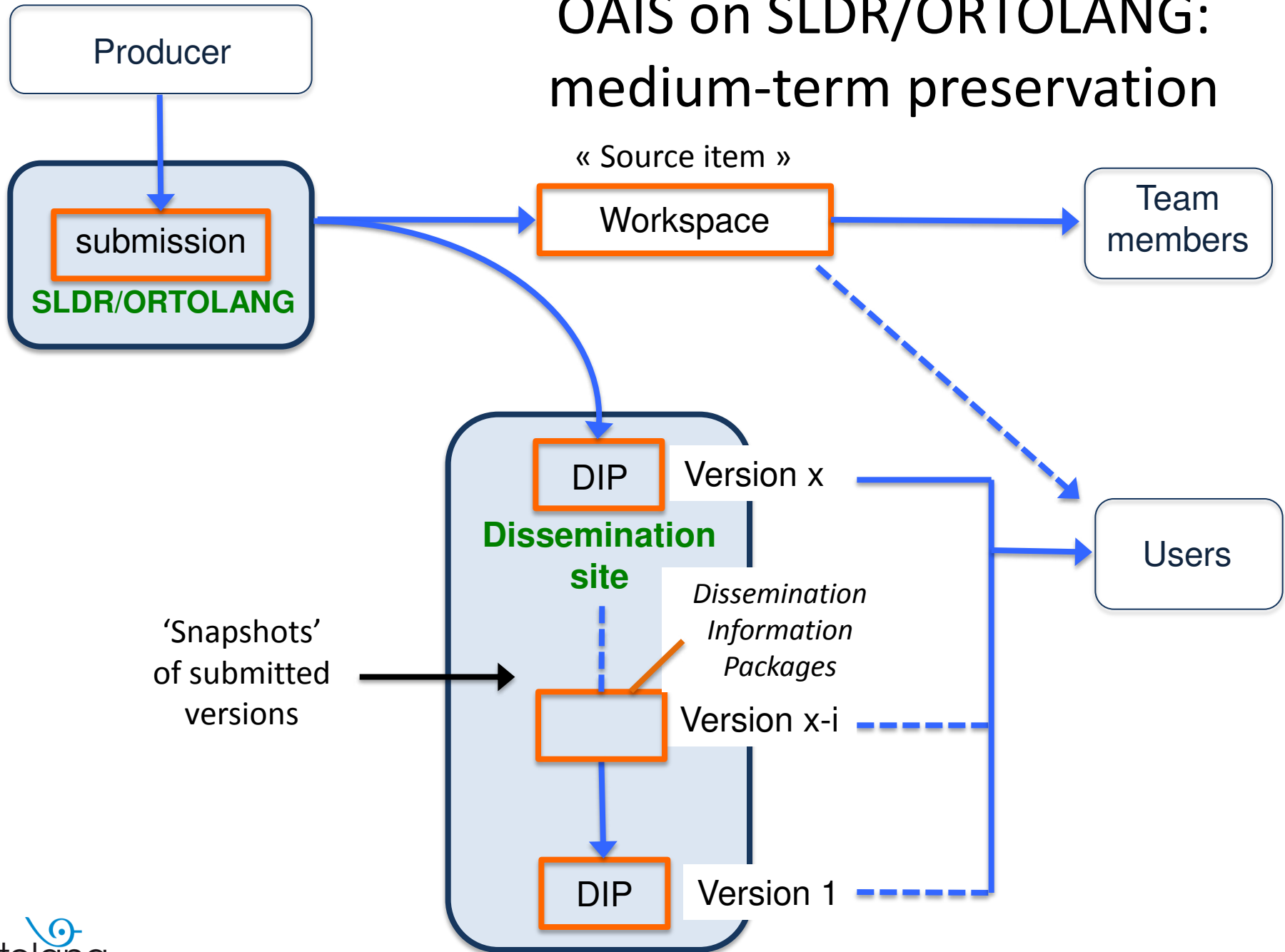
Solution worked out with CINES

- Submission information packages contain a **dissemination folder** whose content is ignored by the archive site. It is forwarded to the dissemination site as part of the *Dissemination Information Package* (DIP).
- This dissemination folder may also contain files requiring **medium-term preservation** because they are subject to updates/deletion and/or their formats are not (yet) recognised by the archive site.

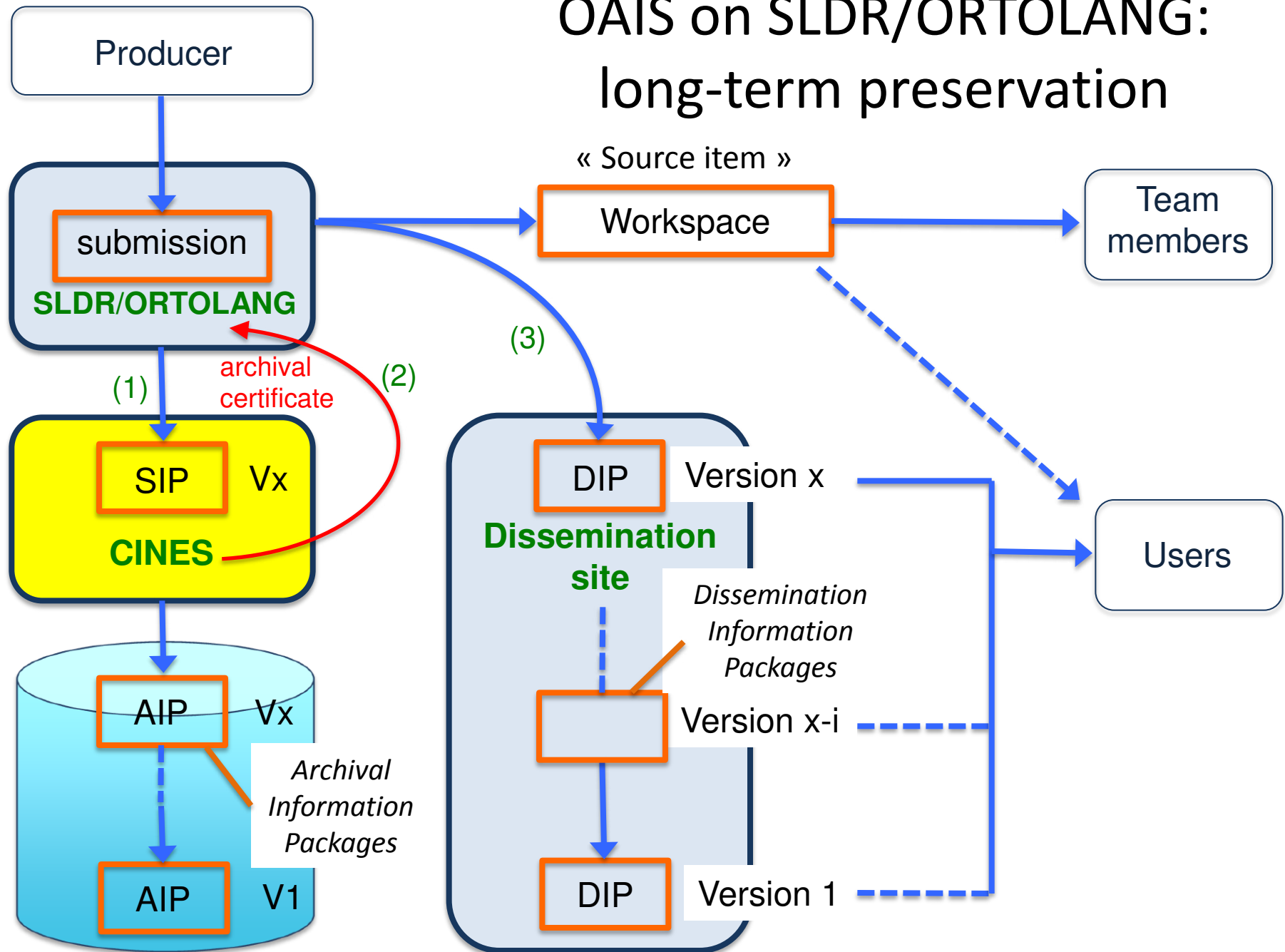
The “SLDR/ORTOLANG” OAIS framework














OAIS on SLDR/ORTOLANG: medium-term preservation



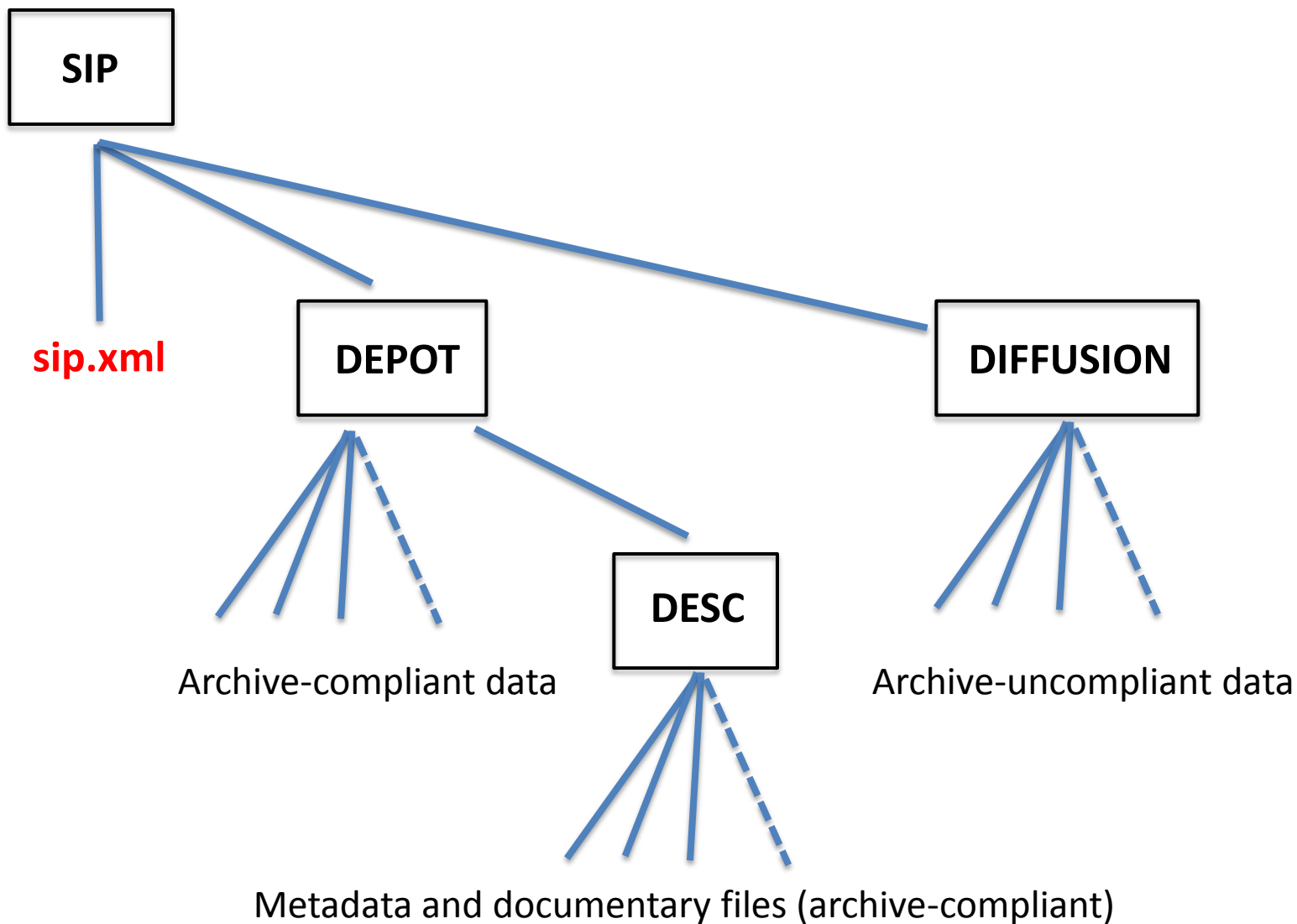
OAIS on SLDR/ORTOLANG: long-term preservation



Creating submission information packages (SIP)

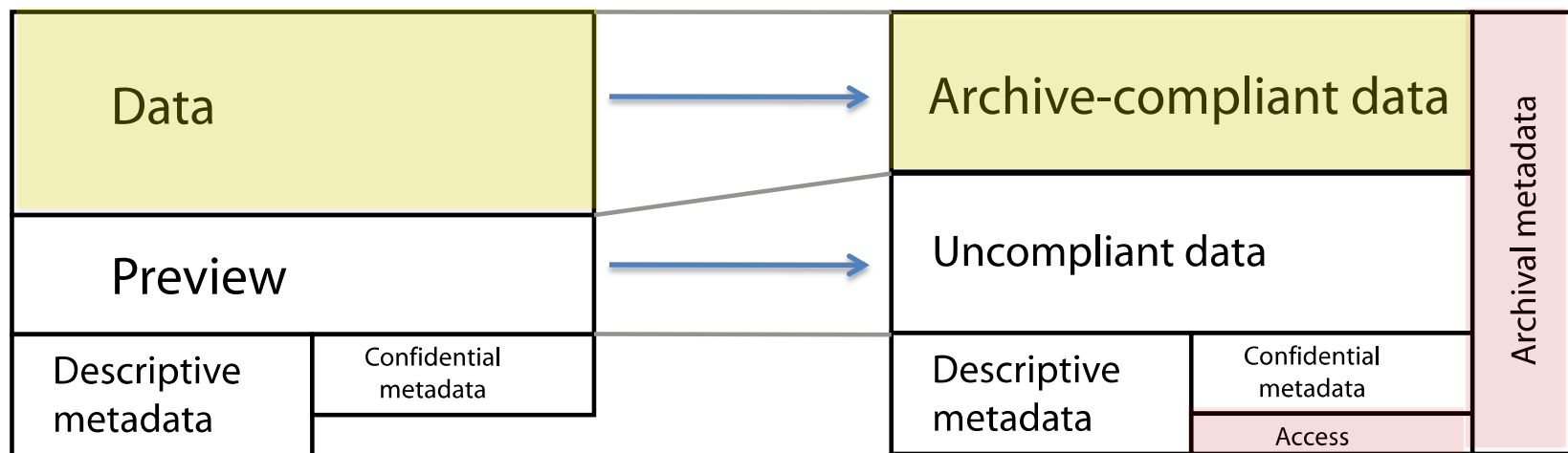
 = 20kgs	 = 20kgs
 = 20kgs	 = 40kgs
 = 30kgs + FREE 	 = 30kgs + FREE 
 = 40kgs	 = 50kgs + FREE 

Structure of the *Submission Information Package*



Creating the submission information package

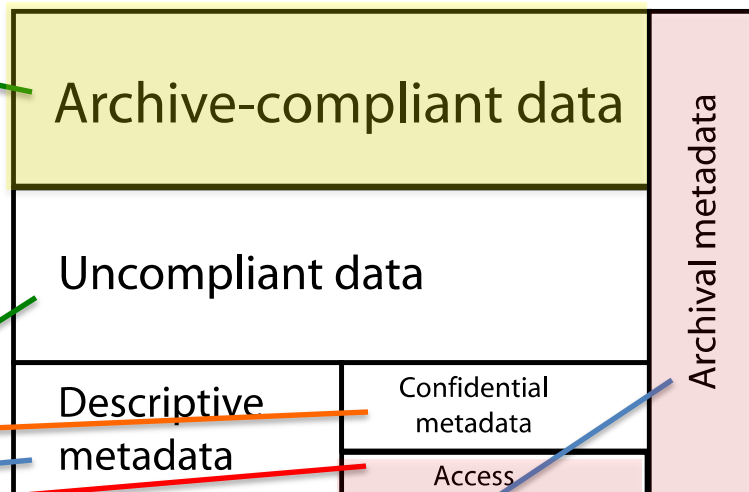
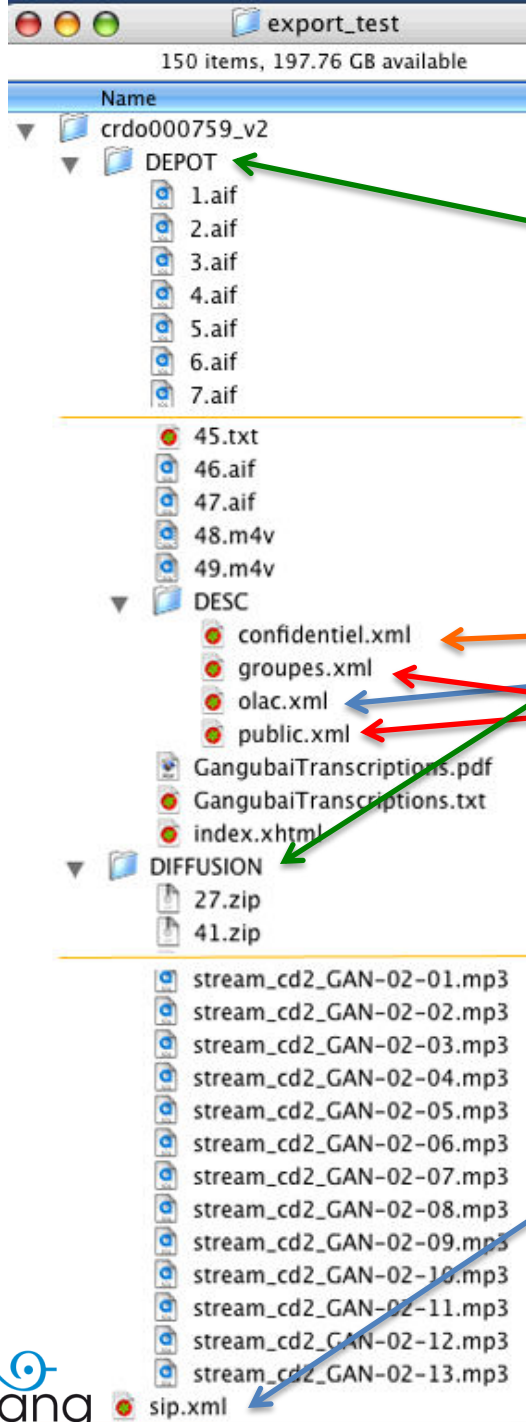
(See: <http://sldr.org/wiki/Packaging-en>)



Item stored at SLDR



Submission Information Package (SIP)



(See: <http://sldr.org/wiki/Packaging-en>)

Archival metadata



Archival metadata: the 'sip.xml' file

(Old version)

```
=== SIP de crdo000759_v2 ===
Nouvelle version = 2
<?xml version="1.0" encoding="UTF-8" ?>
<pac xmlns:pac="http://www.cines.fr/pac/test/avis" xmlns="http://www.cines.fr/pac/test/sip" xml
http://www.cines.fr/pac/test/sip.xsd">
<DocDC>
<title>en: Gangubai</title>
<title>es: Gangubai</title>
<title>fr: Gangubai</title>
<title>mr: गंगुबाई</title>
<creator>Bernard BEL</creator>
<creator>Hema RAIRKAR†</creator>
<subject>en: subaltern studies, cooperative action-research, musicology, social anthropology, lep
<subject>fr: Etudes sur la subordination, recherche-action coopérative, musicologie, anthropolog
<subject>mr: दबलेल्या गटांचा अभ्यास, सहकारी पध्दतीने केलेले कृती-संशोधन, सामाजिक मानव वंशशास्त्र,
<subject>mar</subject>
<subject>linguistics_and_literature</subject>
<subject>pragmatics</subject>
<description>en: An interview with Gangu Ambore, a leper woman in Tadkalas, Parbhani district
support of grindmill and devotional songs borrowed to a popular bhakti tradition.</description>
<description>fr: Entretiens avec Gangu Ambore, une chanteuse lépreuse de Tadakalas, district de
la mouture et de poèmes dévotionnels empruntés à une forme populaire de bhakti.</description>
<description>mr: गंगुबाई अंबोरे या ताडकळस, जिल्हा परभणी, महाराष्ट्र, भारत, इथे रहाणार्या महारोगी बा
भक्तिपरंपरेतील अभंग, गौळणीच्या आधारावर अभिव्यक्त करतात.</description>
<publisher>Centre for Cooperative Research in Social Sciences (CCRSS, Pune IN)</publisher>
<contributor>Bernard BEL (creator)</contributor>
<contributor>Hema RAIRKAR† (creator)</contributor>
<contributor>Gangu AMBORE (singer)</contributor>
<contributor>Jitendra MAID (interviewer)</contributor>
<contributor>Bhimsen NANEKAR (interviewer)</contributor>
<contributor>Andréine BEL (photographer)</contributor>
<contributor>Datta SHINDE (research_participant)</contributor>
<contributor>Malavika TALUKDAR (video) (photographer)</contributor>
<contributor>Bernard BEL (video) (photographer)</contributor>
<contributor>Guy POITEVIN (annotator)</contributor>
<contributor>Gajarabai DAREKAR (transcriber)</contributor>
<contributor>Rajani KHALADKA (transcriber)</contributor>
<date>2010-11-02</date>
<type>corpus</type>
<format>AIFF</format>
```

This file contains archival metadata: a brief description of the item, relations to other items, versioning, storage instructions and specifications of all files submitted for long-term preservation.

Archival metadata: the 'sip.xml' file

(Old version)

=== SIP de crdo000759_v2 ===

Nouvelle version = 2

<?xml version="1.0" encoding="UTF-8" ?>

<pac xmlns:pac="http://www.cines.fr/pac/test/avis" xmlns="http://www.cines.fr/pac/test/sip" xml

http://www.cines.fr/pac/test/sip.xsd">

<DocDC>

<title>en: Gangubai</title>

<title>es: Gangubai</title>

<title>fr: Gangubai</title>

<title>mr: गंगुबाई</title>

<creator>Bernard BEL</creator>

<creator>Hema RAIRKAR†</creator>

<subject>en: subaltern studies, cooperative action

<subject>fr: Etudes sur la subordination, recherche

<subject>mr: दबलेल्या गटांचा अभ्यास, सहकारी पध्दत

<subject>mar</subject>

<subject>linguistics_and_literature</subject>

<subject>pragmatics</subject>

<description>en: An interview with Gangu Ambo

support of grindmill and devotional songs borrow

<description>fr: Entretiens avec Gangu Ambore,

la mouture et de poèmes dévotionnels empruntés

<description>mr: गंगुबाई अंबोरे या ताडकळस, जिल्हा

भक्तिपरंपरेतील अभंग, गौळणीच्या आधारावर अभिव्यक्

<publisher>Centre for Cooperative Research in S

<contributor>Bernard BEL (creator)</contributor>

<contributor>Hema RAIRKAR† (creator)</contributor>

<contributor>Gangu AMBORE (singer)</contributor>

<contributor>Jitendra MAID (interviewer)</contributor>

<contributor>Bhimsen NANEKAR (interviewer)</contributor>

<contributor>Andréine BEL (photographer)</contributor>

<contributor>Datta SHINDE (research_participant)</contributor>

<contributor>Malavika TALUKDAR (video) (photographer)</contributor>

<contributor>Bernard BEL (video) (photographer)</contributor>

<contributor>Guy POITEVIN (annotator)</contributor>

<contributor>Gajarabai DAREKAR (transcriber)</contributor>

<contributor>Rajani KHALADKA (transcriber)</contributor>

<date>2010-11-02</date>

<type>corpus</type>

<format>AIFF</format>

<coverage>Box: name=Maharashtra</coverage>

<coverage>TGN: 1001895</coverage>

<rights>CRDO licence; rightsHolder = Bernard BEL</rights>

<rights>license:URI http://crdo.fr/licenceCRDO.php?version=1</rights>

</DocDC>

<DocMeta>

<authenticite>oui</authenticite>

<dureeConservation>P10000Y</dureeConservation>

<identifiantDocProducteur>crdo000759</identifiantDocProducteur>

<docRelation>

<typeRelation>version</typeRelation>

<sourceRelation>PAC</sourceRelation>

<identifiantSourceRelation>17714</identifiantSourceRelation>

</docRelation>

<docRelation>

<typeRelation>version</typeRelation>

<sourceRelation>Producteur</sourceRelation>

<identifiantSourceRelation>crdo000759</identifiantSourceRelation>

</docRelation>

<serviceVersant>CRDO-Aix</serviceVersant>

<structureDocument>index.xhtml</structureDocument>

<version>2</version>

</DocMeta>

File description in the SIP

```

- <FichMeta>
  <encodage>PCM</encodage>
  <formatFichier>AIFF</formatFichier>
  <nomFichier>6.aif</nomFichier>
  <empreinteOri type="MD5">e06f2782f2992b15336a5734ba0d39c7</empreinteOri>
</FichMeta>
- <FichMeta>
  <encodage>PCM</encodage>
  <formatFichier>AIFF</formatFichier>
  <nomFichier>7.aif</nomFichier>
  <empreinteOri type="MD5">01503f8e495885a1ac18207bced2a7d8</empreinteOri>
</FichMeta>
- <FichMeta>
  <encodage>PCM</encodage>
  <formatFichier>AIFF</formatFichier>
  <nomFichier>8.aif</nomFichier>
  <empreinteOri type="MD5">5dff69adcaa8fe53335af284b5860f9b</empreinteOri>
</FichMeta>
- <FichMeta>
  <encodage>PCM</encodage>
  <formatFichier>AIFF</formatFichier>
  <nomFichier>9.aif</nomFichier>
  <empreinteOri type="MD5">5cc33a222833db40c193e5d65135d040</empreinteOri>
</FichMeta>
- <FichMeta>
  <formatFichier>PDF</formatFichier>
  <nomFichier>BhaktiAFaithForRehabilitation.pdf</nomFichier>
  <empreinteOri type="MD5">6c5814b22e48c3cc87a5db35cebe991f</empreinteOri>
</FichMeta>
- <FichMeta>
  <formatFichier>XML</formatFichier>
  <nomFichier>DESC/confidentiel.xml</nomFichier>
  <empreinteOri type="MD5">ac69593a1922738218ff7365cf6e5a6d</empreinteOri>
</FichMeta>
- <FichMeta>
  <formatFichier>XML</formatFichier>
  <nomFichier>DESC/groupe.xml</nomFichier>
  <empreinteOri type="MD5">ebe5c11589dc2e708f245bf4163b6c17</empreinteOri>
</FichMeta>
- <FichMeta>
  <formatFichier>XML</formatFichier>
  <nomFichier>DESC/olac.xml</nomFichier>
  <empreinteOri type="MD5">3ff9bb6ca93d1e50884decba5217b2e9</empreinteOri>
</FichMeta>

```


Integrated access rights management



Integrated management of access rights

1. France takes advantage from a significant advance on archive law, owing to its legal framework establishing the status of ‘public archive’ and a **uniform set of rules** regulating the access to archived documents: *Code du patrimoine* (the Heritage Code), act of 15 July 2008, articles L213 1-5.
2. This framework implies a radical change of practice with respect to the long-term preservation and sharing of digital data, as any public archive shall be **immediately in open access** with the exception of 24 derogations applicable to certain categories of documents.
3. Most ethnographic and oral/linguistic material may be covered by a derogation (AR048) dealing with the **protection of private information**. In this case, access will be restricted to documents for which specific free consent forms have been signed by informants.
4. Art. L213-5: “Any administration [...] is compelled to **give reasons for denying access** to a public archive.”

AR043

25 years from the date of death of the individual. Documents disclosure of which would undermine medical confidentiality. (Code du Patrimoine, art. L. 213-2, I, 2)

AR044

25 years from the date of death of the person, whichever is shorter than the normal limit of 75 years. Marital status (birth or marriage). (Code du Patrimoine, art. L. 213-2, I, 4, e)

AR045

25 years from the date of death of the person, whichever is shorter than the normal period of 75 years. Minutes and inventories of public and ministerial officials. (Code du Patrimoine, art. L. 213-2, I, 4, d)

AR046

25 years from the date of death of the person, whichever is shorter than the normal period of 75 years. Documents relating to investigations conducted by the staff of the judicial police. (Code du Patrimoine, art. L. 213-2, I, 4, b)

AR047

25 years from the date of death of the person, whichever is shorter than the normal period of 75 years. Documents relating to cases before the courts, subject to special provisions relating to judgments, and execution of court decisions. (Code du Patrimoine, art. L. 213-2, I, 4, c)

AR048

50 years. Documents disclosure of which undermines the protection of privacy or for appreciation or value judgments about a person named or easily identifiable, or which reveal the behavior of a person under circumstances which might bring him/her injury. (Code du Patrimoine, art. L. 213-2, I, 3)

In this corpus, access to **source video files** is restricted to research scholars, teachers and students identified on the site and agreeing with the SLDR non-commercial licence. This restriction is legally justified by derogation code AR048 dealing with private life.



<http://sldr.org/sldr000764/toc>

- [364] [V05f.m4v](#) (1358 Mb) 2010-01-06 10:40:38
- [365] [V05m.m4v](#) (1372 Mb) 2010-01-06 10:40:38
- [366] [V07.m4v](#) (1372 Mb) 2010-01-06 10:40:38
- **SECRET**
 - [367] (no name) (154 Mb) 2010-04-14 12:24:36
 - [368] (no name) (154 Mb) 2010-04-14 12:24:36
 - [369] (no name) (154 Mb) 2010-04-14 12:24:36
- **§autorisations**
 - [371] [AnnaGhimenta.pdf](#) (1121552 Mb) 2011-04-18 08:12:47
 - [372] [GiovanniDepau.pdf](#) (1020646 Mb) 2011-04-18 08:12:57

AR048 (50 ans) – Documents dont la communication porte atteinte à la protection de la vie privée ou portant appréciation ou jugement de valeur sur une personne physique nommément désignée, ou facilement identifiable, ou qui font apparaître le comportement d'une personne dans des conditions susceptibles de lui porter préjudice. (Code du Patrimoine, art. L. 213-2, I, 3) – Jusqu'au 2061-01-26

Public access to **edited video** will be envisaged after collecting the permissions of all participants in the documentation.

Making a digital document 'public' **does not imply that it should be directly downloadable**. This status is compatible with a non-commercial transaction granting access to any person registered as a 'user'.

Corpus Représentations linguistiques Marseille 2007 - A corpus of linguistic interactions in Marseille, 2007

Cécile PETITJEAN

↳ Département de sciences du langage, Université d'Aix-Marseille (Aix-en-Provence FR)
↳ Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR) -> [source](#) - [publications](#)

Persistent Identifier: [hdl:11041/sldr000019](https://hdl.handle.net/11041/sldr000019)

SLDR/ORTOLANG id: <http://sldr.org/sldr000019>

ARK: [ark:/87895/1.4-126697](https://nbn-resolving.org/urn:nbn:fr:ark:/87895/1.4-126697)

OAI: oai.sldr.org/sldr000019 [olac](#) - [oai_dc](#) - [VLO](#) - [language-archives](#)

[Google earth](#)

[OpenStreetMap](#)

[Misc publications](#)

Example of file sharing
in high and low resolutions

[\[Edit\]](#) [\[versions\]](#) [\[Discussion\]](#) [Download version 4](#) [More](#) [Watch this item](#) [\[?\]](#)
[Download source](#)

Type of item	Primary data (corpus)
Identifier	sldr000019 (version 4/4)
Status	long-term preservation
Table of contents (copied from French)	<ul style="list-style-type: none">* Fichiers WAV* Transcription* Article de <i>Journal of Language Contact</i> – THEMA 1 (2008), p.29-51.

- AAC
 - o [1] AAC/accessRights.xml => DEPOT_DESC_accessRights1.xml
 - o [2] [ALF0207F.m4a](#) => [hdl:11041/sldr000019/ALF0207F.m4a](https://hdl.handle.net/11041/sldr000019/ALF0207F.m4a) (8 Mb) public 2011-08-10 09:28:42
 - o [3] [BEN0207M.m4a](#) => [hdl:11041/sldr000019/BEN0207M.m4a](https://hdl.handle.net/11041/sldr000019/BEN0207M.m4a) (14 Mb) public 2011-08-10 09:32:57
 - o [4] [CAY0207F.m4a](#) => [hdl:11041/sldr000019/CAY0207F.m4a](https://hdl.handle.net/11041/sldr000019/CAY0207F.m4a) (13 Mb) public 2011-08-10 09:29:11
 - o [5] [FRA1107M.m4a](#) => [hdl:11041/sldr000019/FRA1107M.m4a](https://hdl.handle.net/11041/sldr000019/FRA1107M.m4a) (13 Mb) public 2011-08-10 09:29:40
 - o [6] [ISN0107M.m4a](#) => [hdl:11041/sldr000019/ISN0107M.m4a](https://hdl.handle.net/11041/sldr000019/ISN0107M.m4a) (11 Mb) public 2011-08-10 09:30:06
 - o [7] [LES0107F.m4a](#) => [hdl:11041/sldr000019/LES0107F.m4a](https://hdl.handle.net/11041/sldr000019/LES0107F.m4a) (13 Mb) public 2011-08-10 09:30:33
 - o [8] [PET0107M.m4a](#) => [hdl:11041/sldr000019/PET0107M.m4a](https://hdl.handle.net/11041/sldr000019/PET0107M.m4a) (15 Mb) public 2011-08-10 09:31:06
 - o [9] [RIO0107M.m4a](#) => [hdl:11041/sldr000019/RIO0107M.m4a](https://hdl.handle.net/11041/sldr000019/RIO0107M.m4a) (13 Mb) public 2011-08-10 09:31:37
 - o [10] [SAN0107F.m4a](#) => [hdl:11041/sldr000019/SAN0107F.m4a](https://hdl.handle.net/11041/sldr000019/SAN0107F.m4a) (23 Mb) public 2011-08-10 09:32:25
 - o [11] [THO0207F.m4a](#) => [hdl:11041/sldr000019/THO0207F.m4a](https://hdl.handle.net/11041/sldr000019/THO0207F.m4a) (13 Mb) public 2011-08-10 09:28:19
- [12] accessRights.xml => DEPOT_DESC_accessRights2.xml
- [13] PERMANENT/accessRights.xml => DEPOT_DESC_accessRights3.xml
- [14] [JLC_Varia_1_2008_Cecile_Petitjean.pdf](#) => [hdl:11041/sldr000019/JLC_Varia_1_2008_Cecile_Petitjean.pdf](https://hdl.handle.net/11041/sldr000019/JLC_Varia_1_2008_Cecile_Petitjean.pdf) (0.309 Mb) public
- [15] [Marseille_Transcription.pdf](#) => [hdl:11041/sldr000019/Marseille_Transcription.pdf](https://hdl.handle.net/11041/sldr000019/Marseille_Transcription.pdf) (1 Mb) public 2010-03-17 16:07:46
- WAV
 - o [16] WAV/accessRights.xml => DEPOT_DESC_accessRights4.xml
 - o [17] [ALF0207F.wav](#) => [hdl:11041/sldr000019/ALF0207F.wav](https://hdl.handle.net/11041/sldr000019/ALF0207F.wav) (23 Mb) 2010-03-17 13:54:36
 - o [18] [BEN0207M.wav](#) => [hdl:11041/sldr000019/BEN0207M.wav](https://hdl.handle.net/11041/sldr000019/BEN0207M.wav) (35 Mb) 2010-03-17 13:54:44
 - o [19] [CAY0207F](#) AR048 (50 years) – Documents disclosure of which undermines the protection of privacy or for appreciation or value judgments about a person named or easily identifiable, or which reveal the behavior of a person under circumstances which might cause him/her prejudice. (Code du Patrimoine, art. L. 213-2, I, 3)
 - o [20] [FRA1107M](#) – Up to 2060-03-16 – Restricted dissemination of high-resolution file
 - o [21] [ISN0107M](#)
 - o [22] [LES0107F](#)
 - o [23] [PET0107M](#)
 - o [24] [RIO0107M](#)

Editing accessRights

File mapping :
[Follow this link](#)

Participants agree with unrestricted access to low-resolution sound files (AAC) but their sources (WAV) are reserved to registered users approving the SLDR non-commercial licence.

page 37


```

- <dc:publisher>
  Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR)
</dc:publisher>
<dc:publisher xsi:type="dcterms:URI">http://lpl-aix.fr</dc:publisher>
- <dc:creator>
  Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR)
</dc:creator>
- <dc:contributor xsi:type="olac:role" olac:code="depositor">
  Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR)
</dc:contributor>
<dc:contributor xsi:type="dcterms:URI">http://lpl-aix.fr</dc:contributor>
- <dcterms:accessRights xml:lang="en">
  SLDR licence; rightsHolder = Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence F
</dcterms:accessRights>
<dcterms:accessRights xml:lang="en">Privileged user: [2065] CID-user</dcterms:accessRights>
<dcterms:license xsi:type="dcterms:URI">http://sldr.org/licence_v1/en</dcterms:license>
<dcterms:license xsi:type="dcterms:URI">http://sldr.org/licence_v1/es</dcterms:license>
<dcterms:license xsi:type="dcterms:URI">http://sldr.org/licence_v1/fr</dcterms:license>
<dcterms:license xsi:type="dcterms:URI">http://sldr.org/licence_v1/zh</dcterms:license>
<dcterms:provenance xml:lang="en">long-term preservation</dcterms:provenance>
<dcterms:provenance xml:lang="es">archivo</dcterms:provenance>
<dcterms:provenance xml:lang="fr">archive pérenne</dcterms:provenance>
<dcterms:provenance xml:lang="zh">长期档案</dcterms:provenance>
- <dcterms:accessRights xml:lang="fr">
  Restriction AR048 (50 ans à partir de 2008-05-02) - Documents dont la communication porte atteinte à la protection de la vie privée ou portant
  appréciation ou jugement de valeur sur une personne physique nommément désignée, ou facilement identifiable, ou qui font apparaître le
  comportement d'une personne dans des conditions susceptibles de lui porter préjudice. (Code du Patrimoine, art. L. 213-2, I, 3)
</dcterms:accessRights>
- <dcterms:accessRights xml:lang="en">
  Restriction AR048 (50 years from 2008-05-02) - Documents disclosure of which undermines the protection of privacy or for appreciation or value
  judgments about a person named or easily identifiable, or which reveal the behavior of a person under circumstances which might cause him/her
  prejudice. (Code du Patrimoine, art. L. 213-2, I, 3)
</dcterms:accessRights>
- <dcterms:accessRights xml:lang="zh">
  制约 AR048 (从2008-05-02准入限制组50年) - 提供破坏隐私保护或欣赏或关于容易辨认的人的价值判断的命名或, 或者在情况也许带来他或她的伤害下显
  露人行为的透露。 (Code du Patrimoine, 艺术。 L. 213-2, I, 3)
</dcterms:accessRights>
- <dcterms:accessRights xml:lang="es">
  Restricción AR048 (50 years from 2008-05-02) - Documentos de divulgación de lo que perjudica la protección de la intimidad o de los juicios de valor
  acerca de apreciación o una persona con nombre o fácilmente identificables, o que revelan el comportamiento de una persona en circunstancias que
  podrían llevarle lesión. (Code du Patrimoine, art. L. 213-2, I, 3)
</dcterms:accessRights>
<dcterms:temporal xsi:type="dcterms:Period" start="2002" end="2003"></dcterms:temporal>
<dcterms:spatial xsi:type="dcterms:ISO3166">FR</dcterms:spatial>
<dcterms:spatial xsi:type="dcterms:Point">name=Aix-en-Provence; east=5.448; north=43.516</dcterms:spatial>
<dcterms:spatial xsi:type="dcterms:TGN">7010786</dcterms:spatial>
<dc:subject xsi:type="olac:linguistic-field" olac:code="phonology"/>

```

Access rights to documents are displayed in the descriptive metadata of an interoperable system.

Users communities



Following up resource usage: users' communities

Downloaded (32) secondary data (resource) **VfrLPL - hdl:11041/sldr000533**

First name and family name	Institution	Field of research	Date of download	Version
M Stéphane RAUZY Contact	Ingénieur de Recherche Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR)	Informatique et Linguistique	2007-05-21 licence #0	1
Mme Morgane ADER Contact	Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR)	informatique appliqué à la linguistique	2007-05-23 licence #0	1
M Stéphane RAUZY Contact	Ingénieur de Recherche Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR)	Informatique et Linguistique	2013-12-02 licence #1	1
M Cedric GENDROT Contact	Institut de linguistique et de phonétique générales et appliquées (ILPGA, Paris FR)	Phonétique	2007-06-06 licence #0	1
M Bruno GUILLAUME Contact	Laboratoire lorrain de recherche en informatique et ses applications - UMR 7503 (Loria, Nancy FR)	TAL	2007-06-06 licence #0	1
M Yoshiyuki FUKUSHIMA Contact	Osaka University, Japan	linguistique française, analyse de la communication, didactique du français	2007-06-17 licence #0	1
M Matthieu HERMET Contact	Universite d'Ottawa - EITI http://www.site.uottawa.ca/index.shtml	CALL, Informatique	2007-07-20 licence #0	1

This page displays the downloadings of an item and profiles of its users.

Access to this list is restricted to persons having produced items on SLDR or the ones who downloaded this particular item.