

Mutualisation et archivage de données de la recherche : l'expérience du SLDR

<http://sldr.org/doc/show/SLDR-PR2I.pdf>

Bernard Bel
Laboratoire Parole et Langage (LPL)
CNRS – Université Aix-Marseille

Journée PR2I : Les Humanités Numériques à AMU
Aix-en-Provence, le 26 novembre 2015



ORTOLANG bénéficie d'une aide de l'Etat au titre du
programme « Investissements d'avenir » (ANR-11-EQPX-0032)



Qui nous sommes...

La Banque de données parole et langage, *Speech & Language Data Repository* (SLDR – www.sldr.org) est un **service générique** de **mutualisation et archivage** de données numériques créé par le Laboratoire Parole et Langage (LPL, UMR 7309 CNRS-AMU).

L'objectif du SLDR était de regrouper les ressources linguistiques associées à la communication orale/multimodale. Une telle centralisation permet tout d'abord d'éviter que les laboratoires travaillant dans ce domaine ne soient conduits à recréer sans cesse des données. La mutualisation de ressources en sciences du langage répond par ailleurs au besoin stratégique de **rapprocher les connaissances** aujourd'hui dispersées dans des domaines variés comme la linguistique descriptive, formelle et computationnelle, la sociolinguistique, la littérature, la traductologie, les neurosciences, la psycholinguistique etc.

Le SLDR est ouvert aux projets de recherche et aux chercheurs du **monde entier**. Son interface de navigation et de dépôt est en **quatre langues** : français/anglais/espagnol/chinois.

Fin 2015, le SLDR est en voie d'achever son **intégration à la plateforme ORTOLANG** (*Open Resources and TOols for LANGuage*) qui offrira de nouveaux outils pour le traitement des données écrites/orales. [ORTOLANG](#) est un service spécialisé pour la langue, complémentaire de l'offre générale proposée par la TGIR [Huma-Num](#), visant une **interopérabilité** avec les réseaux européens [CLARIN](#) (linguistique) et [DARIAH](#) (*Digital Humanities*).

L'architecture du site est **générique**, basée sur un modèle ouvert de système d'information pour l'archivage (**OAIS**, *Open Archival Information System*). Le site garantit un accès permanent aux objets par le biais d'**identifiants pérennes**. La plateforme de dépôt et de diffusion est hébergée par l'[INIST](#). Les dépôts en archive pérenne sont soumis au [CINES](#) selon la procédure mise en place par [Huma-Num](#).

Archiver : pourquoi et comment ?

Pourquoi archiver ?

(Source : [CINES](#))



- L'archivage numérique n'est **pas** une simple **sauvegarde**.
- Motivation première : **préserver** des données
- Motivation seconde : donner la possibilité de les **localiser** et de les **réutiliser** dans un avenir indéterminé. C'est l'objet de l'archivage pérenne (*long-term preservation*).
- L'archivage numérique pérenne **n'est pas l'ultime étape du stockage des données avant l'oubli ou la perte définitive !**
- Trois objectifs principaux :
 - 1) **conserver** le document,
 - 2) le rendre **accessible**,
 - 3) en préserver **l'intelligibilité**.

Comment archiver ?

- Ces objectifs sont inscrits sur le **très long terme**, c'est-à-dire plus de 30 ans.
- Pour ces raisons il faut confier les données à une **archive institutionnelle** — le [CINES](#) — plutôt qu'à un consortium de centres informatiques.
- Ceci entraîne des contraintes **techniques** et **juridiques**.

Partager et archiver : quoi ?

- Système **générique** : **pas de structure imposée** aux objets déposés sur le SLDR. Le plan de classement est déterminé par les métadonnées et les identifiants pérennes des documents.
- Les objets peuvent être en cours d'élaboration : le SLDR diffuse des **données courantes** et des **archives intermédiaires**.
- Les types d'objets en linguistique (et disciplines apparentées) :
 - Données primaires** : les corpus de signaux associés à la production langagière (audio/vidéo/données articulatoires) ainsi que les documents créés ou collectés lors de l'expérimentation/enquête ;
 - Données secondaires** : les ressources fabriquées à partir de données primaires (transcriptions, traductions, annotations, analyses...), ainsi que des objets facilitant l'analyse : lexiques, grammaires, tables de fréquences etc.
 - Outils** : logiciels et descriptions de matériels utiles à l'analyse et l'enrichissement des données.
 - Collections** : un ensemble d'objets des types précités – y compris collections de collections...

[ARK] **Données primaires (corpus) Derrida - Artaud** (Bernard BEL)
Centre de sciences humaines (CSH, New Delhi IN)
Conférence de Jacques Derrida sur Antonin Artaud. Séminaire *In homage to Antonin Artaud* organisé à New Delhi par National School of Drama, Delhi University et l'Ambassade de France, le 23 janvier 1997. [Détails]
(philosophy_of_language)
anglais (English)

embargo -> 2047

[ARK] **Outil Anonymisation de fichiers sonores** (Daniel HIRST)
Laboratoire parole et langage (LPL, Aix-en-Provence FR) -> source
Script PRAAT.
Objet : remplacer les fragments d'un "long sound" étiquetés par un mot-clé par un son murmuré dont les caractéristiques prosodiques sont identiques à celles du son original.
Le fichier source peut être mono ou stéréo, sans incidence sur le résultat. [Détails]
(applied_linguistics, cognitive_science, language_documentation, spr
computational_linguistics)
>> Collection **Outils LPL crdo000763**

public - licence Creative Commons

[ARK] **Ressource Valjouffrey - travaux** (Audrey THOMAS, Clément GIRARD, Médéric GASQUET-CYRUS)
Laboratoire parole et langage (LPL, Aix-en-Provence FR) -> source
Département de linguistique et phonétique générales, Université de Provence (Aix-en-Provence FR)
1) Traductions, transcriptions et annotations des corpus de patois de Valjouffrey
2) Fichiers sonores, vidéos et annotations/sous-titrage accessibles publiquement pour le streaming [Détails]
(language_documentation, lexicography, sociolinguistics, anthropological_linguistics, text_and_corpus_linguistics, phonology, phonetics)
occitan (post 1500); occitan (post 1500) -> patois de Valjouffrey
>> Collection **Valjouffrey crdo000007**

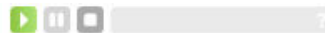
libre accès

[ARK] **Données primaires (corpus) PSH/DISPE - Parole subaquatique et/ou hyperbare** (Alain MARCHAL)
Laboratoire parole et langage (LPL, Aix-en-Provence FR) -> source
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. [Détails]
(phonetics, phonology)
anglais; français

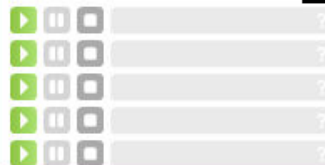
licence non commerciale

[ARK] **Données primaires (corpus) Mờ Piu** (Geneviève CAELEN-HAUMONT)
Multimedia, Information, Communication and Applications (MICA, Hanoi VN)
Données collectées dans le cadre d'un premier terrain à **Nậm Tu Thự ợng** chez les **Mờ Piu** entre le **8 et 12 Juin 2009**. Les corpus parole, chants, morceaux de musique, enregistrés sont répertoriés par jour, thème, locuteur, question (chant ou morceau instrumental de musique). **A chaque couple question/réponse correspond un fichier wav**. Les enregistrements ont été effectués en stéréo. Sur une piste le jour, le thème, le locuteur et le numéro de la question sont rappelés, suivis par une question en vietnamien (posée par un Vietnamien). Sur la seconde piste, en Mờ Piu, la même question est ensuite posée suivie de la réponse.
4 locuteurs, 3 locutrices, 7 chanteuses et 2 chanteurs ont été enregistrés. Soit au total, **7 heures de paroles et 1 heure de chants**. [Détails]
(anthropological_linguistics)
vietnamien -> mờ piu

en cours



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

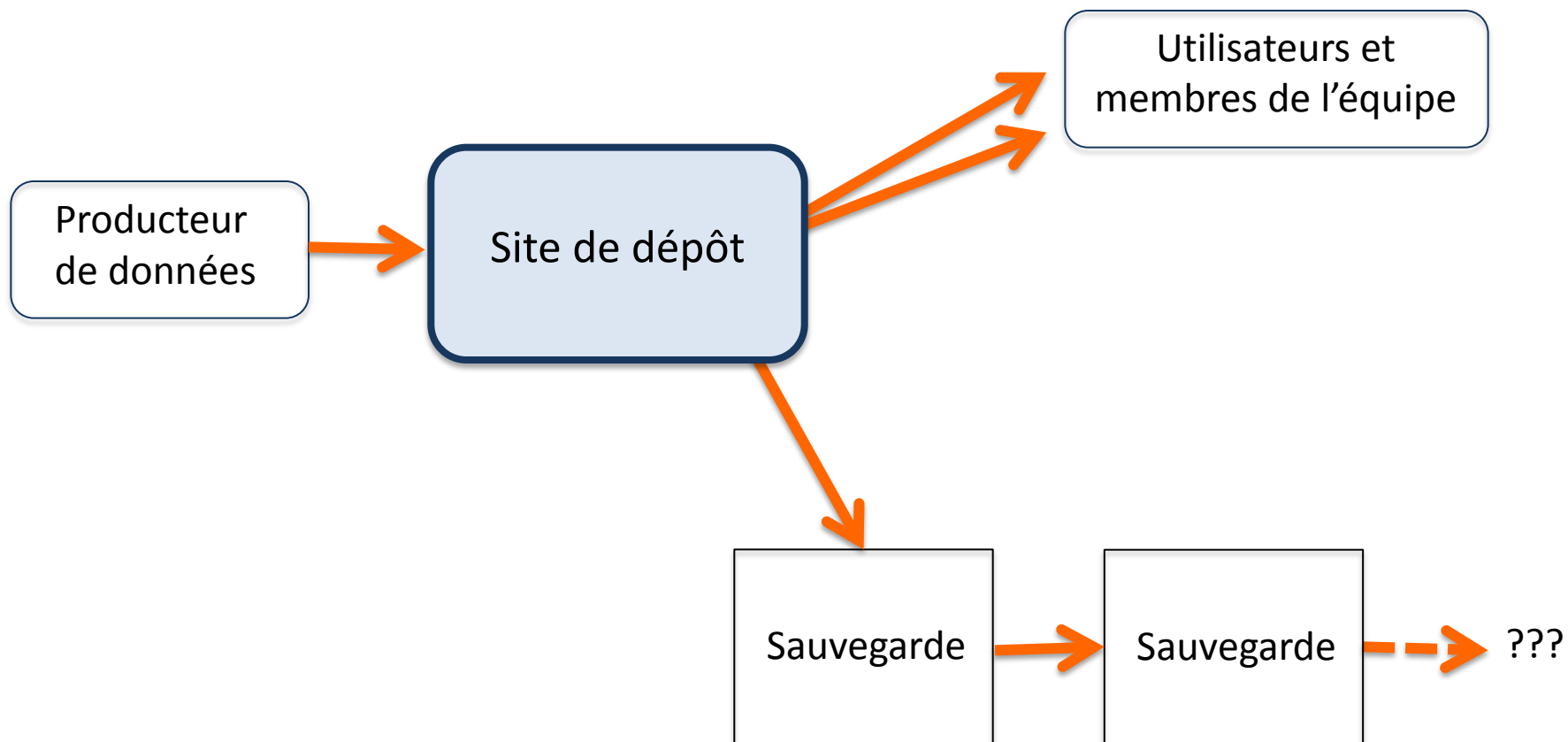


Rapport.pdf



Les ressources du SLDR comprennent des **corpus** de parole ou de chant, leurs **enrichissements**, des lexiques et autres **bases de connaissances**, ainsi que des **outils** adaptés au traitement de ces données. Les corpus peuvent être constitués d'enregistrements **audio/vidéo** mais aussi de **mesures** de l'activité physiologique.

Solution classique pour la mutualisation de données numériques



Les contraintes techniques de l'archivage pérenne

- Préserver la **lisibilité d'un document** :
Nombre limité de **formats durables** dont les spécifications sont librement accessibles.
Migration systématique des formats au fur et à mesure de leur obsolescence : *c'est le travail des archivistes, pas des déposants !*
- Éviter la perte de la **signification du contenu** :
 1. Les métadonnées **descriptives** (« métier ») ;
 2. Les métadonnées **d'archivage** ;
 3. La description formelle des **conventions** utilisées dans l'archive.
PPDI = *Project Preservation Description Information*.
<http://sldr.org/ppdi>

Deux problèmes résolus par le
projet pilote TGE-
Adonis/CINES/IN2P3/CRDO
(2008-2010)

Pb 1 : Comment modifier des objets archivés ?

Problème

- Un objet contient à la fois des **données figées** (ex. les données primaires son/vidéo) et des **données documentaires** (métadonnées descriptives, annotations, conditions d'accès...) modifiables à tout moment.
- Une archive institutionnelle doit préserver **toutes les versions** des paquets d'information qui lui ont été confiées.
- Il serait regrettable de redéposer l'intégralité d'un objet après chaque modification de données documentaires.

Solution élaborée avec le CINES

- Chaque *Submission Information Package* (SIP) contient deux répertoires de données archivables : l'un (**DEPOT**) que nous utilisons pour les données figées et le second (**DEPOT/DESC**) les données documentaires.
- Il est possible de déposer une nouvelle version, soit du paquet entier, soit de la partie documentaire — on parle alors de « mise à jour de métadonnées ».

Pb 2 : Que faire des documents non-archivables ?

Problème

- Certains formats de fichiers (ex. MP3, FLV, ZIP) sont **incompatibles** avec l'archivage pérenne mais requis pour certains usages : *streaming* audio ou vidéo, téléchargement de répertoires entiers etc.
- Dans le modèle OAIS du Projet pilote, toutes les données étaient déposées sur le site d'archivage où elles étaient validées, archivées **puis transmises au site de diffusion**.

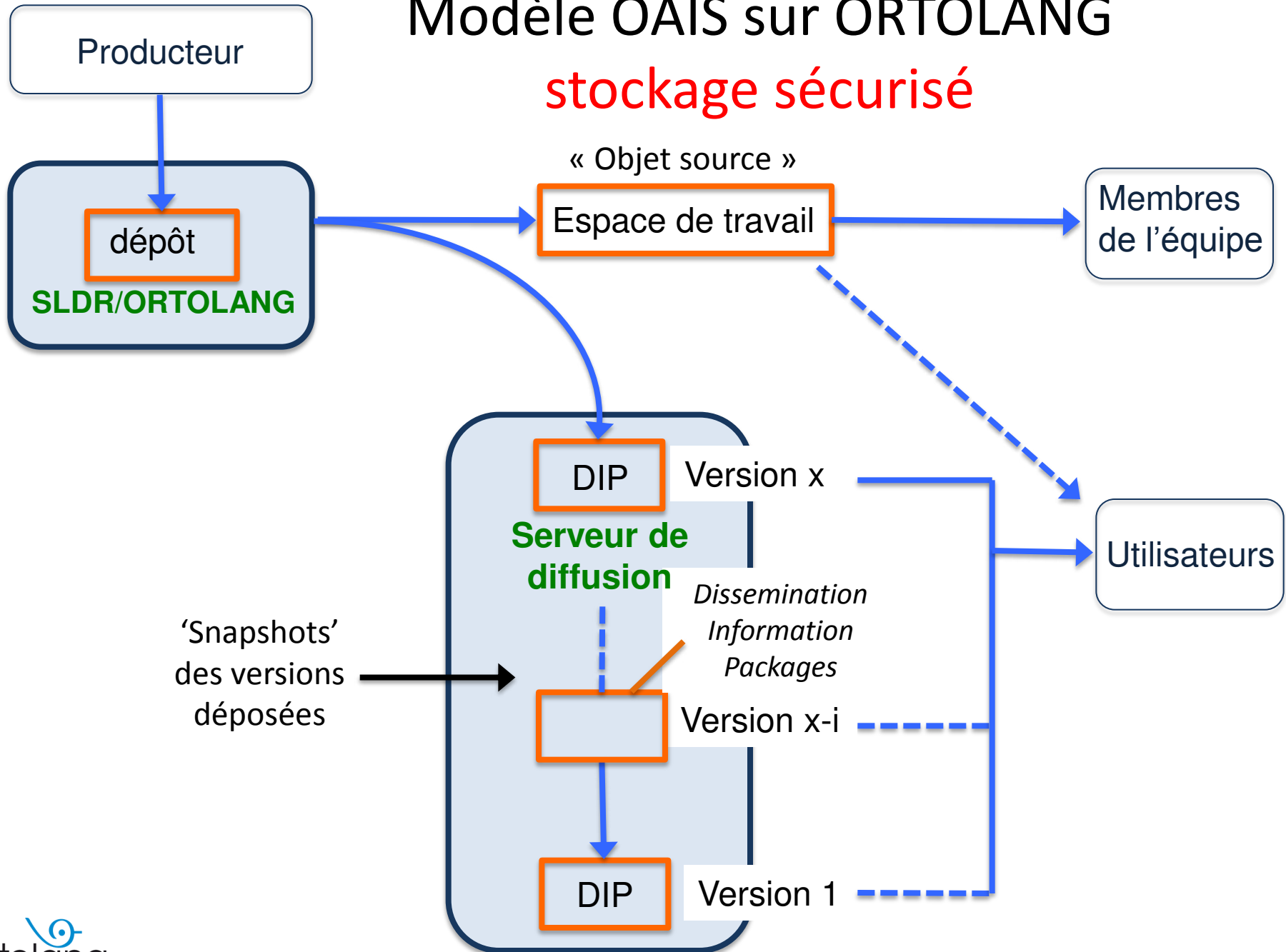
Solution convenue avec le CINES

- Chaque SIP contient un répertoire **DIFFUSION** dont le contenu est ignoré par le service d'archivage et transmis au service de diffusion à l'intérieur du *Dissemination Information Package* (DIP).
- Ce répertoire peut aussi contenir des documents destinés à un 'archivage à court terme' parce que susceptibles d'être modifiés, ou encodés dans des formats pas (encore) reconnus par le service d'archivage.

Mutualisation et
archivage : le modèle
OAS selon
ORTOLANG

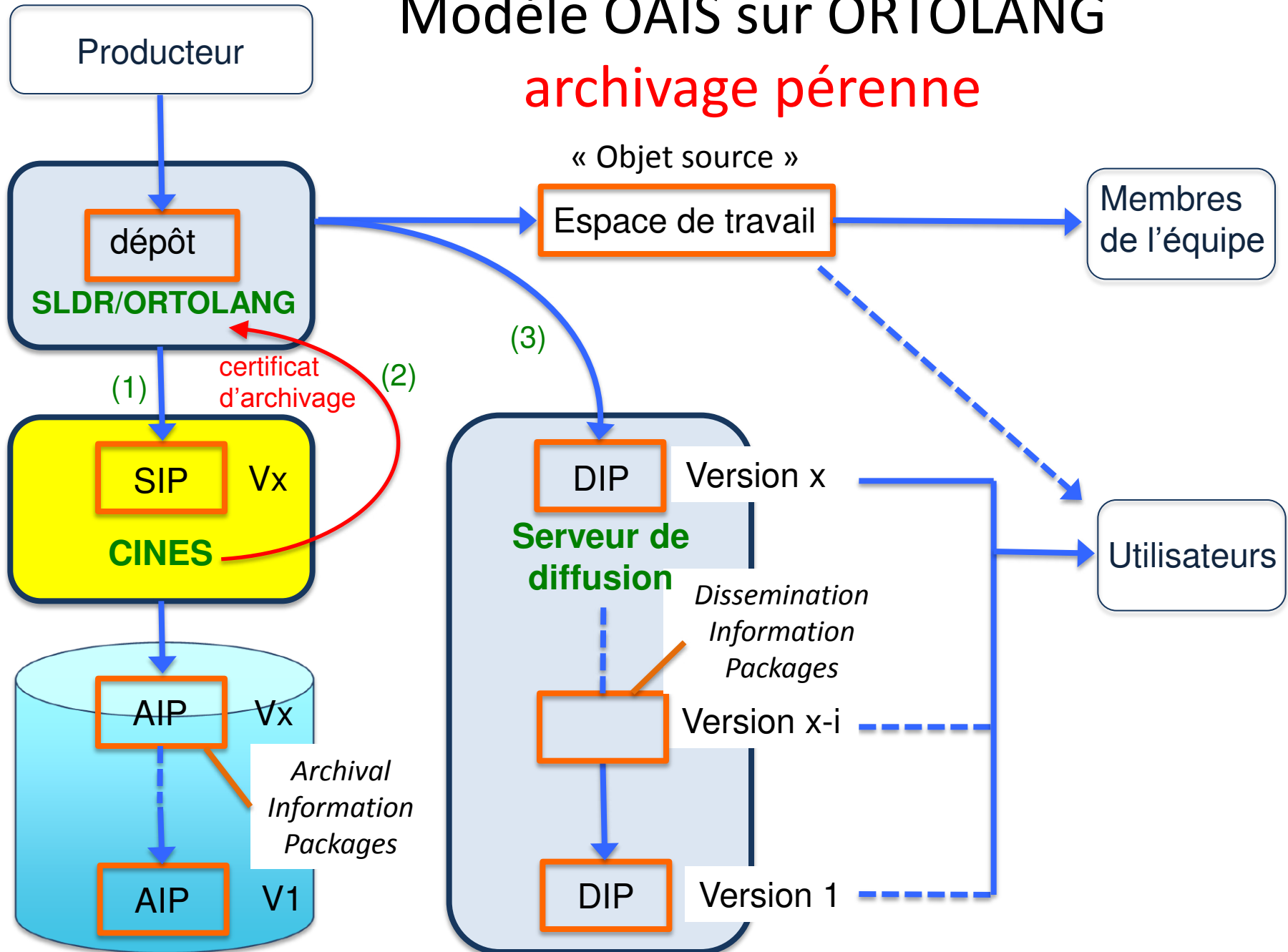
Modèle OAIS sur ORTOLANG

stockage sécurisé



Modèle OAIS sur ORTOLANG

archivage pérenne



Accès aux données

Identifiants pérennes (PID, *Persistent IDentifiers*)

- La **granularité** de l'identification des ressources est variable : chaque objet reçoit un **identifiant pérenne** (*persistent identifier*, PID), ainsi que – par défaut – tous les documents qu'il contient.
- Les objets archivés au CINES reçoivent un identifiant [ARK](#) (*Archival Resource Key*).
- SLDR/ORTOLANG utilise le *Handle system* (www.handle.net) et construit les identifiants de manière **déterministe**, exemple :
 - [hdl:11041/swedia-000788](http://hdl.handle.net/11041/swedia-000788)
(<http://hdl.handle.net/11041/swedia-000788>)
—> corpus *swedia-000788* localisé à sldr.org/swedia-000788
 - [hdl:11041/alipe-000853/ali-prune-071122.mp3](http://hdl.handle.net/11041/alipe-000853/ali-prune-071122.mp3)
—> fichier 'ali-prune-071122.mp3' appartenant au corpus *alipe-000853* localisé à sldr.org/alipe-000853

Voir détails techniques : sldr.org/wiki/Handle_fr
et article : <http://tinyurl.com/n29wc9j>

Données brutes et sites thématiques

- SLDR/ORTOLANG stocke et partage des **données brutes** (*source data*) sous la forme de « **paquets d'information** » (*information packages*) qui forment des arborescences sans limite de complexité ni de volumétrie.
- Ces données peuvent être exploitées selon plusieurs modalités :
 1. Téléchargement
 2. Lecture sur un navigateur web
 3. Streaming audio ou vidéo
 4. Traitement par un logiciel externe
- SLDR ne gère que la mise à disposition des données (en intégrant les droits d'accès). Son interface est minimale.
- Il est possible de construire des **sites thématiques** exploitant les données de SLDR avec une interopérabilité maximale grâce à l'utilisation d'**identifiants pérennes** (PIDs). Ces sites utilisent leurs propres interfaces et systèmes de requêtes : pages statiques, base de données etc.
- Exemple de site thématique multilingue : <http://grindmill.org>
Les fichiers MP3 sont exploités en *streaming* et les images sont affichées. La base de données MySQL est externe, mais son export au format CSV est archivé sur SLDR.

Conditions d'accès

- Les objets de SLDR/ORTOLANG (archives publiques) sont accessibles sur la base de **transactions non-commerciales**. Des liens sont fournis pour accéder à d'autres sites (ELRA, LDC...) dans le cas d'archives privées, exemple : <http://hdl.handle.net/11041/sldr000035>.
- Le téléchargement d'une ressource est conditionné par l'**acceptation d'une licence** (voir sldr.org/wiki/Licences_fr) ainsi que, le cas échéant, d'une licence particulière pour cette ressource.
- Certains objets peuvent être diffusés de pair à pair selon les dispositions d'une licence **Creative Commons**. Si l'on évite les options 'nc' et 'nd' ils bénéficient du label *Open Data* : <http://opendefinition.org>.
- L'accès à un objet peut être réservé à une catégorie d'utilisateurs (ex. enseignants/chercheurs/étudiants) mais il peut être étendu à des institutions ou des individus désignés par le biais d'**accès privilégiés**. Voir sldr.org/wiki/PrivilegedAccess.
- La plateforme SLDR privilégie la **communication entre producteurs et utilisateurs** de ressources : suivi des mises à jour, commentaires sur les objets, profils d'utilisateurs détaillés (liens avec réseaux sociaux), communication entre pairs, newsletter...

Communauté d'utilisateurs (communication entre pairs)

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

Téléchargements d'un objet et
profils des utilisateurs

*Cette liste n'est accessible qu'aux
utilisateurs ayant eux-mêmes déposé
des objets et/ou téléchargé cet objet.*

Moissonnage du serveur OAI-PMH

Les métadonnées de SLDR sont moissonnées par des **portails** et **moteurs de recherche**...

1. **spécifiques** de la linguistique :

- *Language-Archives* (OLAC) (<http://www.language-archives.org>)
- *Virtual Language Observatory* (CLARIN) (<https://vlo.clarin.eu>)

2. **nationaux** :

- ISIDORE de Huma-Num (<http://www.rechercheisidore.fr>)
- Université de Shanghai

3. **internationaux** :

- SHARE (<https://osf.io/share/registration>) - *en cours*
- *Open Archives* OAI (<http://www.openarchives.org>)
- Google (<http://google.com>)
- etc...

Gestion intégrée
des droits d'accès

Le Code du patrimoine

- Les données du SLDR/ORTOLANG sont versées pour leur archivage pérenne dans un centre d'archivage **institutionnel** (le [CINES](#)). Les conditions d'accès doivent donc respecter les récentes dispositions (juillet 2008) du [Code du patrimoine](#) régissant le statut d'**archive publique** et les droits d'accès aux documents archivés (loi du 15 juillet 2008, articles L213 1-5).
- Ce dispositif a induit un changement radical de pratique pour ce qui concerne l'archivage des données numériques de la recherche : toute archive publique est **(immédiatement) communicable** de plein droit, dans la limite de 24 dérogations applicables à certaines catégories de documents (voir sldr.org/wiki/table_derogations_fr).
- La plupart des données orales/multimodales font l'objet de la dérogation 'AR048' protégeant les données personnelles. Dans cette configuration, l'accès aux documents peut être réservé pendant 50 ans aux personnes ou institutions bénéficiant d'**autorisations** signées par les participants.
- L'article L213-5 oblige SLDR/ORTOLANG à **motiver tout refus** opposé à une demande de communication de documents d'archives.

Déroptions au principe de libre-communicabilité

Par dérogation aux dispositions de l'article L. 213-1 :

I.-Les archives publiques sont communicables de plein droit à l'expiration d'un délai de :

1° Vingt-cinq ans à compter de la date du document ou du document le plus récent inclus dans le dossier :

...

2° Vingt-cinq ans à compter de la date du décès de l'intéressé, pour les documents dont la communication porte atteinte au **secret médical** (AR043). Si la date du décès n'est pas connue, le délai est de cent vingt ans à compter de la date de naissance de la personne en cause (AR061) ;

3° Cinquante ans à compter de la date du document ou du document le plus récent inclus dans le dossier, pour les documents dont la communication porte atteinte au **secret de la défense nationale**, aux intérêts fondamentaux de l'Etat dans la conduite de la politique extérieure, à la **sûreté de l'Etat**, à la **sécurité publique**, à la **sécurité des personnes** (AR049) ou à la **protection de la vie privée** (AR048), à l'exception des documents mentionnés aux 4° et 5°. Le même délai s'applique aux documents qui portent une appréciation ou un 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 (AR048).

...

4° Soixante-quinze ans à compter de la date du document ou du document le plus récent inclus dans le dossier, ou un délai de vingt-cinq ans à compter de la date du décès de l'intéressé si ce dernier délai est plus bref :

a) Pour les documents dont la communication porte atteinte au **secret en matière de statistiques** lorsque sont en cause des données collectées au moyen de questionnaires ayant trait aux faits et comportements d'ordre privé (AR041, AR052) ;

b) Pour les documents relatifs aux enquêtes réalisées par les services de la **police judiciaire** (AR046, AR056) ;


...

5° Cent ans à compter de la date du document ou du document le plus récent inclus dans le dossier (AR058), ou un délai de vingt-cinq ans à compter de la date du décès de l'intéressé si ce dernier délai est plus bref, pour les documents mentionnés au 4° qui se rapportent à une **personne mineure** (AR041, AR046, AR047, AR045).

... etc.

Dérogations au principe de libre-communicabilité ([version SIAF](#))

Element `clmDAFAccessRestrictionCode:AccessRestrictionCode`

Namespace	urn:un:unece:unefact:codelist:draft:DAF:accessRestrictionCode:2009-08-18	
Diagram		
Type	<code>clmDAFAccessRestrictionCode:AccessRestrictionCodeType</code>	
Properties	Content:	simple
Facets	Enumeration	AR038 0 an Documents administratifs librement communicables. (Code du Patrimoine, art. L. 213-1)
	Enumeration	AR039 25 ans Documents dont la communication porte atteinte au secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif, à la conduite des relations extérieures, à la monnaie et au crédit public, au secret en matière commerciale et industrielle, à la recherche par les services compétents des infractions fiscales et douanières. (Code du Patrimoine, art. L. 213-2, I, 1, a)
	Enumeration	AR040 25 ans Documents qui ne sont pas considérés comme administratifs et mentionnés au dernier alinéa de l'article 1er de la loi n° 78-753 du 17 juillet 1978 : actes et documents élaborés ou détenus par les assemblées parlementaires, avis du Conseil d'Etat et des juridictions administratives, documents de la Cour des comptes mentionnés à l'article L. 140-9 du code des juridictions financières et les documents des chambres régionales des comptes mentionnés à l'article L. 241-6 du même code, documents d'instruction des réclamations adressées au Médiateur de la République, documents préalables à l'élaboration du rapport d'accréditation des établissements de santé prévu à l'article L. 6113-6 du code de la santé publique, rapports d'audit des établissements de santé mentionnés à l'article 40 de la loi de financement de la sécurité sociale pour 2001 (n° 2000-1257 du 23 décembre 2000). (Code du Patrimoine, art. L. 213-2, I, 1, b)
	Enumeration	AR041 25 ans Secret statistique. (Code du Patrimoine, art. L. 213-2, I, 1, a)
	Enumeration	AR042 25 ans Documents élaborés dans le cadre d'un contrat de prestation de services exécuté pour le compte d'une ou de plusieurs personnes déterminées. (Code du Patrimoine, art. L. 213-2, I, 1, b)
	Enumeration	AR043 25 ans à compter de la date de décès de l'intéressé Documents dont la communication est susceptible de porter atteinte au secret médical. (Code du Patrimoine, art. L. 213-2, I, 2)
	Enumeration	AR044 25 ans à compter de la date de décès de l'intéressé si ce délai est plus court que le délai ordinaire de 75 ans Etat civil (actes de naissance ou de mariage). (Code du Patrimoine, art. L. 213-2, I, 4, e)

... etc. (25 cas)



[table derogations es](#)

AR038

Documentos libremente comunicables. (Code du Patrimoine, art. L. 211-1, L. 211-4, L. 213-1)

AR039

25 años. Documentos de la divulgación de que viola la confidencialidad de las deliberaciones del Gobierno y las autoridades del Poder Ejecutivo para llevar a cabo las relaciones exteriores, moneda y público crédito, comercial e industrial secreto, la investigación de departamentos pertinentes de los delitos aduaneros y fiscales. (Code du Patrimoine, art. L. 213-2, I, 1 a)

AR040

25 años. Documentos que no se consideran administrativos y mencionados en el último párrafo del artículo 1 de la Ley N ° 78-753 del 17 de julio de 1978: el procedimiento y los documentos preparados o conservados por las asambleas parlamentarias, el aviso del Consejo de Estado y los tribunales administrativos, documentos de la corte y de las cuentas mencionadas en el artículo L. 140-9 del Código de las jurisdicciones y documentos financieros de las cámaras regionales de cuentas a que se refiere en el artículo L. 241-6 del Código, los documentos de investigación de las denuncias al Defensor del Pueblo de la República, los documentos antes de preparar el informe para la acreditación de los centros de salud previstos en el artículo L. 6113-6 del Código de Salud Pública, los informes de auditoría de los establecimientos de salud mencionados en el artículo 40 de la Ley de Finanzas de la Seguridad Social de 2001 (N ° 2000-1257, de 23 diciembre de 2000). (Code du Patrimoine, art. L. 213-2, I, 1, b)

AR041

25 años. Secreto de las estadísticas. (Code du Patrimoine, art. L. 213-2, I, 1 a)

AR042

25 años. Documentos elaborados en virtud de un contrato de prestación de los servicios prestados en nombre de una o varias personas determinadas. (Code du Patrimoine, art. L. 213-2, I, 1, b)

AR043

25 años a partir de la fecha de la muerte de los documentos individuales cuya divulgación perjudique a la confidencialidad del secreto medical. (Code du Patrimoine, art. L. 213-2, I, 2)

AR044

25 años a partir de la fecha de la muerte de la persona, lo que es más corto que el límite normal de 75 años Estado civil (nacimiento o matrimonio). (Code du Patrimoine, art. L. 213-2, I, 4, e)

Les dérogations sont affichées dans les quatre langues de navigation du site.

Gestion intégrée des droits d'accès : procédure

Catégorie de dérogation au principe de libre communicabilité des archives publiques (voir documentation)	1	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) À partir de 2011-01-26
Autorisations d'accès	2	HubertBalmes.pdf JulienGaillard.pdf RobertBois.pdf
	3	Jusqu'au 2031-01-26
Licence	4	Licence particulière pour cet objet (Suivre ce lien)

1. Le refus d'accès (dérogation au principe de libre-communicabilité) doit être **explicitement motivé**.
2. Les ayant-droit peuvent signer des **autorisations** rendant possible l'accès aux documents avant l'expiration de leur période de protection.
3. Ces autorisations peuvent être accordées pour une **durée limitée**.
4. Ces autorisations peuvent comporter des **conditions particulières**.

Dans ce corpus, l'accès aux **vidéos sources** est réservé aux chercheurs, enseignants ou étudiants identifiés qui ont approuvé la licence SLDR. Cette restriction est ici justifiée par une dérogation (code AR048, 50 ans) au principe de libre diffusion.

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



○ [364] [V05f.m4v](#) (1358 Mb) 2010-01-06 10:40:38

○ [365] [V05m.m4v](#) (1372 Mb) 2010-01-06 10:51:31

○ [366] [V07.m4v](#)

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

• **SECRET**

○ [367] (no name)

○ [368] (no name)

○ [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

L'accès public aux **vidéos éditées** peut être envisagé une fois obtenues les autorisations des personnes filmées.

Attention : « public » signifie **l'absence de restrictions liées au statut de l'utilisateur**. Toutefois, la visualisation ou le téléchargement peuvent faire l'objet d'une transaction non-commerciale — identification, acceptation de licence(s) etc.

En général, les **travaux universitaires** sont en libre accès car aucune dérogation n'est applicable. Aucune transaction n'est exigée.

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

- [391]  P65.c (14005 Mb) **public** 2011-05-30 12:09:44
- [392]  P65.p (40304 Mb) **public** 2011-05-30 12:09:14

○ **DIVERS**

- [395]  trim. (78 Mb) **public** 2010-06-22 08:00:02
- [396]  [Master2AudreyThomas.pdf](#) (2486820 Mb) **public** 2011-06-22 09:00:26
- [397]  [VideoCassettes000764.t](#) (662 Mb) **public** 2011-07-10 08:15:10

Documents librement communicables. (Code du Patrimoine, art. L. 211-1, L. 211-4, L. 213-1)

○ **TEXTES**

■ **P59-60**

- [398] Fables patois graphie RB modif AT.odt (12734 Mb) 2011-02-14 17:08:32
- [399] Fables patois graphie RB modif AT.pdf (53446 Mb) 2011-02-14 17:32:58
- [400] Liste_de_mots_a_eliciter.odt (9705 Mb) 2011-02-14 17:16:20
- [401] Liste_de_mots_a_eliciter.pdf (23994 Mb) 2011-02-14 17:33:24
- **phonologie**
 - [402] 1.jpg (611118 Mb) 2011-02-18 09:52:56
 - [403] 2.jpg (549067 Mb) 2011-02-18 09:52:50
 - [404] 3.jpg (313716 Mb) 2011-02-18 09:52:53



= icône servant à indiquer que le lien est en libre accès.

« Haute » versus « basse » résolution

• AAC

- [1] AAC/accessRights.xml => DEPOT_DESC_accessRights1.xml
- [2] [ALF0207F.m4a](#) (8 Mb) **public** 2011-08-10 09:28:42
- [3] [BEN0207M.m4a](#) (14 Mb) **public** 2011-08-10 09:32:57
- [4] [CAY0207F.m4a](#) (13 Mb) **public** 2011-08-10 09:29:11
- [5] [FRA1107M.m4a](#) (13 Mb) **public** 2011-08-10 09:29:40
- [6] [ISN0107M.m4a](#) (11 Mb) **public** 2011-08-10 09:30:06
- [7] [LES0107F.m4a](#) (13 Mb) **public** 2011-08-10 09:30:33
- [8] [PET0107M.m4a](#) (15 Mb) **public** 2011-08-10 09:31:06
- [9] [RIO0107M.m4a](#) (13 Mb) **public** 2011-08-10 09:31:37
- [10] [SAN0107F.m4a](#) (23 Mb) **public** 2011-08-10 09:32:25
- [11] [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](#) (0.309 Mb) **public** 2010-03-22 12:23:14
- [15] [Marseille_Transcription.pdf](#) (1 Mb) **public** 2010-03-17 16:07:46

<http://sldr.org/sldr000019/toc>

• WAV

- [16] WAV/accessRights.xml => DEPOT_DESC_accessRights4.xml
- [17] [ALF0207F.wav](#) (23 Mb) 2010-03-17 13:54:36
- [18] [BEN0207M.wav](#) (25 Mb) 2010-03-17 13:54:44
- [19] [CAY0207F.wav](#) (23 Mb) 2010-03-17 13:54:52
- [20] [FRA1107M.wav](#) (23 Mb) 2010-03-17 13:55:00
- [21] [ISN0107M.wav](#) (23 Mb) 2010-03-17 13:55:08
- [22] [LES0107F.wav](#) (23 Mb) 2010-03-17 13:55:16
- [23] [PET0107M.wav](#) (23 Mb) 2010-03-17 13:55:24
- [24] [RIO0107M.wav](#) (23 Mb) 2010-03-17 13:55:32
- [25] [SAN0107F.wav](#) (23 Mb) 2010-03-17 13:55:40
- [26] [THO0207F.wav](#) (23 Mb) 2010-03-17 13:55:48

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 2060-03-16 – Fichiers haute résolution, distribution restreinte

Les mêmes enregistrements audio sont en **libre accès** pour la version **AAC** (comprimée) mais en **accès réservé** pour la version **WAV** (source).

Index d'un objet avec identifiants et droits

Les fichiers 'zip/tar' permettent de télécharger en une seule fois l'ensemble des fichiers listés dans le répertoire qui précède.

La table des matières d'un objet affiche les valeurs des **identifiants** (hdl) attribués aux documents.

Les **liens publics** permettent le téléchargement direct. Les autres liens déclenchent une **procédure d'identification** et de contrôle des droits d'accès.

La justification du refus d'accès est affichée dans la langue de navigation du site.

• Session1

- o [2] [mix.mp3](#) => hdl:11041/ortolang-000900/mix_1.mp3 (26 Mb) **public** 2014-10-08 10:27:40 audio/mpeg
- o [3] [mix.wav](#) => hdl:11041/ortolang-000900/mix_1.wav (191 Mb) **public** 2014-10-08 06:26:21 audio/x-wav
- o **tracks_mp3**
 - [4] [TRACK0_0.mp3](#) => hdl:11041/ortolang-000900/TRACK0_0_1.mp3 (12 Mb) **public** 2014-10-08 10:27:48 audio/mpeg
 - [5] [TRACK1_0.mp3](#) => hdl:11041/ortolang-000900/TRACK1_0_1.mp3 (12 Mb) **public** 2014-10-08 10:27:56 audio/mpeg
 - [6] [TRACK2_0.mp3](#) => hdl:11041/ortolang-000900/TRACK2_0_1.mp3 (12 Mb) **public** 2014-10-08 10:28:04 audio/mpeg
 - ...
- o [13] [tracks_mp3.zip](#) => hdl:11041/ortolang-000900/tracks_mp3_1.zip (102 Mb) **public** 2014-10-10 07:33:38 application/zip
- o [14] [video.m4v](#) => hdl:11041/ortolang-000900/video_1.m4v (428 Mb) **public** 2014-10-09 13:17:45 video/mpeg
- o **tracks_wav**
 - [16] [TRACK0_0.WAV](#) => hdl:11041/ortolang-000900/TRACK0_0_1.wav (95 Mb) 2006-09-15 22:00:00 audio/x-wav
 - [17] [TRACK1_0.WAV](#) => hdl:11041/ortolang-000900/TRACK1_0_1.wav (95 Mb) 2006-09-15 22:00:00 audio/x-wav
 - [18] [TRACK2_0.WAV](#) => hdl:11041/ortolang-000900/TRACK2_0_1.wav (95 Mb) 2006-09-15 22:00:00 audio/x-wav
 - ...
 - [27] [settings.txt](#) => hdl:11041/ortolang-000900/settings_1.txt (0.000 Mb) 2014-10-09 06:26:18 text/plain
- o [28] [tracks_wav.zip](#) => hdl:11041/ortolang-000900/tracks_wav_1.zip (481 Mb) 2014-10-10 07:33:33 application/zip

• Session2

- o [30] [mix.mp3](#) => hdl:11041/ortolang-000900/mix_2.mp3 (26 Mb) **public** 2014-10-08 10:27:40 audio/mpeg
- o [31] [mix.wav](#) => hdl:11041/ortolang-000900/mix_2.wav (191 Mb) **public** 2014-10-08 06:26:21 audio/x-wav
- o **tracks_mp3**
 - [32] [TRACK0_0.mp3](#) => hdl:11041/ortolang-000900/TRACK0_0_2.mp3 (12 Mb) **public** 2014-10-08 10:27:48 audio/mpeg
 - [33] [TRACK1_0.mp3](#) => hdl:11041/ortolang-000900/TRACK1_0_2.mp3 (12 Mb) **public** 2014-10-08 10:27:56 audio/mpeg
 - [34] [TRACK2_0.mp3](#) => hdl:11041/ortolang-000900/TRACK2_0_2.mp3 (9 Mb) **public** 2014-10-08 10:36:43 audio/mpeg
 - ...
- o [41] [tracks_mp3.zip](#) => hdl:11041/ortolang-000900/tracks_mp3_2.zip (72 Mb) **public** 2014-10-10 07:38:21 application/zip
- o [42] [video.m4v](#) => hdl:11041/ortolang-000900/video_2.m4v (240 Mb) **public** 2014-10-09 13:34:19 video/mpeg
- o **tracks_wav**
 - [44] [TRACK0_0.WAV](#) => hdl:11041/ortolang-000900/TRACK0_0_2.wav (67 Mb) 2006-09-15 22:00:00 audio/x-wav

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 2064-10-07

Licences et autorisations

Licence utilisateur (SLDR)

Licence SLDR

(Version 1)

Préambule

Cette licence est destinée à garantir :

1. l'intégrité des données diffusées par le SLDR ;
2. le suivi des utilisations de ces données, au bénéfice de leurs créateurs.

Termes de la licence

En téléchargeant du site du SLDR un corpus, une ressource linguistique ou un outil (ci-après désignés comme « la ressource »), l'utilisateur s'engage sur les points suivants :

1. ne pas distribuer la ressource à de tierces personnes ; la diffusion se fait uniquement, de manière nominative, par téléchargement du site SLDR, après accord sur la présente licence ;
2. signaler clairement l'origine (identifiant unique "slrxxxxxx") dans toute publication ou utilisation (même non commerciale) de la ressource, et reproduire les références bibliographiques des articles mentionnés sur la fiche descriptive, le cas échéant ;
3. informer le SLDR de cette publication ou utilisation. Cette information sera saisie dans l'espace de l'utilisateur (lien "Déposer -> Publication") ;
4. informer le SLDR de tout enrichissement de la ressource et mettre cet enrichissement au service de la communauté scientifique, via le SLDR, en accord avec les auteurs de la ressource.

Signature d'autorisations



Avec Victorine DUMAS et Claude JOUVAL au Mas de la Pyramide (St-Rémy-de-Provence) le 22 avril 2011



Informateur/locuteur

Consentement éclairé

Je soussigné(e),
domicilié à, autorise l'utilisation des enregistrements audio et vidéo auxquels j'ai participé dans le cadre du projet XXXXX dans les conditions suivantes :

1. Le projet XXXXX est consacré à la collection de données à des fins de recherche scientifique (mémoires, thèses, articles scientifiques, exposés à des congrès, séminaires), d'enseignement universitaire, qui seront partagées par la communauté des chercheurs. Toute autre finalité d'utilisation des documents ne pourra être envisagée qu'après révision du présent accord.
2. Seuls les documents contenus dans des répertoires étiquetés 'PUBLIC' pourront être diffusés publiquement par un dispositif d'audition ou de visualisation en ligne.
3. Ces documents destinés à une diffusion publique m'ont été soumis pour approbation. Je me réserve le droit d'exiger à tout moment que certains passages de ces documents soient masqués ou anonymisés.
4. Les documents autres que publics ne pourront être distribués que par téléchargement à partir du site de la Banque de données parole et langage (SLDR, <http://sldr.org>).
5. Ces téléchargements seront réservés à des personnes authentifiées ayant prouvé leur appartenance à des organismes de recherche ou d'enseignement, et approuvé la licence utilisateur du SLDR, ainsi que la licence complémentaire relative à XXXXX, licences dont j'ai pris connaissance.
6. Clauses supplémentaires :
.....
.....

Lieu et date :

Signature :

Autorisation et licence complémentaire

Utilisateur

Licence complémentaire pour « Valjouffrey-Valbonnais 2012 »

Cette licence contient des clauses complémentaires à celles de la licence utilisateur SLDR http://sldr.fr/wiki/Licences_fr régissant l'utilisation des documents audio, vidéo, graphiques, numériques ou textuels téléchargés avec l'objet [sldr000787](#) (Valjouffrey-Valbonnais 2012).

Je m'engage à respecter les règles de confidentialité suivantes :

1. Seuls les documents contenus dans des répertoires étiquetés 'PUBLIC' peuvent être diffusés librement.
2. Les documents contenus dans des répertoires étiquetés 'CONFIDENTIEL' peuvent être utilisés pour des travaux de recherche mais ne doivent en aucun cas être diffusés, y compris lors de la présentation de ces travaux à un auditoire sélectionné.

Quelques références

BEL, Bernard (2015). The OAIS model in Digital Humanities: feedback from archival practice in linguistics. Proceedings of *Speech audio archives: preservation, restoration, annotation, aimed at supporting the linguistic analysis* (2015 septembre 14 : Rome, ITALIE)

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

BEL, Bernard (2014). Mutualisation et archivage pérenne des données orales : un nouveau cadre technique et juridique au service de la recherche en linguistique. Actes, Colloque *Les Archives de la Recherche : problèmes et enjeux de la construction du savoir scientifique* (2012 janvier 18-20 : Paris, FRANCE)

<http://hal.archives-ouvertes.fr/hal-00983697>

BEL, Bernard (2012). A Trusted Digital Repository based on the OAIS model with integrated management of access rights. Proceedings of *Cultural Heritage On Line - Trusted Digital Repositories & Trusted Professionals* (2012 décembre 11-12 : Florence, ITALIE). 2012.

<http://hal.archives-ouvertes.fr/hal-00983703>

BEL, Bernard (2012). Implementing the OAIS for oral/linguistic resources: the Speech & Language Data Repository venture.

<http://hal.archives-ouvertes.fr/hal-01053214>