

SSHOC Workshop

Data Citation in Practice

Cocoon a French repository dedicated to oral resources



Nicolas Larrousse
Michel Jacobson

Huma-Num (CNRS)
Huma-Num (CNRS)



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782



Cocoon repository in brief

- Creation in 2006
- Co-management by two laboratories

LACITO (LANGUES et CIVILISATIONS à Tradition Orale

<https://lacito.vjf.cnrs.fr/>)

LLL (Laboratoire Ligérien de Linguistique - <https://lll.cnrs.fr/>)

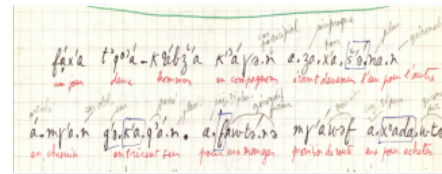
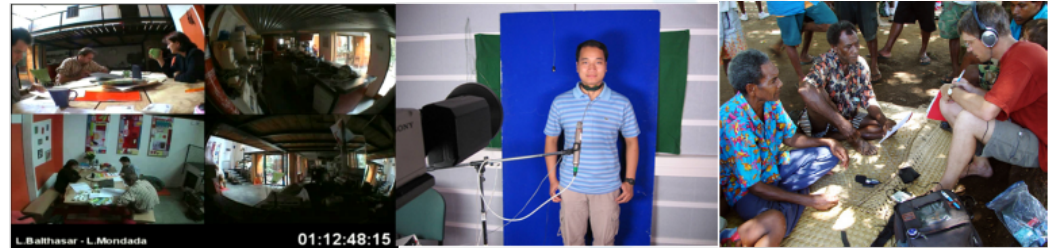
- Hosted by Huma-Num
- CLARIN Center in 2015



Cocoon repository in brief

Oral data for SSH in France

- Recordings (audio & video)
- Annotations
Transcriptions
Translations
Measures
- Metadata expressed in OLAC
- Extended curation
- Triple Store
- Long term preservation at CINES



Plan de classement

- Collection de l'Institut d'histoire du temps présent (IHTP)
 - Archives orales de l'IHTP
 - Fonds Michel Trebitsch (IHTP)
 - Fonds Pierre Mendès France (IHTP). Entretiens oraux et colloques "L'expérience Mendès France (1954-1955) et le..."
 - Série Colloques IHTP
 - Série Entretiens IHTP
 - Série Séminaires IHTP

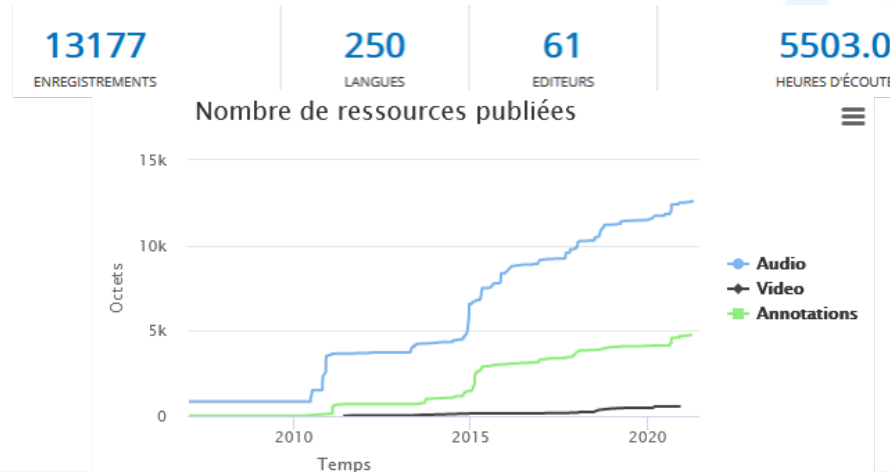
```
<dc:contributor olac:code="depositor" xsi:type="olac:role">Ozanne-Rivierre, Françoise</dc:contributor>
<dcterms:created xsi:type="dcterms:W3CDTF">1973</dcterms:created>
<dc:format xsi:type="dcterms:IMT">audio/x-wav</dc:format>
<dc:type xsi:type="dcterms:DCMIType">Sound</dc:type>
<dcterms:extent>PTOM57S</dcterms:extent>
<dcterms:isRequiredBy xsi:type="dcterms:URI">
  oai:crdo.vjf.cnrs.fr:cocoon-e46923ef-9b3a-3340-976e-2edca9d39a8f
</dcterms:isRequiredBy>
<dc:rights>Copyright (c) Ozanne-Rivierre, Françoise</dc:rights>
<dc:subject xml:lang="fr" olac:code="nem" xsi:type="olac:language">Nemi</dc:subject>
```



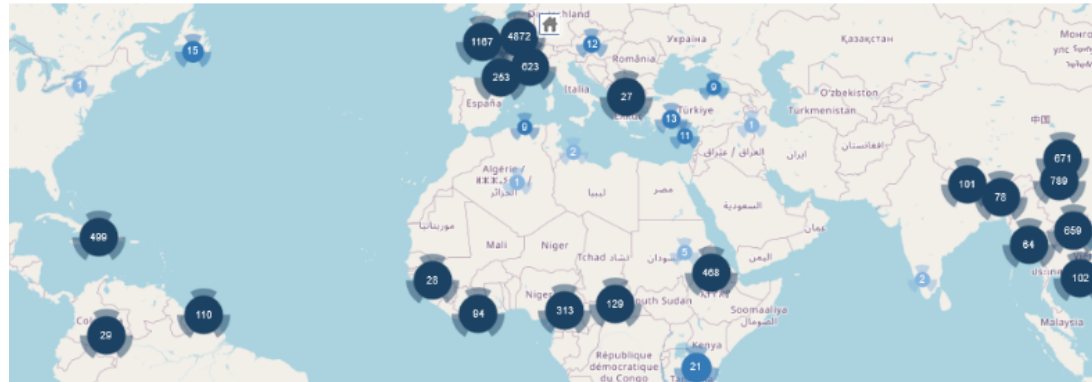
Cocoon repository in brief

A few figures

- Around 14000 recordings
- 250 languages
- 5000 hours



Sont représentés dans cette carte les régions dans lesquelles les chercheurs ont enregistré leurs locuteurs.



Cocoon repository in brief

The screenshot shows the 'Trésors de la parole' website interface. At the top, it says 'Un site du ministère de la Culture'. The main navigation includes 'RECHERCHE', 'LE PROJET', and 'LANGUES'. Below the navigation, there's a search bar and a 'RECHERCHE' section. The search results for 'D'ICI ET DE LÀ-BAS' are displayed, showing a list of items with details like 'Éditeur: Centre de Recherche Berbère', 'Langue: Kabyle', and 'Sujets: cuisine, relations humaines, Kabyle, famille, culture, variation linguistique, développement durable'. A play button and a 'Notice' link are visible for each result.

The screenshot shows the 'Pangloss' website interface. The main heading is 'The Sahu Hari Das has many troubles'. Below the heading, there's a video player with a play button and a progress bar. To the right of the video player, there's a section for 'Options d'affichage' with checkboxes for 'Lecture en continu', 'Transcription du texte', 'Traduction du texte', 'Transcription des phrases', and 'Traduction des phrases'. Below this, there's a list of items with their respective text and a play button. To the right of the video player, there's a section for 'Fichier média original (wav)', 'Fichier média compressé (mp3)', and 'Annotation 1 (xml)'. Below this, there's a section for 'Metadonnées' with links for 'Ressource (xml)' and 'Annotation 1 (xml)'. At the bottom right, there's a section for 'Chercheur(s)', 'Locuteur(s)', 'Date d'enregistrement', and 'Lieu d'enregistrement'.

Reuse of data in several portals



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Cocoon and citation

Use of different types of PIDs

- **DOIs**
<https://doi.org/10.34847/COCOON.ce8a4a03-5083-3144-88ce-7cc4606e8352>
- **ARK**
<https://cocoon.huma-num.fr/exist/crdo/ark:/87895/1.17-1037964>
- **PURL**
<http://purl.org/poi/crdo.vjf.cnrs.fr/cocoon-ce8a4a03-5083-3144-88ce-7cc4606e8352>

And formally “handles” ...



Cocoon and citation

Different types of resources

- **Collection**

<https://doi.org/10.34847/cocoon.262c6f48-fc06-375b-9313-dd4142d55732>

- **Recording**

<https://doi.org/10.34847/cocoon.03343f2a-1c6b-382d-83fd-d66013d6b072>

- **Annotation**

<https://doi.org/10.34847/cocoon.e50826ee-32ff-3935-be2f-8f3f80148e57>

Notice de collection

Corpus d'Orléans: ESLO1: la recette de l'omelette

Baude, Olivier (compilier)

(mise à disposition: 2015-04-22; archivage: 2015-07-09T09:53:07+02:00; dernière modification de la notice: 2021-06-04)

Plan de classement

Corpus d'Orléans

Corpus d'Orléans: ESLO1: la recette de l'omelette

Notice d'enregistrement

ESLO - Omelette bande 007

Ross, John; Kay, Jack; Blanc, Michel; Biggs, Patricia; Dalwood, Mary

Centre Orléanais de Recherche en Anthropologie et Linguistique (depositor, researcher); Crown, Richard (speaker); FU200 (speaker); M'Charek, Linda (transcriber); Petit, Mélanie (transcriber)

(création: 1969-04-15; mise à disposition: 2010-06-30; archivage: 2010-06-30T10:39:36+02:00; dernière modification de la notice: 2020-11-27)

Position dans le plan de classement

Corpus de la parole

Notice d'annotations

ESLO - Omelette bande 007

Dalwood, Mary; Ross, John; Kay, Jack; Blanc, Michel; Biggs, Patricia

Centre Orléanais de Recherche en Anthropologie et Linguistique (depositor, researcher); Crown, Richard (speaker); FU200 (speaker); Petit, Mélanie (transcriber); M'Charek, Linda (transcriber)

(dernière modification: 2018-04-10T12:14:16+02:00; mise à disposition: 2011-02-04; archivage: 2011-02-04T23:50:20+01:00)


Editeur(s): Centre Orléanais de Recherche en Anthropologie et Linguistique



Cocoon and citation

Recording

<https://doi.org/10.34847/cocoon.03343f2a-1c6b-382d-83fd-d66013d6b072>



The screenshot shows a web interface for 'ESLO - Omelette bande 007'. At the top, there are navigation tabs: 'RECHERCHER / DEPOSER', 'DOCUMENTATION', 'A PROPOS', and 'Corpus'. Below the title, there are citation style options: APA, Harvard, MLA, Vancouver, Chicago, IEEE, BibTeX, and RIS. The selected citation text is: 'Blanc, Michel, Biggs, Patricia, Ross, John, Kay, Jack, Dalwood, Mary, & Centre Orléanais de Recherche en Anthropologie et Linguistique. (2010, juin 30). *ESLO - Omelette bande 007* (Version 1). Centre Orléanais de Recherche en Anthropologie et Linguistique. <https://doi.org/10.34847/COCOON.03343F2A-1C6B-382D-83FD-D66013D6B072>'. A 'Copier' button is next to the citation. Below the citation, there are fields for 'Format(s): [fr] 1 bande magnétique audio', 'Droits: Librement accessible', and 'Copyright (c) 1993 Université d'Orléans/CORAL'. A section titled 'Identifiant(s):' lists various identifiers: doi:10.34847/cocoon.03343f2a-1c6b-382d-83fd-d66013d6b072, hdl:10670/1.vf95yr, oai:crdo.vjf.cnrs.fr:cocoon-03343f2a-1c6b-382d-83fd-d66013d6b072, ark:/87895/1.5-124608, and [fr] Ancienne cote: crdo-FRA_ESLO_omelette007_SOUND. A 'Pour citer la ressource:' section shows a 'doi' icon and the same URL with a 'citation' button. On the right, a 'Lieu(x) d'enregistrement' section provides location information: 'France, Centre-Val de Loire, Loiret, Orléans', 'Identifiant TGN: 7008337', 'Identifiant Geonames: 6454159', and 'Pays/région (ISO-3166): France'. A small map snippet is visible at the bottom of this section.



Cocoon and citation

Citing part of a recording

Use of specifications from W3C "Media Fragments URI"

<https://www.w3.org/TR/media-frags/>

Use of the "temporal tag"

#t=start,end

<https://doi.org/10.34847/COCOON.ce8a4a03-5083-3144-88ce-7cc4606e8352#t=4.3201,9.3600>

<http://purl.org/poi/crdo.vjf.cnrs.fr/cocoon-ce8a4a03-5083-3144-88ce-7cc4606e8352#t=53.3600,56.3200>



Cocoon and citation

Citing part of a recording

<https://cocoon.huma-num.fr/exist/crdo/meta/cocoon-03343f2a-1c6b-382d-83fd-d66013d6b072#t=4,9>



The screenshot shows the Cocoon website interface. At the top, there is a navigation bar with the Cocoon logo and menu items: RECHERCHER / DEPOSER, DOCUMENTATION, A PROPOS, Corpus, and TR. A 'Connexion' button is visible on the right. Below the navigation bar, the main content area displays a notice titled 'ESLO - Omelette bande 007'. The notice lists authors: Ross, John; Kay, Jack; Blanc, Michel; Biggs, Patricia; Dalwood, Mary. It also lists the depositor and researchers: Centre Orléanais de Recherche en Anthropologie et Linguistique. The notice includes creation, availability, archiving, and modification dates. A 'Position dans le plan de classement' section shows a breadcrumb trail: 'Corpus de la parole' > 'ESLO - Omelette bande 007'. On the right side, there is a sidebar with a play button icon and the text 'Consulter le document lire l'enregistrement audio'. Below this is an audio player showing a progress bar at 0:04 / 0:54. Further down, there is a section 'consulter les annotations' with a button 'Afficher' and the text 'ESLO - Omelette bande 007'. At the bottom of the sidebar, there is a 'Partager' button.



Cocoon and citation

Citing version with specifiers

PID of the recording

<http://purl.org/net/crdo/data/cocoon-f031bcb4-ae4e-3d7a-8966-91ad43d6e825>

Version(s) of the recording

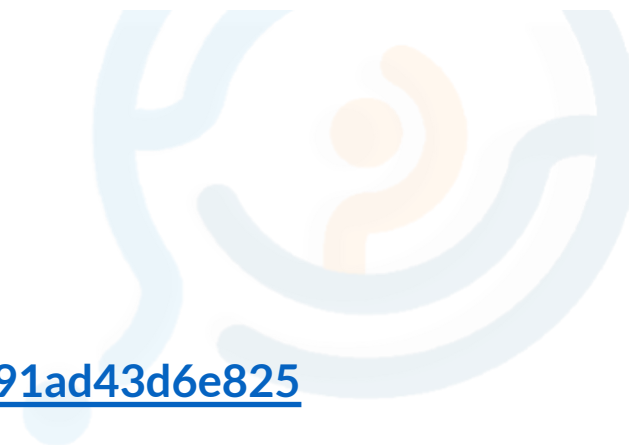
<http://purl.org/net/crdo/data/cocoon-f031bcb4-ae4e-3d7a-8966-91ad43d6e825.version1>

<http://purl.org/net/crdo/data/cocoon-f031bcb4-ae4e-3d7a-8966-91ad43d6e825.version2>

<http://purl.org/net/crdo/data/cocoon-f031bcb4-ae4e-3d7a-8966-91ad43d6e825.version3>

Other version

<http://purl.org/net/crdo/data/cocoon-f031bcb4-ae4e-3d7a-8966-91ad43d6e825.diffusion>



Cocoon and citation

Different granularities for the citation

- Collection
 - Recording
 - Annotation
 - Fragment
 - Version
- Etc.

One of the benefits of putting your data in a repository



Thank you for your attention!

Nicolas.Larrousse@huma-num.fr
Michel.Jacobson@huma-num.fr



Join our community



<https://www.sshopencloud.eu>



@SSHOpenCloud



info@sshopencloud.eu



/in/sshopencloud



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

