

Un plan de gestion de données rétrospectif pour un programme financé par l'ANR : COLOSTRUM

Rédigé le 2 octobre 2020, révision 21 mai 2021

Claire Cialone-Grégoire & Véronique Ginouvès, archivistes

Le programme “L'alimentation pré-lactée, don et consommation néonatale du colostrum : pratiques, représentations, et enjeux de santé publique” – dit “ANR Colostrum” car financé par l'Agence Nationale de la Recherche (ANR) – est un programme de recherche interdisciplinaire et international consacré à l'alimentation pré-lactée du nouveau-né. Dès le début, Joël Candau, professeur en anthropologie et coordinateur du projet, a souhaité intégrer la gestion des données qui seraient produites au cours des quatre années de recherche. Sans avoir donné lieu à un document spécifique, cette volonté préfigurait une forme de plan de gestion de données telle que celle qui est demandée depuis 2019 par les programmes financés par l'Agence.

En effet, la rédaction d'un plan de gestion de données est devenue obligatoire pour les programmes financés par l'ANR depuis 2019, et fortement recommandée par le Ministère Français de l'Enseignement Supérieur, de la Recherche et de l'Innovation ([Plan national pour la Science Ouverte](#)) comme par l'Europe (H2020 et European Research Council). Les plans de gestion de données (PGD) s'inscrivent dans un contexte de développement de la science ouverte qui était un des objectifs du porteur du projet, Joël Candau.

L'Institut de Recherche pour le développement les présente ainsi :

« Le Plan de Gestion de Données ou Data Management Plan (PGD ou DMP) est un document formalisé rédigé de préférence au démarrage d'un projet de recherche qui décrit la façon dont les données seront produites ou obtenues, (ré)utilisées, traitées, organisées, stockées, sécurisées, préservées, documentées, partagées, au cours et à l'issue d'un projet, explicitant notamment leur mise à disposition. »¹

Les données de la recherche sont les matériaux produits par les chercheurs et servant de base à leur publication. L'intérêt de traiter et d'archiver ces données est multiple. Il s'agit notamment d'un gage de transparence dans le processus de la recherche : en rendant possible l'accès à ses données, le chercheur rend compte de la validité de ses sources. Par ailleurs, il est tout à fait envisageable que des données produites dans une discipline puissent être réutilisées dans un autre

¹ <https://data.ird.fr/plan-de-gestion-de-donnees/>

contexte, ou bien suivant un autre angle d'approche. Dès lors, meilleure sera leur documentation, plus précise sera leur réexploitation².

Quand Joël Candau a rédigé le projet de l'ANR, il a intégré dès le départ la gestion et l'ouverture des données qui allaient être produites au cours des 4 années de recherche, en insistant sur le volet anthropologique (entretiens et photographies). Dans cet objectif, la phonothèque a été sollicitée pour archiver les données produites sur le long terme et former les chercheurs en amont pour faciliter cette réalisation. Ainsi, si au lancement du programme, en 2012, la rédaction d'un PGD n'était pas encore obligatoire pour les projets financés par l'ANR, la gestion des données produites lors des recherches a été prise en compte dès le début par les acteurs de "l'ANR Colostrum", en particulier en ce qui concerne les données anthropologiques.

Au cours de l'année 2020 le Consortium "[Archives des Ethnologues](#)" a eu pour mission de rendre "FAIR" les données produites au cours de ces dernières années au sein de son réseau. Cet engagement³ a été une occasion favorable pour clôturer l'archivage des données du programme de recherche de l'ANR Colostrum⁴. La démarche exemplaire de Joël Candau peut certainement rendre service à d'autres équipes menant ce type de programme en rendant compte d'une problématique au coeur de préoccupations d'actualité.

Après avoir présenté le programme "ANR Colostrum", nous dresserons un état des lieux des données produites lors des recherches (en particulier les données anthropologiques). Puis avant de présenter les modalités de traitement, nous reviendrons sur le travail de préparation en amont des recherches jusqu'au dépôt des données produites pour leur conservation sur le long terme au Centre Informatique National de l'Enseignement Supérieur - CINES⁵. Ce plan rétrospectif de gestion des données sera aussi l'occasion de mettre en avant les difficultés qui peuvent survenir dans le traitement des données de la recherche, notamment le décalage qui peut exister entre un PGD prévisionnel et la pratique des chercheurs.

1. Description du projet "ANR Colostrum"

² L'objectif est de suivre les 4 principes du FAIR (Findable, Accessible, Interoperable, Reusable) : <https://www.go-fair.org/fair-principles>

³ Claire Cialone-Grégoire, archiviste en auto-entrepreneur, a pu ainsi effectuer une analyse des données et de leur traitement, valider le plan de nommage de fichier, vérifier l'EAD du traitement archivistique avant l'envoi des métadonnées et des fichiers numériques.

⁴ Voir infra "Description du projet"

⁵ Le CINES est un établissement public à caractère administratif national (EPA), basé à Montpellier, et placé sous la tutelle du Ministère de l'enseignement supérieur, de la recherche et de l'innovation (MESRI) qui propose des moyens exceptionnels à l'ensemble de la communauté scientifique dans ses missions statutaires stratégiques nationales en particulier pour l'archivage pérenne de données électroniques. La collaboration entre le CINES et la phonothèque de la MMSH se fait à travers l'équipe de la TGIR Huma-Num.

Les recherches du programme “L’alimentation pré-lactée, don et consommation néonatale du colostrum : pratiques, représentations, et enjeux de santé publique” (ou “ANR Colostrum”), coordonnées par Joël Candau (LAPCOS - Université de Nice Sophia Antipolis), ont été menées sur une période de trois ans du 1er janvier 2013 au 31 décembre 2015, puis prolongées sur une année (2016), avec la collaboration de cinq laboratoires de recherche : le Centre d’anthropologie culturelle (CANTHEL, EA 4545) de l’université de Paris Descartes (Paris, France), le Centre des Sciences du Goût et de l’Alimentation - UMR 6265 (Dijon, France), le laboratoire d’Anthropologie et de psychologie sociale et cognitive (LAPCOS, EA 7278) de l’université de Nice Sophia Antipolis (Nice, France), la phonothèque de la Maison méditerranéenne des sciences de l’homme (MMSH, USR 3125, Aix-en-Provence, France), «Tolérance Immunitaire » (EA 6302 TIM) de l’université de Nice Sophia Antipolis (France). Suivant une approche interdisciplinaire, le programme était composé de quatre volets complémentaires : anthropologique, immunologique, psychobiologique, et un volet “open science” (science ouverte) pour la gestion des données et la restitution des résultats. De façon globale, deux axes majeurs se sont dégagés : un axe anthropologique, dont les données allaient être plus particulièrement traitées dans une optique de science ouverte (description de la collecte anthropologique, conservation et diffusion des données collectées sur le terrain) ; et un axe biologique et psychobiologique.

L’axe anthropologique visait à documenter les pratiques et les représentations relatives à la consommation néonatale du colostrum dans des contextes sociaux et culturels contrastés. Treize partenaires, présents dans sept pays différents (Allemagne, Bolivie, Brésil, Burkina Faso, Cambodge, France, Maroc), ont ainsi participé à la réalisation des enquêtes anthropologiques. La phonothèque, au sein de la médiathèque de la MMSH, en collaboration avec l’équipe du LAPCOS (EA 7278), a été chargée de cataloguer, indexer et documenter le dépôt des données issues des entretiens menés par les anthropologues et d’archiver les données de la recherche, suivant les exigences de la science ouverte.

L’axe psychobiologique et biologique s’attachait quant à lui à deux aspects particuliers du colostrum : ses propriétés sensorielles et fonctionnelles (pour des adultes, parents ou non, et pour des nouveau-nés), mises en lien avec les données récoltées sur le terrain par les anthropologues. Pour ces deux axes, la partie psychobiologique de la recherche a été menée par l’équipe d’éthologie du Centre des sciences du goût et de l’alimentation - CSGA, UMR 6265 à Dijon tandis que des expériences ont été conduites par l’EA 6302 Tolérance immunitaire (TIM) de l’Hôpital L’Archet (Nice) sur des souriceaux portant sur le potentiel immunologique du colostrum et ses effet sur la prévention des allergies.

Dans le cadre du volet “open science” un financement spécifique mis en place la deuxième année du programme a permis l’embauche d’une ingénieure d’étude⁶ qui a collecté, traité et archivé les données anthropologiques produites sur le terrain. Par ailleurs, sur le site de la Maison des sciences de l’homme de Nice, un carnet de recherche collaboratif a été créé et alimenté sur la plateforme Hypotheses : <https://colostrum.hypotheses.org>. En fin de programme, sur un financement plus léger⁷, les données anthropologiques ont été cataloguées sur [Calames](#) et dans la [base de données de la phonothèque](#), tandis qu’une série de photographies ont été mises en ligne sur [MédiHAL](#) (archive ouverte de photographies et d’images scientifiques du CCSD).

2. Etat des lieux des données produites lors des recherches

Le programme « ANR Colostrum » suivait deux axes de recherches, aboutissant tous deux à la production d’une certaine quantité de données de terrain auxquelles se sont ajoutées les données contextualisantes : données administratives, courriers, supports de restitution des résultats (articles, diaporamas…), comptes-rendus de réunion et rapports, supports de formation.

- **Les données anthropologiques**

Dès le début du programme, l’accent avait été mis sur la gestion des données anthropologiques. La phonothèque de la MMSH ainsi reçu les données de terrain en provenance de sept pays, dont voici un état des lieux présenté en tableaux récapitulatifs :

ALLEMAGNE	
Fichiers WAVE	25 entretiens (4940 à 4964)
Fichiers MP3	25 entretiens (4940 à 4964)
Photographies	35 photographies (JPEG)
Transcriptions (DOC ou PDF) : Mères	20 fichiers (T-4940 à T-4959)
Transcriptions (DOC ou PDF) : Pères / couples (+ "éleveurs" pour le Maroc	1 fichier (T-4964)

⁶ Marine Soubrié, archiviste, a obtenu un contrat d’ingénieure de recherche sur 11 mois.

⁷ Giulia Manfredi, archiviste, a obtenu un stage de Master 2 “Métiers des archives et des bibliothèques. Méditations de l’histoire et humanités numériques” Aix-Marseille Université de 3 mois.

Sud)	
Transcriptions (DOC ou PDF) : Personnel soignant	4 fichiers (T-4960 à T-4963)
Contrats d'autorisation auprès des informateurs	AUCUN
Contrat phonothèque / enquêteur	AUCUN
Fiches descriptives	AUCUN
Questionnaires papier	AUCUN
Questionnaires sonores	AUCUN
CV des enquêteurs	AUCUN
Divers	2 fichiers (dates enquêtes et compte- rendu préliminaire)

BOLIVIE	
Fichiers WAVE	30 fichiers pour 27 entretiens (4917 à 4939, 4965 à 4967, 5231. Certains entretiens ont été enregistrés en plusieurs temps : (4918, 4919, 4934)
Fichiers MP3	AUCUN
Photographies	87 photographies (JPEG)
Transcriptions (DOC ou PDF) : Mères	21 fichiers (T-4917 à T-4936, T-4938, T-5231)
Transcriptions (DOC ou PDF) : Pères / couples (+ "éleveurs" pour le Maroc Sud)	1 fichier (T-4937)
Transcriptions (DOC ou PDF) : Personnel soignant	4 fichiers (T-4939 + T-4965 à T-4967)
Contrats d'autorisation auprès des informateurs	AUCUN
Contrat phonothèque / enquêteur	un contrat entre l'enquêtrice et la phonothèque
Fiches descriptives	24 fichiers (4917 à 4939 + 4967) >> manquent les fiches descriptives de 3 entretiens
Questionnaires papier	21 fichiers : Q1 (regroupement de 100 questionnaires auprès des mères) et 20 Q2 (personnel soignant)
Questionnaires sonores	AUCUN
CV des enquêteurs	CV Céline Geffroy
Divers	5 fichiers : prospectus sur l'allaitement, traitement des données (versions de travail)

BRESIL	
Fichiers WAVE	9 entretiens (4868 à 4876)
Fichiers MP3	9 entretiens (4868 à 4876)
Photographies	AUCUNE
Transcriptions (DOC ou PDF) : Mères	17 transcriptions (français portugais) pour 9 enquêtes (T-4868 à T-4876)
Transcriptions (DOC ou PDF) : Pères / couples (+ "éleveurs" pour le Maroc Sud)	
Transcriptions (DOC ou PDF) : Personnel soignant	
Contrats d'autorisation auprès des informateurs	5 contrats d'autorisation d'enquête + 119 contrats d'autorisation de questionnaire
Contrat phonothèque / enquêteur	AUCUN
Fiches descriptives	AUCUNE
Questionnaires papier	142 fichiers : 117 Q1 (auprès des mères, en français ou en portugais) ; 25 Q2 (auprès du personnel soignant, en français ou en portugais)
Questionnaires sonores	103 Q1 (auprès des mères) ; 20 Q2 (personnel soignant)
CV des enquêteurs	AUCUN
Divers	4 fichiers : 2 questionnaires pour entretien (V1 et V2) ; catalogue ; données démographiques pour le Brésil

BURKINA FASO

Fichiers WAVE	26 entretiens (4817 à 4843)
Fichiers MP3	26 entretiens (4817 à 4843)
Photographies	27 photographies (JPEG)
Transcriptions (DOC ou PDF) : Mères	1 fichier (T-4819) >> manquent les transcriptions de tous les autres entretiens
Transcriptions (DOC ou PDF) : Pères / couples (+ "éleveurs" pour le Maroc Sud)	AUCUNE
Transcriptions (DOC ou PDF) : Personnel soignant	AUCUNE
Contrats d'autorisation auprès des informateurs	AUCUN
Contrat phonothèque / enquêteur	AUCUN
Fiches descriptives	3 fichiers : 1 fichier regroupant 21 fiches descriptives des entretiens auprès des mères ; 1 fichier regroupant 5 fiches descriptives des entretiens auprès du personnel soignant ; 1 fichier modèle de fiche descriptive
Questionnaires papier	124 fichiers : 101 Q1 (auprès des mères) ; 21 Q2 (personnel soignant) ; 2 tableurs récapitulatifs des questionnaires
Questionnaires sonores	AUCUN
CV des enquêteurs	AUCUN
Divers	7 "documents annexes" : rapport 2013-2016, rapport-étape, catalogue, données démographiques, bilan de terrain (x2), références

CAMBODGE	
Fichiers WAVE	27 entretiens (4776 à 4802)
Fichiers MP3	3 entretiens (4777, 4787, 4788)
Photographies	52 photographies (JPEG)
Transcriptions (DOC ou PDF) : Mères	20 fichiers (T-4776 à T-4795)
Transcriptions (DOC ou PDF) : Pères / couples (+ "éleveurs" pour le Maroc Sud)	3 fichiers (T-4796 à T-4798)
Transcriptions (DOC ou PDF) : Personnel soignant	4 fichiers (T-4799 à 4802)
Contrats d'autorisation auprès des informateurs	39 fichiers : 20 pour les mères ; 3 pour les pères ; 3 pour le personnel soignant (>> manquant pour enquête 4799). 10 contrats d'autorisation pour les photographies ; 3 pour autorisation de l'hôpital Calmette
Contrat phonothèque / enquêteur	AUCUN
Fiches descriptives	27 fichiers : 20 pour les mères (FD-4776 à FD-4795) ; 3 pour les pères (FD-4796 à FD-4798) ; 4 pour le personnel soignant (FD-4799 à FD-4802)
Questionnaires papier	1 fichier : tableau récapitulatif des données issues des questionnaires Q1 (100 questionnaires passés) et Q2 (20 questionnaires passés)
Questionnaires sonores	AUCUN
CV des enquêteurs	2 CV enquêtrices
Divers	8 fichiers : 1 lettre de motivation ; 4 messages des Pr Mengly et Kry Leang Sim, J. Candau, S. Goyet et E. Bureau ; 3 "documents annexes" (PP de restitution des résultats à l'hôpital, photo et "carte à remettre" = présentation du projet en khmer)

FRANCE	
Fichiers WAVE	24 entretiens (4845 à 4867 et 5230)
Fichiers MP3	22 entretiens (4846 à 4867)
Photographies	58 photographies (JPEG)
Transcriptions (DOC ou PDF) : Mères	3 fichiers (T-4848, T-4851 et T-4861)
Transcriptions (DOC ou PDF) : Pères / couples (+ "éleveurs" pour le Maroc Sud)	AUCUNE
Transcriptions (DOC ou PDF) : Personnel soignant	1 fichier (T-4864)
Contrats d'autorisation auprès des informateurs	22 fichiers : 13 pour les mères (manquants pour 4845 à 4848, 4850, 4851) ; 4 pour le personnel soignant (manquant pour 5230) + 4 contrats d'autorisation pour les photos + 1 autorisation maternité
Contrat phonothèque / enquêteur	AUCUN
Fiches descriptives	AUCUNE
Questionnaires papier	1 fichier : tableau récapitulatif des données issues des questionnaires Q1 (100 questionnaires passés) et Q2 (20 questionnaires passés)
Questionnaires sonores	AUCUN
CV des enquêteurs	CV Céline Verguet
Divers	4 fichiers : 4 "documents annexes" ("empathelien", bibliographie données de terrain)

MAROC (Fès)	
Fichiers WAVE	22 enquêtes : 18 entretiens en arabe (5210 à 5227) ; 2 entretiens en français (5208 et 5209) ; 2 pères (5228 et 5229)
Fichiers MP3	22 enquêtes (5208 à 5229)
Photographies	3 photographies (JPEG)
Transcriptions (DOC ou PDF) : Mères	9 fichiers (T-5210 à T-5218)
Transcriptions (DOC ou PDF) : Pères / couples (+ "éleveurs" pour le Maroc Sud)	AUCUNE
Transcriptions (DOC ou PDF) : Personnel soignant	AUCUNE
Contrats d'autorisation auprès des informateurs	AUCUN
Contrat phonothèque / enquêteur	2 contrats (dont un en double) entre la phonothèque et les enquêteurs
Fiches descriptives	6 fichiers (FD-5210, FD-5211, FD-5213 à FD-5215, FD-5217) >> autres fiches descriptives manquantes
Questionnaires papier	3 fichiers : tableaux récapitulatifs des données issues des questionnaires Q1 (6 questionnaires passés) et Q2 (20 questionnaires passés)
Questionnaires sonores	AUCUN
CV des enquêteurs	AUCUN
Divers	AUCUN

MAROC (Sud)	
Fichiers WAVE	28 enquêtes (4877 à 4904)
Fichiers MP3	28 enquêtes (4877 à 4904)
Photographies	34 photographies (JPEG)
Transcriptions (DOC ou PDF) : Mères	25 fichiers (T-4877 à T-4901)
Transcriptions (DOC ou PDF) : Pères / couples (+ "éleveurs" pour le Maroc Sud)	2 fichiers (T-4903 et T-4904) >> manque T-4902
Transcriptions (DOC ou PDF) : Personnel soignant	AUCUNE
Contrats d'autorisation auprès des informateurs	AUCUN
Contrat phonothèque / enquêteur	AUCUN
Fiches descriptives	28 fichiers (FD-4877 à FD-4904)
Questionnaires papier	AUCUN
Questionnaires sonores	AUCUN
CV des enquêteurs	CV Marie-Luce Gélard
Divers	2 fichiers : Compte-rendu provisoire, "approche linguistique"

- ***Les données des sciences de la vie***

Au début du programme "ANR Colostrum", le projet était d'archiver toutes les données produites lors des recherches. Toutefois, la notion de science ouverte est beaucoup moins présente dans l'axe biologique et psychobiologique, la question de la gestion des données n'est pas vraiment posée dans le domaine des sciences de la vie. Les données issues de cet axe n'ont pas été transmises à la phonothèque pour être associées aux données anthropologiques. Ces données sont traitées et conservées dans les laboratoires producteurs :

- Tolérance Immunitaire : les données produites dans le laboratoire TIM de l'Hôpital de l'Archet de Nice concernent l'impact du colostrum sur la croissance et le développement immunitaire. Ces données ne sont pas

publiques, et il n'y a encore pas de publication issue de leur exploitation [Date probable de publication : 2022] .

- Centre des Sciences du Goût et de l'Alimentation : destruction des données produites au CSGA pour des raisons éthiques et juridiques. En effet, le travail reposait essentiellement sur des films de nouveaux-nés, films qui ne peuvent être conservés que quelques mois, le temps de l'exploitation⁸.

- **Les données contextualisant la recherche**

Il a semblé important, à l'occasion de la rédaction de ce plan rétrospectif de gestion des données, de faire un point sur les données qui ont été produites au cours du programme "ANR Colostrum", même celles qui ne sont pas issues directement des recherches.

Il s'agit des documents qui ont accompagné et préparé la production des données sur le terrain et la publication finale des résultats : préparation du dossier scientifique présenté à l'ANR, notifications financières, éléments de méthodologie, publications, guides d'enquête, comptes-rendus scientifiques, comptes-rendus des réunions, supports de restitution et de formation (présentations PowerPoint), enregistrements et photographies des réunions, rapports intermédiaires des enquêtes de terrain, documents concernant le dépôt et le traitement des données à la phonothèque de la MMSH... Toutes ces données font partie intégrante du programme "ANR Colostrum" et sont représentatives de la façon dont "se fait la recherche", surtout dans le cadre d'un projet mettant en jeu autant d'acteurs.

La bibliographie n'est pas archivée avec les autres données. Elle est consultable à la fin du chapitre d'ouvrage "Une approche bioculturelle du premier aliment du nouveau-né. Le colostrum" présentant les résultats de l'ANR Colostrum dans l'ouvrage collectif *Premiers cris, premières nourritures*⁹; les articles, chapitres d'ouvrage, rapport et autres documents de la bibliographie ont été déposés sur l'espace de stockage partageable Sharedocs¹⁰, un gestionnaire de fichiers mis en oeuvre par la TGIR Huma-Num, sur ses propres serveurs, pouvant être utilisé via un navigateur web. Il correspond aux usages de l'espace privé créé au départ du projet sur "DropBox"¹¹.

⁸ Joël Candau dans un entretien mené par Michel Roland, « Gestion des données de la recherche dans un projet de recherche multidisciplinaire », Urfist de Nice - Université Nice Sophia Antipolis, mai 2016

<https://doranum.fr/2016/10/12/gestion-des-donnees-de-la-recherche-dans-un-projet-de-recherche-multidisciplinaire/>

⁹ Joël Candau, Eve Bureau, Karine Durand, Céline Geffroy, Véronique Ginouvès, et al.. Une approche bioculturelle du premier aliment du nouveau-né. Le colostrum. *Premiers cris, premières nourritures*, Presses Universitaires de Provence, pp.123-153, 2019, *Premiers cris, premières nourritures*, 979-1032002049. CC-BY [en ligne] <https://hal-amu.archives-ouvertes.fr/hal-02126733/>

¹⁰ Présentation de Sharedocs : <https://documentation.huma-num.fr/1024>

¹¹ Dropbox est un service de stockage et de partage de copies de fichiers locaux en ligne proposé par Dropbox, Inc., entreprise localisée à San Francisco, en Californie. L'espace gratuit est lié

Ces données sont réparties au sein de quatre répertoires et stockées sur le serveur de la phonothèque de la MMSH :

- Traitement-phonothèque-MMSH
- Documents-de-travail
- Dossier-scientifique-et-administratif
- Reunions-et-restitutions

Les captures d'écran suivantes donnent un aperçu du contenu de ces répertoires :

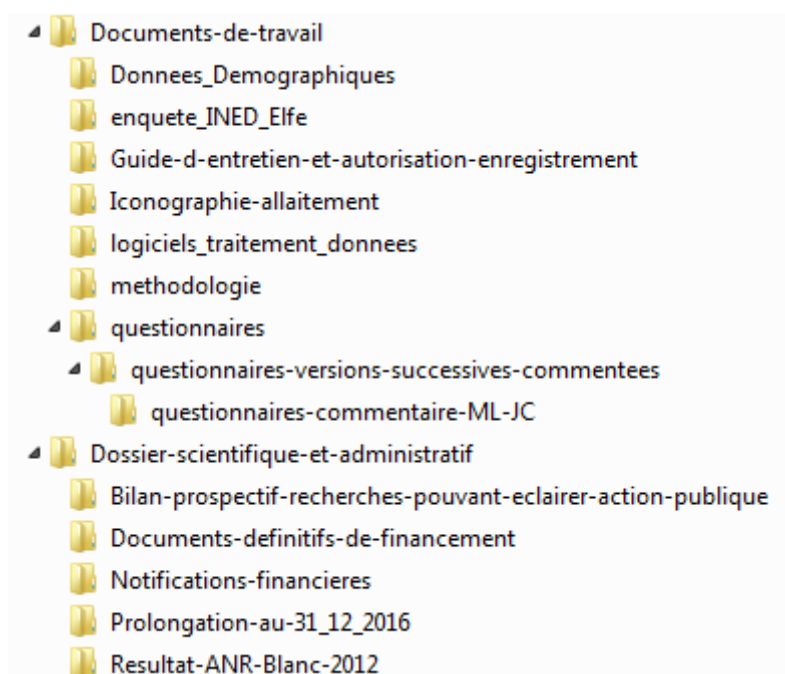


image 1 : contenus des répertoires "Documents-de-travail" et "Dossier-scientifique-et-administratif"

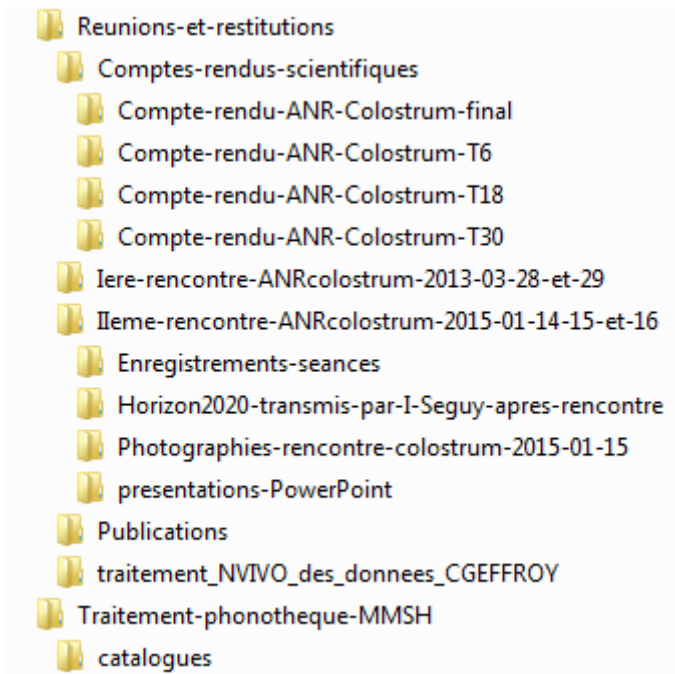


image 2 : contenus des répertoires “Reunions-et-restitutions” et “Traitement-phonotheque-MMSH”

Un tableur en annexe de ce document liste tous les fichiers contenus dans ces répertoires et sous-répertoires¹². Les noms des répertoires comme ceux des fichiers textes suivent des règles de nommage simples (pas d’espaces, pas de caractères accentués, de ponctuation ou exotiques). Il n’y a pas eu de plan de nommage constitué au départ car, excepté pour les archives sonores qui suivent le plan de nommage de la phonothèque, l’ensemble a été rassemblé à des périodes variées, avec des modalités de création hétérogènes et dans des langues multiples.

3. Le travail en amont

Présente dès le début du projet, portée par une préoccupation de Joël Candau, la question de l’*open science* est née d’une réflexion en anthropologie sur les sources premières. Ainsi, le programme “ANR Colostrum” a intégré cet objectif en prévoyant la gestion des données produites au cours des recherches, avant même que cela ne devienne une obligation pour les projets financés par l’ANR. Dans l’entretien mené auprès de Joël Candau par Michel Roland, le professeur en anthropologie souligne l’importance de “verrouiller” en amont la gestion des données, afin d’avoir une cohérence et une homogénéité au moment du traitement et de l’archivage¹³. Dans cette optique, Véronique Ginouvès de la phonothèque de la MMSH et Thierry Rosso du LASMIC (Université de Nice-Sophia Antipolis) ont tenu une formation auprès des

¹² Annexe 1 : Etat des lieux des données contextualisant la recherche [en ligne] <https://www.nakala.fr/nakala/data/11280/50d3e682>

¹³ *idem*

chercheurs participants, à l'occasion de la deuxième rencontre des partenaires de l'ANR Colostrum en janvier 2015¹⁴.

Ainsi, au moment du lancement du projet, plusieurs principes ont été transmis aux chercheurs lors des journées de présentation du programme¹⁵:

1. Le format des fichiers son pour la conservation et la méthode d'enregistrement pour obtenir une qualité de son supérieure était le wave
2. Tout élément d'enquête (enregistrement, photographie, prise d'information) devait faire l'effet de l'obtention d'un consentement par les informateur·trice·s et les institutions au sein desquelles les informations étaient collectées. Les questions juridiques et éthiques qui pouvaient se poser au moment de l'enregistrement ou de la collecte d'image comme d'information devaient *a minima* être décrites, et résolues dans la mesure du possible. Un contrat d'utilisation rédigé spécifiquement pour le programme a été diffusé.
3. La transmission de l'information sur le contexte de production des documents (enregistrements, images, documents connexes) ainsi que les éléments nécessaires au catalogage des entretiens à fournir au moment (date, lieu, nom des informateurs/trices, contexte d'enregistrement, traduction en langue française si nécessaire, contrat d'utilisation) devaient être fournis aux archivistes par les anthropologues.

Prévoir la gestion des données signifie inviter les chercheurs à renseigner au mieux les données qu'ils produisent en vue de leur documentation et de leur archivage (1). Il s'agit aussi de définir quels formats informatiques seront les mieux adaptés pour la production de données numériques (2). Enfin, il s'agit de prendre en compte toutes les composantes éthiques et juridiques susceptibles d'interférer lors de la diffusion de ces données (3).

(1) Contextualiser une enquête de terrain dès sa réalisation permet un traitement plus efficace par la suite. Il est ainsi recommandé de préciser systématiquement certaines informations sur les documents collectés sur le terrain (en particulier les enregistrements audio et vidéo et les photographies) :

- le lieu de l'enquête
- la date de l'enquête
- les noms et prénoms des intervenants, avec quelques éléments sur leur biographie (âge, profession, origines, langues parlées...)
- le contexte de l'enquête
- les informations juridiques et éthiques

¹⁴ Véronique Ginouvès et Thierry Rosso, « Bonnes pratiques pour l'archivage des sources de l'ANR Colostrum », 16 janvier 2015. Le diaporama de la formation a été déposé en libre accès sur Nakala : <https://www.nakala.fr/nakala/data/11280/a56a5914>

¹⁵ Programme des journées des 28 et 29 mars 2013 : <https://colostrum.hypotheses.org/368>

- pour un entretien, les grandes lignes de l'échange si l'entretien n'a pas été fait en langue française

(2) Outre les éléments de contextualisation des enquêtes de terrain, il faut aussi veiller à fournir des fichiers numériques de bonne qualité. Cette qualité permettra un traitement plus simple des données produites, en évitant de manipuler à nouveau les fichiers numériques pour les conformer aux standards de conservation¹⁶. Plusieurs points de vigilance à prendre en considération :

- qualité du magnétophone et du micro = Zoom H4
- qualité de la prise de son (orientation du micro, maîtrise du fonctionnement de l'appareil, vérification des niveaux)
- qualité du format d'enregistrement = WAVE (44.1 Hzt, 16 Bits, format non compressé)
- carte son de bonne capacité
- les sons perturbants pendant l'entretien
- la charge du magnétophone (piles ou transformateur)
- nommage simple du fichier (ex : AAAMMJJ-N+1)
- copie sur deux supports différents et préservation intégrale de l'enregistrement original
- pour les photographies, utilisation d'un format non compressé (TIFF)

(3) L'attention portée aux questions juridiques et éthiques dans la collecte de données sur le terrain permet une meilleure diffusion des données, notamment en évitant un long travail de recherche des ayants-droits (travail qui n'aboutit pas toujours).

Au cours du programme, plusieurs documents ont ainsi été élaborés avec pour objectif d'anticiper les problèmes éthiques et juridiques :

- Un guide d'entretien¹⁷, qui rappelle la façon dont mener un entretien en anthropologie et les éléments à reporter pour anticiper la gestion des données : nécessité d'enregistrer le "consentement éclairé" de l'informateur (à l'oral ou par voie de contrat écrit), présentation du programme aux informateurs, tenue d'une fiche descriptive de l'entretien, choix du nommage des fichiers enregistrés...
- Des exemples de contrats d'autorisation pour l'enregistrement, l'exploitation et la diffusion des données : photographies¹⁸, enregistrements sonores

¹⁶ Le CINES liste les formats validables pour la conservation des données numériques. Voir FACILE - Service de validation de formats : <https://facile.cines.fr/>

¹⁷ Le guide d'entretien est téléchargeable et consultable en ligne : <https://www.nakala.fr/nakala/data/11280/575a9ea1>

¹⁸ « Demande d'autorisation d'utilisation de l'image d'une personne dans le cadre de l'ANR COLOSTRUM Volet France » : <https://www.nakala.fr/nakala/data/11280/56c4bec0>

- Des questionnaires, traduits dans les différentes langues employées pour les enquêtes
- Des contrats de dépôt entre les chercheurs et la phonothèque de la MMSH.

4. Le traitement et l'archivage des données

Le traitement documentaire des données anthropologiques s'est fait à la phonothèque de la MMSH en lien avec Thierry Rosso de l'équipe de Joël Candau (LASMIC, EA 7278, Université de Nice Sophia Antipolis), au cours de missions et stages effectués entre 2015 et 2017 : Marine Soubrié (ingénieur d'études, CDD financé par l'ANR, à la phonothèque de la MMSH), Anissa-Claire Adgharouamane (stage financé par le CNRS à la phonothèque de la MMSH), Giulia Manfredi (stage financé par l'ANR à l'Université de Nice sous la direction de Thierry Rosso).

Les moyens alloués au traitement correspondent à l'aide accordée par l'ANR à la MMSH. Cette aide s'élevait à 26 395 euros.

Les données contextualisant la recherche ont été traitées en septembre 2020 au moment de la rédaction du plan de gestion de données par Claire Cialone-Grégoire sur un financement du Consortium Archives des ethnologues.

Le traitement documentaire a donné lieu à deux instruments de recherche en format XML :

- en format Dublin Core (XML) sur Ganoub, la base de données de la phonothèque : <http://phonotheque.mms.huma-num.fr/dyn/portal/index.seam?page=alo&alold=10820&fonds=0&nat=3&cid=338>
- en format EAD (XML) sur Calames (Catalogue en ligne des archives et manuscrits de l'enseignement supérieur, développé par l'[ABES](http://www.abes.fr)) : <http://www.calames.abes.fr/pub/ms/FileId-2009>

Chaque enregistrement sonore a fait l'objet d'une notice documentaire indexée et d'une mise en ligne.

Par ailleurs, 123 photographies de six des huit "terrains" (Maroc-Sud, Fès, France, Bolivie, Cambodge, Allemagne) ont été déposées sur MédiHAL, l'archive ouverte du CCSD : https://medihal.archives-ouvertes.fr/search/index/?q=anr+colostrum&submit=&submitType_s%5B%5D=file

Lors de la préparation des fichiers pour l'archivage sur le long terme au CINES, il a fallu standardiser les documents textes, les tableurs et les diaporamas PowerPoint. En effet, ces documents étaient enregistrés sous des formats non ouverts (.DOC, XLS, PPT) qui ne font pas partie de la liste des formats validables établie par le CINES (<https://facile.cines.fr/>). Standardiser ces fichiers a consisté en leur

conversion vers des formats ouverts. Ainsi, les documents textes sous format .DOC ont été convertis en format .ODT (Open Document Texte) et les tableurs sous format .XLS ont été convertis en format .ODS (OpenDocument Spreadsheet). Les diaporamas sous format PPT ont quant à eux été convertis en .PDF (Portable Document Format) car le format .ODP (OpenDocument Presentation), “équivalent” ouvert de .PPT, n’est pas retenu par le CINES pour la conservation des diaporamas ; de plus, la conversion de .PPT à .ODP altère sensiblement la forme de la présentation.

Première publication du PGD (travail en cours) :

<https://ethiquedroit.hypotheses.org/3306>