



**HAL**  
open science

## Plan de gestion des données du projet ANR DIASCO-TIB (Convergences diasporiques : le cas des réfugiés tibétains Acronyme)

Camille Simon, Xénia de Heering, Françoise Robin, Anne-Sophie Bentz

### ► To cite this version:

Camille Simon, Xénia de Heering, Françoise Robin, Anne-Sophie Bentz. Plan de gestion des données du projet ANR DIASCO-TIB (Convergences diasporiques : le cas des réfugiés tibétains Acronyme). 2024. hal-04566967

**HAL Id: hal-04566967**

**<https://hal.science/hal-04566967v1>**

Preprint submitted on 2 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# PGD DU PROJET "CONVERGENCES DIASPORIQUES : LE CAS DES REFUGIES TIBETAINS"

Plan de gestion des données basé sur le modèle ANR structuré fourni par l'Agence nationale de la recherche (ANR) en date du 09/04/2024. ([https://dmp.opidor.fr/public\\_templates#content](https://dmp.opidor.fr/public_templates#content))

## RENSEIGNEMENTS SUR LE PLAN

Titre du plan	PGD du projet "Convergences diasporiques : le cas des réfugiés tibétains "
Version du PGD	Version initiale
Domaines de recherche (selon classification de l'OCDE)	5.4 sociologie ; 5.9 autres sciences sociales ; 6.2 langues et littérature
Langue de rédaction du plan	fra
Responsable du plan	Camille Simon, Xénia de Heering
Date de création	2024-05-02
Type d'identifiant	
Licence	CC4.0 BY-NC-SA

## RENSEIGNEMENTS SUR LE PROJET

Titre du projet	Convergences diasporiques : le cas des réfugiés tibétains
Acronyme	Diasco-Tib
Résumé	<p>DiasCo-Tib propose d'analyser des formes de convergences linguistique, spatiale et sociale à l'œuvre au cours d'un « moment diasporique », c'est-à-dire un moment crucial de réactivation et de reconfiguration d'une diaspora. La recherche se fonde sur le cas des réfugiés tibétains, qui font actuellement l'expérience d'un tel « moment diasporique », lié à la disparition anticipée de leur chef spirituel, le Dalaï-lama (né en 1935), Les récents mouvements migratoires des réfugiés tibétains de l'Asie du Sud vers l'Europe et l'Amérique du Nord entraînent une reconfiguration spatiale à grande échelle de ce réseau diasporique multipolaire, dont la France est désormais un lieu clef. Notre hypothèse est que, dans le contexte d'un moment diasporique, une dispersion spatiale accrue peut, paradoxalement, intensifier les processus de convergence sociale. L'équipe interdisciplinaire de DiasCo-Tib étudiera des phénomènes sociaux concomitants et évaluera leur degré d'interdépendance dans les domaines des langues et des pratiques linguistiques, des réseaux sociaux et économiques, des formes de représentation collective (dans les sphères politiques, civiques ou encore artistiques), de l'évolution des rôles genrés et des pratiques religieuses. Une recherche multi-située permettra d'observer et comprendre la circulation de normes et pratiques sociales au sein de la diaspora, en prenant en compte l'inscription des réfugiés tibétains dans leurs différentes sociétés d'accueil comme de leurs réseaux sociaux transnationaux. Outre les convergences attendues, DiasCo-Tib identifiera des lignes de segmentation qui se cristallisent dans la reconfiguration de pratiques linguistiques et sociales partagées mais plurielles. L'étude de cas choisie permettra donc de mettre en lumière les multiples dimensions du processus du diasporisation, tels qu'ils sont vécus et mis en œuvre par les individus et les communautés dans la vie quotidienne et tout au long des parcours de vie.</p>
Source de financement	Agence Nationale de la Recherche : ANR-23-CE41-0017
Date de début	2024-02-01
Date de fin	2028-01-31
Partenaires	IFRAE (UMR 8043) CESSMA (UMR 245)

**Produits de recherche :**

Jeux de données collectées par les membres du projet Diasco-Tib

Jeux de données multimédia et multilingues produites par les membres dans le cadre du projet Diasco-Tib

**Contributeurs**

Nom	Affiliation	Rôles
Françoise Robin	IFRAE	Coordinatrice du projet
Camille Simon	LACITO	Responsable du PGD Personne contact pour les données
Xénia de Heering	CESSMA	Responsable du PGD Personne contact pour les données

**1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES**

**1.1. Description générale du produit de recherche**

Les données produites par le projet Diasco-Tib seront de différentes nature : documents textuels (notes de terrain et d'entretiens manuscrites et numériques), enregistrements audio et audiovisuels, documents visuels (photos, schémas, documents cartographiques), tableaux et bases de données.

Une sélection de ces données sera déposée sur des plateformes d'archivage au cours et à l'issue du projet : les données textuelles, tableaux et bases de données seront déposées sur HAL, les images seront déposées sur Nakala et Medi-HAL, les données audio seront déposées sur Pangloss, et les documents audiovisuels seront déposés selon les cas soit sur Pangloss, soit sur Nakala.

Les identifiants pérennes seront attribués lors du dépôt des données sur les plateformes d'archivage : DOI sur Nakala et Pangloss et idHal sur HAL et Medi-Hal. L'archive Pangloss prévoit en outre l'attribution d'un DOI pour le document audio complet, mais aussi pour chaque segment annoté.

Le dépôt des données dans ces archives ouvertes permettra leur réutilisation ultérieure. Les métadonnées et annotations prévues (voir *infra*) faciliteront leur identification et leur réutilisation par des chercheurs dans diverses disciplines. Par exemple, la mention des codes ISO-639-3 et Glottolog pour des documents audiovisuels produit dans le cadre du volet socio-anthropologique des investigations permettra, grâce à ce référencement conforme aux standards internationaux, leur repérage par des linguistes, et contribuera ainsi à augmenter l'accessibilité de ressources pour une langue peu documentée comme l'est le tibétain parlé.

**1.2. Est-ce que des données existantes seront réutilisées ?**

Nous envisageons de réutiliser des données issues de trois types de sources différentes :

- (1) Des données disponibles en accès ouvert en ligne. Il s'agit principalement d'annuaires (pages jaunes, répertoire national des associations, répertoires d'entreprises en ligne comme la base SIRENE de l'INSEE), de sources numérisées disponibles en ligne (articles de presses, images photographiques ou filmées), et des données bibliographiques issues de catalogues de bibliothèques etc.
- (2) Des données issues d'archives officielles pour lesquelles une autorisation d'accès dans le cadre du projet de recherches devra être demandé. Il s'agit notamment des archives administratives de l'OFPPRA, des archives de l'INED (en particulier l'enquête « trajectoires et origines »), d'archives municipales, d'archives préfectorales, OFII...
- (3) Des données issues d'archives commerciales et associatives, pour lesquelles des accords écrits de consultation et de réutilisation seront établis et signés par les porteuses du projet et les responsables de ces structures. La réutilisation de ces données demandera un processus d'anonymisation préalable.

En outre, quelques jeux de données issues de recherches antérieures des membres du projet, et en particulier celles collectées lors de la phase préparatoire du projet (entretiens audio transcrits, documents visuels) seront réutilisés : la réutilisation de ces données ne présente pas de contraintes particulières.

### 1.3. Comment seront produites/collectées les nouvelles données ?

Les nouvelles données seront principalement collectées et produites par les membres de l'équipe au moyen d'enquêtes qualitatives, d'enquêtes par questionnaires, et de recherches documentaires. Ces données incluent des documents textuels, des enregistrements audio et vidéo, des documents iconographiques (photos, schémas, cartes...), et des bases de données.

Le traitement et l'annotation des données se feront principalement à l'aide de logiciels libres. Les logiciels suivants seront utilisés :

- (1) ELAN pour la transcription et l'annotation linguistique des données audio et audio-visuelles
- (2) Audacity pour le traitement et le montage des fichiers audio
- (3) PRAAT pour l'analyse de phonologique
- (4) Tropy pour la gestion des corpus iconographiques
- (5) Finalcut Pro pour le traitement des données audiovisuelles
- (6) Zotero et JurisM pour la gestion des données bibliographiques
- (7) QGIS pour traiter les données géoréférencées
- (8) Des logiciels de tableurs (Calc et Excel)
- (9) Des logiciels de traitement de texte (Open Office et Word)
- (10) MIRO (pour réaliser des schémas)

Les métadonnées préciseront systématiquement, outre le nom du chercheur déposant : a) pour les données nouvelles produites dans le cadre du projet, la date et le lieu de production ; b) pour les données réutilisées, la source et la date de collecte.

## 2. DOCUMENTATION ET QUALITE DES DONNEES

### 2.1. Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?

Les archives choisies pour déposer les données du projet respectent les standards internationaux pour répondre aux exigences de la science ouverte et des données FAIR. Les métadonnées demandées par ces plateformes d'archives seront donc systématiquement enregistrées au cours de la recherches, pour l'ensemble des données produites.

- (1) « La Collection Pangloss a d'emblée été conçue comme un maillon dans un dispositif mondial autour de normes pour les métadonnées, en particulier OLAC (Open Language Archive Community), et de protocoles ouverts de partage (OAI-PMH : Open Archives Initiative Protocol for Metadata Harvesting). » (Aurelia Vasile, Séverine Guillaume, Mourad Aouini, Alexis Michaud). Le Digital Object Identifier, une impérieuse nécessité ? L'exemple de l'attribution de DOI à la Collection Pangloss, archive ouverte de langues en danger. *I2D – Information, données & documents*, 2020, 2, pp.156-175. ([halshs-02870206](https://halshs.archives-ouvertes.fr/halshs-02870206))
- (2) Nakala utilise le modèle Dublin Core qualifié pour décrire les données.
- (3) Les principes fondateurs de HAL et Medi-HAL garantissent un archivage FAIR (<https://about.hal.science/principles/>)

Les langues des documents seront systématiquement identifiées par les codes ISO-639-3 et Glottolog, en usage parmi les linguistes.

Au cours du projet, les données seront stockées sur Sharedocs (TGIR HumaNum), qui permet l'enregistrement systématique de métadonnées selon le format Dublin Core. Ces indications permettront de classer et gérer efficacement les jeux de données et d'en différencier les versions successives. De plus, pour les données partagées entre les membres de l'équipe, des conventions de nommage communes ont été définies.

Pour faciliter la réutilisation interdisciplinaire des données archivées, outre les métadonnées prévues par les formats standards mentionnés ci-dessus, des informations supplémentaires pourront être fournies dans un document lié par sa convention de nommage au jeu de données principales. Ces informations pourraient être, par exemple, le niveau de scolarisation ou quelques éléments clés du parcours migratoire.

## **2.2. Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?**

Les documents audio et vidéo seront enregistrés en formats non-compressés (.wav et .mkv), pour garantir une bonne qualité d'enregistrement. Les membres du projet qui possèdent leur propre matériel d'enregistrement (acquis à l'occasion de projets de recherche antérieurs) le réutiliseront. D'autres appareils (notamment une caméra, des dictaphones) seront empruntés ponctuellement au Lacito UMR7107 dont le personnel d'appui à la recherche pourra aider les membres du projet pour les questions techniques.

Les membres du projet prévoyant de la documentation photographique veilleront à paramétrer leurs appareils au format haute qualité.

Les protocoles de recueil des données dans le cadre des enquêtes qualitatives menées par les membres du projet seront discutées lors des séances du séminaire interne Diasco-Tib. Des discutants externes y seront régulièrement invités.

## **3. EXIGENCES LEGALES ET ETHIQUES, CODE DE CONDUITE**

### **3.1. Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?**

Le projet Diasco-Tib respecte le Règlement Général sur la Protection des Données (RGPD) de l'Union Européenne et les principes directeurs d'une recherche responsable. Les membres du projet ont tous déjà travaillé avec des réfugiés tibétains avec lesquels ils entretiennent des relations de confiance et de longue date, et sont conscient de la nécessité de traiter les données personnelles d'une manière éthique et culturellement sensible. Les personnes enquêtées étant des réfugiés politiques, une attention particulière sera accordée à l'anonymisation des données dans le contexte de l'ingérence du gouvernement chinois parmi les "minorités exilées sensibles" telles que les réfugiés tibétains et ouïghours. De plus les membres du projet respectent la confidentialité des données sur le terrain (ne pas divulguer les informations personnelles dans le réseau d'interconnaissances) et durant le processus de recherche (respect de la confidentialité des données). Enfin, chaque fois que nécessaire, les données seront stockées dans des fichiers cryptés, protégés par mot de passe.

L'accès aux données personnelles recueillies sera aussi restreinte que possible : en tout état de cause, aucune personne extérieure au projet n'aura accès à ces données. En outre, afin de minimiser les risques de fuite ou de perte de ces données, les données personnelles seront traitées par chaque chercheur individuellement ou organisées en sous-groupe de travail (notamment dans le cas de la cartographie de la population tibétaine en France, 1<sup>ère</sup> tâche « Mapping the Tibetan Diaspora in France » du WP1).

Le consentement éclairé des personnes enquêtées pour le recueil, la préservation et/ou le partage de données sera recueilli selon le protocole suivant : une explication orale détaillée du projet de la part des membres du projet, la signature d'un formulaire de consentement en anglais, français ou tibétain ou l'enregistrement du consentement (notamment pour les personnes illettrées). Le formulaire de consentement sera validé par le service juridique de l'Inalco, qui pourra également appuyer le projet pour tout autre question juridique.

Le choix de pseudonymiser ou d'anonymiser les métadonnées dépendra du contenu des données recueillies. Par exemple, pour des entretiens portant sur des sujets anodins, non sensibles politiquement, les personnes enquêtées pourront choisir d'apparaître dans le projet sous leur nom, sous pseudonyme, ou de façon anonyme.

Pour d'autres jeux de données, par exemple ceux issus de la cartographie de la population tibétaine en France, ou des entretiens sur des sujets politiquement sensibles, la pseudonymisation sera automatique au cours du projet, et l'anonymisation sera la règle en cas de publication. Les pseudonymes et leur décodage seront conservés sur papier (carnet de recherche).

Les jeux de données pourront être détruits (immédiatement ou à tout moment pendant ou à l'issue du projet) si la personne enquêtée le demande.

Dans le cadre de la cartographie des Tibétains en France, une analyse d'impact relative à la protection des données sera effectuée à l'aide de l'outil proposé par la CNIL pour s'assurer d'un traitement des données conforme au RGPD et respectueux de la vie privée.

### **3.2. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?**

Seule une partie des données produites au cours du projet Diasco-Tib sera sélectionnée pour une préservation sur le long terme via le dépôt dans Nakala, Pangloss, HAL ou Medi-HAL, sous licence Créative Commons. Le choix des données à conserver se fera en concertation entre les membres du projet, la décision finale revenant aux trois coordinatrices. Ce choix sera basé sur l'intérêt scientifique des données et les potentialités de réutilisation, mais tiendra compte aussi des restrictions par le caractère confidentiel de certaines données. Ainsi pour cette sélection de données archivées, l'accès à certains jeux de données pourra être protégé par embargo. De cette façon, les données seront gérées suivant le principe « aussi ouvert que possible, aussi fermé que nécessaire ».

Les droits de propriété intellectuelle sur les données du projet et leurs dérivés restent acquis par les différentes parties ayant créé ou produit ces données. Cependant, les différents membres du programme de recherche Diasco-Tib partageront entre eux une partie des données produites. Dans le cadre de l'utilisation des données à des fins de publications, l'auteur sera toujours crédité pour les données qu'il a collectées si elles sont utilisées de façon ponctuelle par un collègue dans une publication.

Toute publication ou communication scientifique réalisée dans le cadre du projet créditera ainsi l'auteur ou les auteurs des données, ainsi que s'ils le souhaitent et si les données n'ont pas été anonymisées (voir *supra*), les personnes enquêtées auprès de qui ces données ont été recueillies. En effet, les personnes ayant choisi d'apparaître sous leur nom dans les différentes enquêtes pourront être créditées comme co-auteurs des produits de la recherche.

Les données ou jeux de données déposés par les membres du projet Diasco-Tib sur les plateformes Nakala, Hal, Medi-Hal et Pangloss se verront attribuer des identifiants persistants (DOI et idHAL), par lesquels ils pourront être cités dans les publications scientifiques et les productions de valorisation de la recherche à destination du grand public.

D'éventuelles restrictions sur la réutilisation de données fournies par des tiers ne sont pas connues actuellement et seront documentées, si nécessaire, dans les prochaines versions du plan de gestion des données.

### **3.3. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?**

Les modalités de stockage et le partage des données au cours du projet, ainsi que la publication des données en accès ouvert à l'issue du projet tiendront compte des questions éthiques et des nécessités de protéger la confidentialité des données personnelles (voir *supra*). On se référera notamment au guide « Pratiquer une recherche intègre et responsable » publié par le comité d'éthique (COMET) du CNRS en 2017.

En complément des principes FAIR pour l'archivage des données, les membres du projet Diasco-Tib tiendront également compte des principes CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics, <https://www.gida-global.org/care>) de gouvernance des données issues de recherches avec des populations minorisées. Ces questions d'éthique et de déontologie sont régulièrement abordées au fil du déroulement du projet, lors des séminaires internes.

## **4. TRAITEMENT ET ANALYSE DES DONNEES**

### **4.1. Comment et avec quels moyens seront traitées les données ?**

Les données seront analysées à l'aide des méthodes d'analyse quantitatives et qualitatives propres aux disciplines de sciences sociales des membres du projet. Le traitement prévu ne nécessite qu'un équipement informatique de base et aucun logiciel propriétaire particulier.

Le traitement et l'analyse des données sera principalement effectué par les membres du projet. Certaines tâches ponctuelles comme la transcription d'enregistrements pourront être confiées à des prestataires externes. Le coût lié à ces prestations externes a été prévu dans le budget initial, de même que le coût de l'équipement informatique.

## **5. STOCKAGE ET SAUVEGARDE DES DONNEES PENDANT LE PROCESSUS DE RECHERCHE**

### **5.1. Comment les données seront-elles stockées et sauvegardées tout au long du projet ?**

Pour assurer la pérennité des données au cours du projet les membres déposeront leurs données régulièrement sur un compte projet Sharedocs (voir *infra*). Le compte projet Sharedocs sera prioritairement dévolu au stockage des données destinées à être immédiatement partagées avec les autres membres du projet. L'espace individuel de stockage permettra également de stocker les données non-partagées, en complément des plateformes numériques individuelles mises à disposition par les institutions de chaque chercheur. Ces sauvegardes seront effectuées individuellement par les chercheurs au moins une fois par mois.

En parallèle, des sauvegardes sur disque-durs externes seront également effectuées et les documents audiovisuels seront initialement enregistrés, puis conservés, sur les cartes mémoire des appareils, ce qui constituera une sauvegarde supplémentaire. La multiplication des supports de sauvegarde permettra de minimiser le risque de perte des données.

Une partie des données sera recueillie au format papier (notes de terrains, schémas etc.) et conservés par les chercheurs au cours du projet. Ces documents pourront être numérisés au cas par cas, et stockés informatiquement.

L'accès aux données stockées sur Sharedocs au cours du projet sera restreint aux seuls membres du projet, ce qui en garantira la sécurité. De plus, la gestionnaire du compte-projet sur Sharedocs en accord avec les trois responsables scientifiques du projet déterminera des droits d'accès spécifiques pour chacun des dossiers stockés.

La sécurité des données à caractère personnel sera assurée par une pseudonymisation ou une anonymisation qui pourra intervenir, si nécessaire, dès le moment du recueil des données sur le terrain, et au plus tard au moment où les données seront rendues publiques. Dans le cas où des données à caractère sensible ne pourraient pas être suffisamment anonymisées ou que leur publication n'aurait pas été autorisée par l'interlocuteur ou pourrait lui être préjudiciable, elles ne seront pas utilisées et seront détruites.

Le recueil et le traitement des données sera effectué conformément à la RGPD, et dans le respect des codes déontologiques de la recherche scientifique. De plus, la population au cœur de notre étude est une population de réfugiés politiques, ce qui impose de prendre des précautions particulières, pour protéger les données personnelles recueillies au cours de l'enquête. Si nécessaire, en fonction de la nature sensible des données, les fichiers seront chiffrés par mots de passe. Les services juridiques des institutions partenaires du projet apporteront leur appui pour assurer la meilleure sécurité possible des données tout au long du processus de recherches.

## 6. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

### 6.1. Comment les données seront-elles partagées ?

Les données produites dans le cadre du projet Diasco-Tib seront codées selon des formats standards et ouverts, facilitant le partage et la réutilisation à long terme des données, en privilégiant l'usage de formats issus de logiciels libres. Certaines données seront recueillies sur papier. Pour les documents numériques, les formats utilisés seront :

- **Vidéos** : **.mp4** et **.mkv** (un format compressé adapté au partage et à la diffusion des fichiers et un format non-compressé garantissant la qualité des données) ; combinés, le cas échéant, à des fichiers d'annotations en **.xml**
- **Documents audio** : **.wav** et **.mp3** (un format compressé adapté au partage et à la diffusion des fichiers et un format non-compressé garantissant la qualité des données) ; combinés, le cas échéant, à des fichiers d'annotations en **.xml**
- **Images** : le format **.jpeg**
- **Textes** : les formats **.docx** et **.odt**
- **Bases de données** : les bases de données constituées à l'aide de Zotéro seront enregistrées au format **.rdf** et les autres bases de données sous forme de tableurs seront aux formats **.xlsx**, **.ods** et **.csv**
- Les documents combinant différents types de données seront conservés au format **.pdf**.

Dans ces documents, les textes écrits en tibétain seront encodés à l'aide de polices de caractère Unicode (UTF-8).

Les volumes nécessaires estimés par types de documents ont vocation à être réévalués dans les versions ultérieures du plan de gestion des données. L'essentiel des besoins en espace de stockage concernera les documents vidéo, audio et les images.

- **Vidéos** : environ 45 heures de rushes
- **Enregistrements audio** : environ 150 heures
- **Images** : 1 800 (comprenant des photographies, des schémas et des documents de type cartographique)
- **Textes, annotations et documents combinant différents types de données** : à évaluer à un stade ultérieur du projet
- **Bases de données** : à évaluer à un stade ultérieur du projet

Une demande d'accès au service HumaNum Box sera faite pour permettre le stockage intermédiaire des données les plus volumineuses

Du fait de la capacité du compte projet sur Sharedocs actuellement limité à 300 Go par la TGIR HumaNum, un premier tri sera effectué par les chercheurs pour déterminer les éléments devant être stockés prioritairement. Parmi toutes ces données sauvegardées une seconde sélection sera effectuée pour déterminer lesquelles auront vocation à être déposées dans un entrepôt de données et (éventuellement), rendues publiques.

Pour les données qui seront rendues publiques via les archives ouvertes (HAL, Medi-HAL, Nakala et Pangloss) le moissonnage des métadonnées mises à disposition de manière interopérable est prévu par les infrastructures de recherche qui administrent ces entrepôts. Cela permet l'exploitation et le référencement des données déposées par des agrégateurs externes tels qu'Isidore ou OLAC. Des identifiants persistants (notamment DOI et idHAL) sont attribués automatiquement lors du versement des données dans ces entrepôts.

Une partie des données sélectionnées pour la publication sera rendue disponible au cours du projet : notamment une partie du corpus linguistique qui sera mis en ligne dans le courant de l'année 2026, ainsi que des données valorisées à travers le blog Hypothèses. A l'issue du projet, les membres proposeront une sélection définitive de données qui seront archivées et mises à disposition. Cette sélection pourra comprendre des données, qui seront placée sous embargo (détails à définir ultérieurement), soit parce qu'il s'agit de données sensibles, soit dans l'attente d'une publication de travaux scientifiques à venir : dans ce cas, seules les métadonnées seront rendues publiques. Comme précisé plus haut, les données trop sensibles pour être partagées ne seront pas déposées dans des entrepôts de données : aucune restriction d'accès ou politique de contrôle d'accès ne sera donc à prévoir pour les données déposées sur Nakala, Medi-HAL, HAL et Pangloss.

Les données rendues publiques sans restriction d'accès seront réutilisables par tout un chacun, puisqu'elles seront publiées sous licence Créative Commons (voir *supra*). L'équipe prévoit d'inclure des métadonnées et des annotations ou mots-clés en langue tibétaine ce qui facilitera leur accessibilité aux populations enquêtées. En effet, il est prévu que les données puissent être réutilisés non seulement par les membres du projet dans leurs publications scientifiques et plus largement par la communauté scientifique en général, mais aussi par les Tibétains eux-mêmes. L'archivage en formats ouverts permettra de garantir autant que possible une accessibilité à long terme, sans recours à des logiciels propriétaires.

## **6.2. Comment les données seront-elles conservées à long terme ?**

Tout au long du projet, les membres se concerteront lors des séances du séminaire interne pour déterminer quelles données méritent d'être conservées à long terme, en fonction de leur intérêt scientifique et documentaire (caractère inédit, valeur et pertinence pour des démarches comparatives futures...).

A terme, les responsables du projet pourront se rapprocher d'un service public d'archives pour y déposer le reste des archives du projet (comptes-rendus de réunions etc.).

Les données déposées dans Pangloss font l'objet d'une archive pérenne puisqu'il s'agit d'une archive ouverte de langues en danger ou peu documentées. Cet archivage se fait au CINES via la plate-forme CoCoon et les services Huma-Num. Pour ce qui est des données partagées via Medi-HAL ou HAL, l'archivage au CINES est également automatique ; pour les données déposées sur Nakala, une demande d'archivage à long terme sera adressée au cas par cas à [cogrid@huma-num.fr](mailto:cogrid@huma-num.fr) par les coordinatrices du projet, en fonction de l'intérêt scientifique des données.

Conformément aux principes déontologiques exposés en 3., à la demande des personnes enquêtées, des données pourront ne pas être divulguées, voire être détruites.