



**HAL**  
open science

# Les données en sciences du patrimoine : le cas d'entretiens audiovisuels réalisés auprès d'archéologues

Gwendoline Torterat

► **To cite this version:**

Gwendoline Torterat. Les données en sciences du patrimoine : le cas d'entretiens audiovisuels réalisés auprès d'archéologues. 2020. hal-03106624

**HAL Id: hal-03106624**

**<https://hal.science/hal-03106624>**

Preprint submitted on 11 Jan 2021

**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 4.0 International License

# LES DONNÉES EN SCIENCES DU PATRIMOINE

## ATELIER DOPAMINE

### **Le cas d'entretiens audiovisuels réalisés auprès d'archéologues Gwendoline Torterat**

#### **INFORMATIONS GÉNÉRALES**

##### **ANTHROP'ARC**

De la production du savoir scientifique à l'appropriation sociale du passé - Le cas des fouilles du gisement paléolithique à Arcy-sur-Cure (Yonne)

Financement : Région Ile de France/DIM/Matériaux Anciens et Patrimoniaux

Appel à projets : AAP 2019-5

Contrat : Post-doctorat (2019-11/2020-10)

Axes de recherche : usages, archives, réflexivité

Responsabilité scientifique : Nejma Goutas et Baptiste Buob

Partenariat : ArScAn ; LESC ; Service des archives de la MSH Mondes ; Musée de l'Avallonnais

Le projet ANTHROP'ARC porte sur des questions épistémologiques et historiographiques concernant la naissance et l'essor d'une Préhistoire à la française héritée du préhistorien André Leroi-Gourhan et ce, à travers une recherche conduite autour d'un gisement archéologique fouillé depuis la fin du XIX<sup>e</sup> siècle et fréquenté depuis des siècles : les grottes d'Arcy-sur-Cure (Yonne).

Le modèle de plan de gestion de données que je vais vous présenter est destiné à préparer l'organisation et la mise en valeur des données brutes que je récolte au cours de cette enquête ; une enquête qui se base principalement sur des entretiens enregistrés soit de façon audio, soit de façon audiovisuelle (selon le souhait de mes interlocuteurs) auprès de préhistoriens ayant fouillé dans les grottes d'Arcy-sur-Cure entre 1946 et 1963. Pour chacun, un formulaire d'autorisation de captation et de diffusion d'images et de son a été créé. Signé et daté en deux exemplaires, il est une première garantie légale en terme de diffusion des données.

Le plan de gestion que j'ai choisi a été mis en place d'abord pour l'archéologie puisqu'il a été élaboré dans le cadre du groupe de travail du consortium « Mémoires des archéologues et des sites archéologiques » (MASA). Il est disponible sur DMP OPIDoR : <https://dmp.opidor.fr/plans>. Basé sur une version initiale réalisée à l'INRAP en 2018, il intègre dans ses recommandations les principaux standards et bonnes pratiques applicables à l'archéologie, ainsi que des liens vers une documentation de référence.

Ce PDG est intégré à l'outil d'aide à la création en ligne de plan de gestion de données DMP OPIDoR hébergé et géré par l'INIST (Institut de l'Information Scientifique et Technique).

Le modèle MASA permet de décrire les données de manière globale (onglet *Généralités sur les données*) puis, plus finement, par jeu de données ou lot de fichiers. Ces ensembles de données peuvent être constitués en référence à une typologie documentaire (principalement des enregistrements bruts et transcriptions d'entretiens dans mon cas) ou selon des ensembles intellectuellement cohérents (étude céramique, d'anthropologie, etc.). Le modèle comporte trois autres onglets pour organiser les renseignements à consigner. Les informations générales sur le projet ou le contexte de production des données pourront être rassemblées dans l'onglet *Contexte*. L'onglet *Gestion des données* permet quant à lui d'expliquer les responsabilités, les moyens et les procédures qualité mis en œuvre pour réaliser cette gestion. Un dernier onglet *À l'issue du projet* permet de décrire les politiques d'archivage et de dissémination envisagées.

## **Description du plan de gestion**

### **Contexte**

- Identification du projet (acronyme ; affiliation du responsable scientifique, institution porteuse du projet, coordinateur personne morale bénéficiaire (nom, pays. Identifiant) ; partenaires du projet ; type de financement ; type de projet)
- Descriptif du projet et mots clés (inscription du projet dans une programmation scientifique financée, chronologie, lieux, sujet ou thème)
- Identification des opérations archéologiques associées (optionnel)
- Historique du document (date de création ; version du document ; nombre de versions prévues ; rédacteur(s) du plan de gestion de données)

### **Gestion des données**

*Informations sur la gestion des données mise en œuvre pour le projet.*

- Responsabilités (responsable(s) de la gestion des données au cours du projet de recherche ; autres intervenants dans la gestion des données)
- Ressources (budget et temps) allouées à la gestion des données.
- Contrôle qualité : quel est le dispositif de contrôle qualité des données et des fichiers mis en place sur ce projet ?

### **Généralités sur les données**

*Les données seront décrites par lot ou jeu de données. Chaque jeu de données sera constitué d'un ensemble de fichiers rassemblés en référence à une typologie documentaire (photographies, images, etc.) ou selon des ensembles intellectuellement cohérents (étude céramique, d'anthropologie, etc.). L'arborescence de classement des fichiers peut servir de base à la constitution des lots. Les données réutilisées en l'état constitueront des lots à part.*

- Constitution du corpus (origine des données)
- Organisation (arborescence de classement des fichiers ; nature des jeux de données décrits)
- Volume (poids attendu et final ; nombre final de fichiers)
- Cadre juridique d'exploitation, de réutilisation et de préservation des données
-

## **Jeux de données**

- **Jeu de données** (nom ; identifiant ; statut ; description ; producteur(s) ; nature ; La production de données a-t-elle nécessité la réutilisation de données existantes par ailleurs ? ; Comment les données ont-elles été produites ? ; format et nommage des données ; traitement des fichiers ; condition de réutilisation)
- **Métadonnées associées** (format ; Quel est leur circuit de production ? ; Avec quel(s) outil(s) sont-elles produites ? ; Quels sont les référentiels utilisés pour les décrire ? ; Y a-t-il une documentation associée aux données ? ; métadonnées d'origine ; métadonnées ajoutées)
- **Stockage des données** (documentation papier produite au cours du projet : volumétrie prévisionnelle ; infrastructure de stockage des données numériques durant le projet)
- **Sécurité des données** (Quels sont les risques ou les menaces qui pèsent sur les données ? ; garantie de confidentialité, d'intégrité et de traçabilité)
- **Accès aux données** (lecture des données ; Comment la disponibilité des données est-elle garantie ? ; échange et partage)
- **Protection des données sensibles au cours du projet** (Y a-t-il des données nécessitant une protection juridique particulière et préciser pourquoi ? Des mesures de protection ont-elles été prises pour protéger les données sensibles ?)

## **À l'issue du projet**

Cette phase vise à décrire les aspects suivants : archivage des données et dissémination des données

- **Politique d'archivage des données produites** (durée de conservation des données produites ; volumétrie finale des données issues du projet ; modalités de conservation des données à l'issue du projet ; modalités d'accès aux données à l'issue du projet)
- **Dissémination des données** (politique de dissémination des données à l'issue du projet ; L'exploitation de ces jeux de données a-t-elle donné lieu à des publications ? ; Certaines données sont-elles sous embargo ? ; potentiel de réutilisation des données (publics cibles, moyens nécessaires...) ; conditions de réutilisation)