



HAL
open science

Le projet Troglolie – La question de la gestion des jeux de données

Anaïs Lamesa

► **To cite this version:**

Anaïs Lamesa. Le projet Troglolie – La question de la gestion des jeux de données. Atelier DOPAMINE – Les données en sciences du patrimoine, May 2020, Paris, France. hal-02593158

HAL Id: hal-02593158

<https://hal.science/hal-02593158>

Submitted on 15 May 2020

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.

Diapo 1

Bonjour,
je souhaite remercier les organisateurs de cette journée, Sophie David et Laurent Romary pour l'organisation de cette table ronde et leur disponibilité.

Je me nomme Anaïs Lamesa et coordonne le projet Troglolie qui se déroule sous la direction scientifique de Marie-Laure Derat et implique deux institutions française, l'université Paris-Sud et le Musée National d'Histoire Naturelle ainsi qu'une société privée, IGECAV. A cette équipe s'est ajouté un quatrième partenaire, l'université d'Addis Ababa. Nous avons, grâce à l'APP valorisation du DIM-Map, pu également associer la société de production Novapista qui doit réaliser un film d'animation.

Ce projet a pour objectif principal d'étudier les dynamiques entre les techniques et les sociétés anciennes. Nous pensons qu'à travers la compréhension du creusement des églises rupestres, certains aspects des communautés chrétiennes les ayant creusés apparaîtront, afin de pallier un manque crucial de sources.

Il se restreint à la province nord de l'Ethiopie, le Tigray et plus précisément le Tigray oriental où de nombreuses églises rupestres datées entre le VIIe et le XXIe siècle ont été creusées dans du grès. Afin de mener dans le temps imparti cette recherche, nous avons décidé avec Marie-Laure Derat de nous concentrer sur une église et de mener une analyse pluridisciplinaire : archéologique, ethnologique, géologique, historique et dendrochronologique. Malheureusement, ce premier programme a été contraint par la situation sanitaire.

Diapo 2

Toutefois j'ai pu, durant un peu plus de deux mois, partir en Ethiopie et mener un mois environ de terrain. De celui-ci j'ai pu extraire différents jeux de données :

- d'images après la visite de sites, de géolocalisation à chaque étape de visite, sonores et vidéos liés aux entretiens menés avec des tailleurs de pierre et manuscrites, donc mes notes et schémas sur mon cahier.

Le traitement de ces données a produit d'autres jeux de données : des données cartographiques avec le logiciel qgis mais je risque de passer sur Arcgis bientôt, des tableurs avec excel, des données issues de DAO principalement avec Photoshop ainsi que des tableurs en format macro par l'intermédiaire d'un logiciel libre Le Stratifiant qui s'appuie sur les formats macro d'excel.

Dans un second temps, si toutefois la mission sur l'église a bien lieu, un dernier jeu de données sera produit, ce sont des nuages de points issus d'images prises par drone pour de la géomorphologie et par un scanner 3D fixe. Ces données seront transformées en modélisation 3D.

Diapo 3

Le cycle de vie de ces jeux de données pose évidemment beaucoup de questions : notamment comment les diffuser dans la mesure où, dans certains cas, je n'ai pas d'autorisations écrites me permettant de les diffuser. Je pense notamment aux photographies de manuscrits ainsi qu'aux vidéos et audios liés aux entretiens de tailleurs de pierre.

De même, certains formats comme le *.xlsm, le *.psd ou encore le *.gpx ne sont pas utilisables sans les logiciels dédiés comme Excel, Adobe ou encore BaseCamp.

Diapo 4

Enfin, où stocker ces données et comment les rendre réutilisables de manière optimum : j'ai, pour cela, réfléchi à différentes options.

M'intégrer à des projets internationaux mais faut-il encore que le retraitement des jeux de données soient possibles, pour les inclure dans ces grandes bases de données.

Dans un cadre plus français, j'ai également pensé à Medihal ou Nakala, qui sont tous deux dans le réseau Isore et donc moissonnés pour apparaître dans les moteurs de recherche dédiés aux chercheurs.

Dans un dernier cas, il m'est également apparu pertinent de les publier dans un journal qui permet le dépôt de données brutes, mais souvent, le comité d'édition privilégie des articles contenant des données 3D, des données cartographiques ou d'analyses quantitatives.

Mes questions concernent donc plusieurs niveaux :

- comment contourner le problème d'autorisations ?
- quelles sont les meilleures options pour diffuser et surtout permettre l'exploitation de ces données dans le futur ?
- enfin de manière plus terre à terre, je n'arrive pas à joindre le consortium ArkeoGis (la demande d'inscription ne fonctionne pas).