



HAL
open science

Plan de Gestion de Données du projet DOM-NAT “ L’invention de la maison durable chez les primo-sédentaires du Proche-Orient ”

Elisa Caron-Laviolette

► To cite this version:

Elisa Caron-Laviolette. Plan de Gestion de Données du projet DOM-NAT “ L’invention de la maison durable chez les primo-sédentaires du Proche-Orient ”. UMR 8068 TempS. 2024. hal-04636976

HAL Id: hal-04636976

<https://hal.science/hal-04636976>

Submitted on 5 Jul 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 4.0 International License

DMP DU PROJET

« L'INVENTION DE LA MAISON DURABLE CHEZ LES PRIMO-SÉDENTAIRES DU PROCHE-ORIENT »

Auteur : Elisa Caron-Laviolette

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "MASA - Modèle de PGD (français) V2" fourni par MASA Consortium.

1. INFORMATIONS GÉNÉRALES

1.1. RENSEIGNEMENTS SUR LE PLAN

Titre du plan	DMP du projet "L'invention de la maison durable chez les primo-sédentaires du Proche-Orient"
Version	Version 2
Objet/périmètre du plan	Le plan porte sur le projet post-doctoral DOM-NAT : "L'invention de la maison durable chez les primo-sédentaires du Proche-Orient : analyse de la matérialité d'un espace domestique pratiqué et transmis au fil des générations", financé par le DIM PAMIR, de février 2023 à juillet 2024
Domaines de recherche (selon classification de l'OCDE)	History and archaeology
Langue	fra
Date de création	2023-02-08
Date de dernière modification	2024-07-05
Licence	<i>Nom</i> Creative Commons Attribution 4.0 International <i>URL</i> http://spdx.org/licenses/CC-BY-4.0.json

1.2.RENSEIGNEMENTS SUR LE PROJET

Titre du projet	L'invention de la maison durable chez les primo-sédentaires du Proche-Orient
Acronyme	DOM-NAT
Résumé	<p>L'adoption durable de la sédentarité par des chasseurs-cueilleurs à la fin du Paléolithique est un tournant majeur des sociétés humaines qui dès lors inventent une nouvelle forme d'habitat, de nouveaux modes de circulation et d'interaction avec le milieu et une nouvelle façon d'être au monde. C'est dans le Natoufien ancien du Levant (13 100-11 600 av. J-C) qu'apparaissent les toutes premières maisons rondes à murs de pierres regroupées en hameaux. L'un des plus remarquables est celui d'Eynan Mallaha, en Galilée, occupé pendant près de 2500 ans. La maison la plus ancienne et la plus massive de ce site, reconstruite à plusieurs reprises, a fait l'objet d'un enregistrement archéologique exemplaire entre 1966 et 1977. Cela permet aujourd'hui un retour sur l'analyse spatiale et diachronique, grâce à des outils et des questionnements scientifiques qui ont largement évolué. Ce projet post-doctoral a pour ambition de reconstituer l'organisation de cet espace domestique durable où un minimum de 5 sols et 3 phases de construction (abris 131-51-62) se superposent. Aujourd'hui, les outils numériques et statistiques permettent, à partir du traitement et de l'analyse des archives archéologiques, de réunir efficacement l'ensemble des données produites sur les objets mobiliers et les éléments architecturaux qui constituent cette succession d'événements. L'objectif est d'abord de mieux définir les périodes d'occupation successives, l'alternance entre vivants et défunts qui partagent cet espace, les phases d'abandon, de restauration et de reconstruction. D'autre part, il s'agit de comprendre la structuration de cet espace : aires d'activité, de repos ou de passage ; placement, circulation et interactions des objets utilitaires et symboliques ; gestion des déchets. Une synthèse comparative permettra de proposer des hypothèses quant aux modalités de construction et d'usage de l'espace et leur transmission au fil des générations, d'en discuter la stabilité et d'identifier ainsi les repères structurant l'espace domestique de ces jeunes sociétés sédentaires.</p>
Sources de financement	Région Île-de-France
Date de début	2023-02-01
Date de fin	2024-07-31
Partenaires	<ul style="list-style-type: none">• Technologie et Ethnologie des Mondes PréhistoriqueS• Maison des Sciences de l'Homme Mondes• Musée d'Archéologie Nationale 02654699X• Modèles et simulations pour l'Architecture et le Patrimoine• Centre de Recherche Français à Jérusalem• Israel Antiquities Authority 0000000406048857

PRODUITS DE RECHERCHE

- Archives papier (Objet physique)
- Documents d'archives numérisés (Jeu de données)
- Documents numériques (Jeu de données)
- Base de données spatiale (Jeu de données)
- Fichiers de code (Jeu de données)

CONTRIBUTEURS

Nom	Affiliation	Rôles
Caron-Laviolette Elisa	UMR 8068 TEMPS	<ul style="list-style-type: none">• Coordinateur du projet• Personne contact pour les données (DOM-NAT_doc_num, DOM-NAT_bdd, DOM-NAT_code, DOM-NAT_scans)• Responsable du plan de gestion de données
Montagne-Bôrras Aurélie	UAR 3225 MSH Mondes	<ul style="list-style-type: none">• Personne contact pour les données (DOM-NAT_archives)

2. CONTEXTE

2.1.DESCRPTION DU PROJET ET MOTS CLÉS

TYPE DE PROJET

Étude

Ce projet post-doctoral a pour objectif de revaloriser les données issues des fouilles archéologiques de l'abri 131-51-62 du site d'Eynan-Mallaha (Israël), considéré comme l'un des sites majeurs du Natoufien. La numérisation et l'analyse des archives de fouille (relevés, notes de terrain et d'analyse du matériel, inventaires détaillés, photographies) permettront de visualiser au mieux ces données primaires acquises sur le terrain, afin d'en mener une analyse fine et de recentrer le débat sur une problématique de fond : comment les Natoufiens construisaient-ils et pratiquaient-ils leur espace au quotidien et sur le temps long à Eynan-Mallaha au début de la sédentarisation ?

STATUT DU PROJET

Projet de recherche post-doctoral

INSCRIPTION DU PROJET DANS UNE PROGRAMMATION SCIENTIFIQUE FINANCÉE

Le projet s'inscrit dans les axes méthodologiques "Données et méthodes numériques" et "Pratiques patrimoniales et histoires des collections" et dans le champ thématique "Objets : production, circulation, usages" du DIM Patrimoines matériels – innovation, expérimentation, résilience (PAMIR).

CHRONOLOGIE

Natoufien (https://pactols.frantiq.fr/opentheso/?idc=12834&idt=TH_1)

LIEUX

Mallaha (Israël) (<https://pactols.frantiq.fr/opentheso/?idc=31085&idt=th17>)

SUJET OU THÈME

sédentarisation

maison

habitat

2.2.IDENTIFICATION DES OPÉRATIONS ARCHÉOLOGIQUES ASSOCIÉES

IDENTIFIANT DE L'OPÉRATION OU DES OPÉRATION(S)

Opération d'archéologie programmée (Etranger)

INTITULÉ DE L'OPÉRATION OU DES OPÉRATIONS

Mission archéologique d'Eynan-Mallaha

NATURE DE(S) L'INTERVENTION(S)

Fouilles

RESPONSABLE(S) D'OPÉRATION

Perrot, Jean / CNRS (1955-1956 puis 1959-1961)

Perrot, Jean / CNRS et Lechevallier, Monique / CNRS (1971-1972)

Perrot, Jean / CNRS et Valla, François / CNRS (1973-1976)

COORDONNÉES GÉOGRAPHIQUES DE L'OPÉRATION

33°04'51.0"N 35°34'25.9"E

3. GESTION DES DONNÉES

3.1.RESPONSABILITÉS

A. ARCHIVES PAPIER

Responsable(s) de la gestion des données au cours du projet de recherche

- Responsabilité réglementaire des archives publiques : Montagne-Bôrras, Aurélie (aurelie.montagne-borras@cns.fr) / UAR 3225 Service des Archives
- Responsabilité de l'exploitation dans le cadre du projet : Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>

Autres intervenants dans la gestion des données

Assistance pour la numérisation : Grimonpont, Clara ; stagiaire en double licence Archéologie et Anthropologie - Préhistoire à l'université Paris - Nanterre

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Responsable(s) de la gestion des données au cours du projet de recherche

- Responsable de la numérisation et de la gestion des données numérisées : Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>
- Responsable de l'archivage des données à l'issue du projet, accompagnement de la numérisation : Montagne-Bôrras, Aurélie (aurelie.montagne-borras@cns.fr) / UAR 3225 Service des Archives

Autres intervenants dans la gestion des données

Assistance pour la numérisation : Grimonpont, Clara ; stagiaire en double licence Archéologie et Anthropologie - Préhistoire à l'université Paris - Nanterre

C. DOCUMENTS NUMÉRIQUES

Responsable(s) de la gestion des données au cours du projet de recherche

Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>

Autres intervenants dans la gestion des données

Participation à la transcription d'inventaires manuscrits sous forme de tableau de données :

- Congnard, Tamaé ; stagiaire de l'école du Louvre
- Morin, Émilie ; stagiaire en master Archéologie à l'université Paris - Nanterre
- Biennu, Gabriel ; stagiaire en double licence Archéologie et Anthropologie - Préhistoire à l'université Paris - Nanterre

D. BASE DE DONNÉES SPATIALE

Responsable(s) de la gestion des données au cours du projet de recherche

Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>

E. FICHIERS DE CODE

Responsable(s) de la gestion des données au cours du projet de recherche

Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>

3.2.RESSOURCES

A. ARCHIVES PAPIER

Ressources (budget et temps) allouées à la gestion des données

Accompagnement par le service des Archives de la MSH Mondes : classement du fonds François Valla, organisation de la numérisation, mise à disposition des documents. Budget temps : 1,5 mois sur 2023-2024 par IEHC

Le fonds d'archives Jean Perrot a déjà fait l'objet d'un traitement archivistique, en dehors du cadre du projet DOM-NAT.

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Ressources (budget et temps) allouées à la gestion des données

Le temps nécessaire à la numérisation a été prévu dans le temps du projet post-doctoral. Budget temps : 2 semaines par la post-doctorante.

Accompagnement du service des Archives. Budget temps : 1,5 mois sur 2023-2024 par IEHC.

Les données suivent les principes FAIR :

- facilement trouvables, car porteuses de métadonnées riches et normées, issues à la fois de leur traitement archivistique et de leur exploitation dans le cadre du projet de recherche archéologique DOM-NAT, et accompagnées d'un tableau d'index ;
- accessibles à l'ensemble de l'équipe scientifique du site d'Eynan-Mallaha via la plateforme ShareDocs, en accès restreint sur une forge Gitlab, et versées *in fine* aux archives nationales ;
- interopérables, car enregistrées au format standard TIFF ;
- réutilisables grâce aux métadonnées et à la documentation associée aux fichiers.

C. DOCUMENTS NUMÉRIQUES

Ressources (budget et temps) allouées à la gestion des données

Le temps nécessaire à la transformation des archives numérisées en documents numériques a été prévu dans le temps du projet post-doctoral. Budget temps : 2 mois par la post-doctorante.

En fonction des besoins, des stagiaires de niveau licence ou master pourront être amenés à participer à la transcription et/ou à la vectorisation.

Les données suivent les principes FAIR :

- facilement trouvables, car porteuses de métadonnées riches et normées, issues à la fois de leur traitement archivistique et de leur exploitation dans le cadre du projet de recherche archéologique DOM-NAT ;
- accessibles à l'ensemble de l'équipe scientifique du site d'Eynan-Mallaha via la plateforme ShareDocs, en accès restreint sur une forge Gitlab ;
- interopérables, car enregistrées dans des formats de fichiers standard (CSV, SVG, GeoTIFF, GeoPackage) ;
- réutilisables grâce aux métadonnées et à la documentation associée aux fichiers.

D. BASE DE DONNÉES SPATIALE

Ressources (budget et temps) allouées à la gestion des données

Le temps de conception et d'alimentation de la base de données a été prévu dans le temps du projet post-doctoral. Budget temps : 1 mois par la post-doctorante.

Les données suivent les principes FAIR :

- facilement trouvables, car associées à des métadonnées riches et normées ;
- accessibles à l'ensemble de l'équipe scientifique du site d'Eynan-Mallaha via la plateforme ShareDocs, en accès restreint sur une forge Gitlab, ou, en fonction des choix qui seront faits, sur le serveur Heurist de la TGIR Huma-Num ;
- interopérables, car enregistrées dans une base de données SQL en utilisant un système de gestion ouvert largement utilisé (PostgreSQL) ;
- réutilisables grâce aux métadonnées et à la documentation associée aux fichiers.

E. FICHIERS DE CODE

Ressources (budget et temps) allouées à la gestion des données

Aucune ressource spécifique n'est nécessaire pour la gestion de ce jeu de données ; la mise en œuvre de bonnes pratiques lors de la rédaction et de la modification du code source est jugée suffisante.

Les données suivront les principes FAIR :

- facilement trouvables, car associées à des métadonnées ;
- accessibles en accès restreint sur une forge Gitlab hébergée par la TGIR Huma-Num ;
- interopérables, car utilisant des langages de programmation largement utilisés par la communauté des chercheurs (R ; Python) ;
- réutilisables, car associées à une documentation interne et externe aux fichiers eux-mêmes.

3.3.CONTRÔLE QUALITÉ

A. ARCHIVES PAPIER

Quel est le dispositif de contrôle qualité des données et des fichiers mis en place sur ce projet ?

Exploitation d'un fonds d'archives déjà classé et d'un autre en cours de classement par le service des Archives de la MSH Mondes

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Quel est le dispositif de contrôle qualité des données et des fichiers mis en place sur ce projet ?

Les fichiers image ont été contrôlés systématiquement et manuellement, avec les critères suivants :

- cadrage ;
- définition ;

- colorimétrie ;
- métadonnées techniques ;
- concordance avec l'instrument de recherche archivistique.

Un tableau de récolement de la numérisation a été réalisé, auquel ont été ajoutées des métadonnées descriptives utiles à l'exploitation scientifique.

C. DOCUMENTS NUMÉRIQUES

Quel est le dispositif de contrôle qualité des données et des fichiers mis en place sur ce projet ?

L'adéquation entre relevés numérisés et relevés vectoriels a été contrôlée systématiquement, selon les critères suivants :

- nombre d'entités dessinées ;
- adéquation entre forme des entités vectorielles et dessin originel ;
- fermeture des formes.

Le géoréférencement des images et des tracés vectoriels a été contrôlé dans le logiciel QGIS, dans lequel a également été importé le système de carroyage utilisé pour les relevés de terrain depuis 1955.

La transcription du texte manuscrit, lorsqu'elle a été effectuée par un stagiaire, a été systématiquement contrôlée manuellement.

D. BASE DE DONNÉES SPATIALE

Quel est le dispositif de contrôle qualité des données et des fichiers mis en place sur ce projet ?

Un certain nombre de contraintes et de scripts ont été mis en place lors de la conception de la base, en amont de la saisie des données, qui ont permis de vérifier la cohérence des données au fur et à mesure de leur saisie. Par exemple, cela a permis d'obtenir une erreur lors de la saisie de deux objets portant le même identifiant, ou encore lors de la saisie d'un objet enregistré en 1959 mais localisé dans une zone non explorée durant cette campagne de fouilles.

Des listes de valeurs permettent de normaliser et de hiérarchiser les données, et de détecter ou d'éviter les erreurs de saisie. Par exemple, dans le champ "Materiau" de la table "Objet" pourront être saisies uniquement les valeurs "silex", "os", "basalte", etc. Ces listes de valeur suivent autant que possible le référentiel PACTOLS, et un champ URI permet de relier chaque valeur avec la valeur correspondante PACTOLS, lorsqu'elle existe.

E. FICHIERS DE CODE

Quel est le dispositif de contrôle qualité des données et des fichiers mis en place sur ce projet ?

Chaque fichier de code ajouté au projet a été testé pour en vérifier le bon fonctionnement.

4. DESCRIPTION DES DONNÉES

4.1.CONSTITUTION DU CORPUS

A. ARCHIVES PAPIER

Origine des données

Documents produits sur le terrain lors des fouilles du site d'Eynan-Mallaha (Israël) de 1955 à 1961 puis de 1971 à 1976, par la mission d'Aïn Mallaha (Eynan) (direction : Jean Perrot puis Jean Perrot et Monique Lechevallier puis Jean Perrot et François Valla), qui sont réutilisés dans le cadre du projet DOM-NAT.

Ces documents d'archives sont actuellement conservés par le service des Archives de la MSH Mondes. Une partie des documents (fonds Jean Perrot) a été classée par le service en amont du projet, le reste (fonds François Valla) a fait l'objet d'un traitement archivistique en cours de projet.

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Origine des données

Numérisation de documents d'archives physiques (papier ; voir produit de recherche *A. Archives papier*)

C. DOCUMENTS NUMÉRIQUES

Origine des données

Transformation de données existantes : passage d'images de documents papiers (voir produit de recherche *B. Documents d'archives numérisés*) à des documents numériques, par transcription et vectorisation des images, et géoréférencement des plans. Seuls les documents qui concernent la zone du site étudiée dans le cadre du projet DOM-NAT a été traitée.

D. BASE DE DONNÉES SPATIALE

Origine des données

Transformation de données existantes : intégration de données issues de documents numériques (tableaux d'inventaire, fichiers graphiques, images géoréférencées ; voir produit de recherche *C. Documents numériques*) dans une base de données relationnelle.

E. FICHIERS DE CODE

Origine des données

Création de nouvelles données

4.2. ORGANISATION

A. ARCHIVES PAPIER

Arborescence de classement des fichiers

Les documents du fonds Jean Perrot ont été classés au service des Archives de la MSH Mondes en amont du projet et ont fait l'objet d'un inventaire contextualisé réalisé par Erwan Le Gueut dans le cadre de son mémoire de M1, et [mis en ligne](#) sur le site du service.

En cours de projet, le fonds François Valla a été classé par Julie Bessenay-Prolonge, et les documents de terrain relatifs aux fouilles d'Eynan-Mallaha, en partie réutilisés dans le cadre du projet DOM-NAT, ont été ajoutés au fonds Jean Perrot, afin que l'ensemble de ces documents de terrain soient regroupés. L'outil de recherche a en outre été complété avec ces documents issus du fonds François Valla.

L'outil de recherche est structuré comme suit, en ce qui concerne les documents réutilisés dans le cadre du projet DOM-NAT :

- JP216-JP218 - Carnets de fouilles
- JP219 - Journaux graphiques
- JP220, JP1003-JP1004 - Catalogues de fouilles
- JP552-JP553, JP1002 - Relevés de terrain
- JP660 - Dossiers d'études

Nature des jeux de données décrits

Documents d'archives papier, notamment des relevés de terrain (plans et coupes), des inventaires manuscrits et de la documentation photographique. Les documents sont issus du fonds Jean Perrot, classé et conservé par le service des Archives de la MSH Mondes.

Les documents du fonds Jean Perrot ont fait l'objet d'un inventaire contextualisé réalisé par Erwan Le Gueut dans le cadre de son mémoire de M1 sous la direction du service des Archives de la MSH Mondes, et [mis en ligne](#) sur le site du service des Archives de la MSH Mondes. Cet inventaire a été complété par Julie Bessenay-Prolonge lors de l'ajout d'une partie des documents du fonds François Valla au fonds Jean Perrot.

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Arborescence de classement des fichiers

L'arborescence des fichiers a été conçue en concertation avec Aurélie Montagne-Bôrras, du service des Archives de la MSH Mondes. Elle se structure comme suit :

- Dossier principal correspondant à l'identifiant du fonds archivistique ("JP" pour Jean Perrot ou "FVA" pour François Valla)
- Sous-dossier correspondant à l'identifiant du dossier ou du sous-dossier (par exemple : "JP552")

Nature des jeux de données décrits

Documents numérisés sous forme de fichiers TIFF, par page individuelle : plans, inventaires de terrain, coupes stratigraphiques

C. DOCUMENTS NUMÉRIQUES

Arborescence de classement des fichiers

Les fichiers correspondant à la documentation numérique ont été classés dans un dossier dédié, divisé comme suit :

- Dossier **Relevés géoréférencés**, contenant plusieurs sous-dossiers :
 - **Plans**, classés par cote archivistique du document d'archive d'origine
 - **Coupes**, classés par cote archivistique du document d'archive d'origine
 - **Journaux graphiques**, classés par cote archivistique du document d'archive d'origine
- Dossier **Relevés vectorisés**, contenant plusieurs sous-dossiers :
 - **Plans**, classés par cote archivistique du document d'archive d'origine
 - **Coupes**, classés par cote archivistique du document d'archive d'origine
- Dossier **Inventaires**, contenant plusieurs sous-dossiers correspondant aux cotes archivistes des documents d'archive d'origine

Nature des jeux de données décrits

- Relevés de terrain, comprenant des plans et des journaux graphiques (plans schématiques) et des coupes, sous forme d'images géoréférencées ;
- Relevés de terrain, comprenant des plans et des journaux graphiques (plans schématiques) et des coupes, sous forme de tracés vectoriels géoréférencés ;
- Tableaux d'inventaire des objets.

D. BASE DE DONNÉES SPATIALE

Arborescence de classement des fichiers

Base de données relationnelle PostgreSQL comportant plusieurs tables liées entre elles.

Nature des jeux de données décrits

Données qualitatives, quantitatives et cartographiques concernant les objets et structures relevées sur le terrain lors de la fouille et de l'étude des abris 62-51-131 d'Eynan-Mallaha

E. FICHIERS DE CODE

Arborescence de classement des fichiers

- Plans
- Projections
- Autres figures

- Tableaux de données
- Analyses statistiques

Nature des jeux de données décrits

Fichiers de code source en langage R ou Python, permettant de générer soit des figures (plans, projections verticales et graphes), soit des tableaux de données, ou de réaliser des analyses statistiques.

4.3.VOLUME

A. ARCHIVES PAPIER

Aucune donnée numérique

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Poids attendu

4 Go

Poids final

3,2 Go

Nombre final de fichiers

Environ 3000

C. DOCUMENTS NUMÉRIQUES

Poids attendu

8 Go

Poids final

60 Go

Nombre final de fichiers

- 600 fichiers GeoTIFF ;
- 600 fichiers GeoPackage ;
- 2 fichiers CSV.

D. BASE DE DONNÉES SPATIALE

Poids attendu

1 Go

Poids final

75 Mo

Nombre final de fichiers

2 fichiers de code "dump" contenant la structure et le contenu de la base PostgreSQL, et permettant de la recréer intégralement

E. FICHIERS DE CODE

Poids attendu

1 Mo

Poids final

1 Mo

Nombre final de fichiers

- 22 de fichiers de code
- 3 fichiers makefile permettant de générer les figures automatiquement lorsque la base de données ou le code source est modifié

4.4.DESCRPTION DES DONNÉES**A. ARCHIVES PAPIER****Producteur(s) des données**

Jean Perrot, François Valla et leur équipe dans le cadre des fouilles du site d'Eynan-Mallaha (Israël).

Les documents proviennent de deux fonds versés au service des Archives de la MSH Mondes : le fonds Jean Perrot et le fonds François Valla.

Les fonds Jean Perrot et François Valla ont été classé respectivement par Erwan Le Gueut dans le cadre d'un mémoire de M1 et par Julie Bessenay-Prolonge, sous la responsabilité d'Aurélié Montagne-Bôrras.

Nature des données

Ensemble de documents papier (plans, inventaires, coupes stratigraphiques)

Traitement des fichiers

Les documents n'ont pas fait l'objet d'une numérisation antérieure au projet. Ceux qui ont été réutilisés dans le cadre du projet ont été numérisés (voir produit de recherche *B. Documents d'archives numérisés*).

Conditions de réutilisation

La réutilisation se fait selon le code du patrimoine (livre II Archives) et le code de la propriété intellectuelle.

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS**Producteur(s) des données**

Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>

Nature des données

Fichiers TIFF issus de la numérisation de plans, inventaires et coupes stratigraphiques

La production de données a-t-elle nécessité la réutilisation de données existantes par ailleurs ?

Réutilisation de documents d'archives physiques (voir produit de recherche *A. Archives papier*)

Comment les données ont-elles été produites ?

Les documents ont fait l'objet d'une numérisation en couleur (RVB 24 bits/pixels) à une résolution de 300 dpi, à l'aide d'un photocopieur Konica Minolta "Bizhub". La numérisation a été faite page par page, en excluant les pages vierges.

Les formats dimensionnels ont été adaptés aux formats, divers et souvent en dehors des formats standards, des documents papier d'origine.

Les documents dépassant les dimensions maximales du scanner utilisé (format A3) ont été numérisés en plusieurs fois. Le géoréférencement ultérieur des relevés de terrain (voir produit de recherche *C. Documents numériques*) permettra de rassembler les portions de ces documents *a posteriori*.

Format des données

Les données ont été enregistrées au format TIFF (Tagged Image File Format) non compressé.

Ce format a été choisi pour assurer une cohérence avec les fichiers obtenus après le géoréférencement d'une partie de ce jeu de données, au format GeoTIFF (voir produit de recherche *C. Documents numériques*).

Nommage des données

Les règles de nommage des fichiers ont été fixées en concertation avec Aurélie Montagne-Bôrras, du service des Archives de la MSH Mondes.

Chaque fichier porte un nom unique ne dépassant pas 31 caractères, et constitué exclusivement de lettres (non accentuées), chiffres, tirets bas et traits d'union.

Celui-ci est composé de plusieurs éléments séparés par des tirets bas :

- un préfixe correspondant à l'identifiant du dossier archivistique (par exemple : "JP552") ou, à défaut, à l'identifiant du fonds d'archive (par exemple : "FVA") ;
- une racine permettant d'identifier le document, correspondant soit à la catégorie documentaire (par exemple : "relevés-coupes"), soit à l'année de production ;
- un numéro suivant l'ordre du classement archivistique, dans le cas où ces deux éléments ne suffisent pas à individualiser le document ;
- un suffixe correspondant au numéro de la page pour les documents multipages (par exemple : "p024") ;
- le suffixe "v" s'il s'agit du verso d'un document.

Voici quelques exemples de noms de fichiers :

- JP552_carnetA_p32.tif
- JP219_1955_p03.tif
- JP1004_p032v.tif

C. DOCUMENTS NUMÉRIQUES

Producteur(s) des données

Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>

Nature des données

Plans géoréférencés, relevés vectoriels, tableaux d'inventaire

La production de données a-t-elle nécessité la réutilisation de données existantes par ailleurs ?

Données produites à partir de documents d'archives numérisés (voir produit de recherche *B. Documents d'archives numérisés*)

Comment les données ont-elles été produites ?

Les plans et coupes numérisés ont été géoréférencés en utilisant le logiciel QGIS, en prenant pour points de contrôle les coins des mètres carrés de fouille, dont les coordonnées sont connues. Ces points de contrôle ont été sauvegardés dans un fichier séparé. Plusieurs algorithmes de transformation ont été préalablement testés sur un échantillon de données ; l'algorithme sélectionné, ainsi que les paramètres utilisés, sont explicités dans la documentation associée aux données. Les images géoréférencées ont été enregistrées au format GeoTIFF pour archivage.

Le tracé des vestiges et structures archéologiques et des couches stratigraphiques ont été vectorisés sous QGIS, à l'aide de l'extension libre Bezier Editing. Ces tracés ont été enregistrés au format GeoPackage.

Les inventaires manuscrits ont été retranscrits manuellement sous forme de tableaux de données au format tableur.

Format des données

- Tableaux d'inventaire au format CSV (Comma-Separated Values), avec un encodage UTF-8 ;
- Fichiers rasters géoréférencés au format GeoTIFF (Geographic Tagged Image File Format) ;
- Fichiers SIG au format ouvert GeoPackage.

Nommage des données

Pour assurer la cohérence des données, les fichiers de relevés géoréférencés et les fichiers de dessin vectoriel portent le même nom que les fichiers TIFF issus de la numérisation (voir produit de recherche *B. Documents d'archives numérisés*). Étant donné que l'extension des fichiers GeoTIFF est la même que celle des fichiers TIFF (.tif), le suffixe "_georef" a été ajouté aux noms des fichiers géoréférencés.

Par exemple, les fichiers issus de la numérisation d'un plan d'archive correspondant à la page 15 du carnet A de la cote JP552 seront :

- JP552_carnetA_p15.tif (fichier image TIFF issu de la numérisation) ;
- JP552_carnetA_p15_georef.tif (fichier raster GeoTIFF correspondant à l'image géoréférencée) ;
- JP552_carnetA_p15.gpkg (fichier contenant les tracés vectoriels géoréférencés des vestiges et des structures archéologiques).

Les fichiers d'inventaire sont nommés en suivant la cote archivistique du document d'origine (par exemple : JP1004.csv).

D. BASE DE DONNÉES SPATIALE

Producteur(s) des données

Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>

Nature des données

Données descriptives et cartographiques

La production de données a-t-elle nécessité la réutilisation de données existantes par ailleurs ?

Base de données alimentée à partir de documents numériques (voir produit de recherche *C. Documents numériques*) et de documents d'archive numérisés (voir produit de recherche *B. Documents d'archives numérisés*).

Comment les données ont-elles été produites ?

La conception de la base de données a été menée en concertation avec l'ensemble de l'équipe scientifique du site d'Eynan-Mallaha (Direction F. Bocquentin et L. Weissbrod). La structure a été documentée par différents moyens explicités ci-dessous.

Les inventaires au format CSV (voir produit de recherche *C. Documents numériques*) ont fait l'objet d'imports dans la base.

Les données SIG vectorielles au format GeoPackage (voir produit de recherche *C. Documents numériques*) ont également fait l'objet d'imports, et chaque tracé a été relié à une entité existante (objet, structure, catalogue, etc.).

Cette méthode a permis d'obtenir des tables de données reliées entre elle, contenant chacune des entités (objets, structures, sépultures, etc.) elles-mêmes associées à des géométries géoréférencées et aux documents d'archives de référence (voir produit de recherche *A. Archives papier*).

Format des données

Base de données relationnelle PostgreSQL contenant plusieurs tables reliées. PostgreSQL étant un SGBD client-serveur, la base est régulièrement exportée sous forme de deux fichiers de code "dump", contenant respectivement la structure de la base et son contenu, et qui permettent de recréer la base.

Nommage des données

L'équipe de fouille et d'étude du site d'Eynan-Mallaha étant une équipe internationale, l'ensemble des éléments de la base de données a été nommé en utilisant des termes en langue anglaise.

Les tables de données, les colonnes, ainsi que les fonctions, vues, et autres éléments de la base ont été nommés en utilisant exclusivement des chiffres, des lettres minuscules et non accentuées et des tirets bas.

Les tables principales ont été nommées par des noms pluriels (par exemple : artifacts, loci, catalogs, etc.). Les tables de relations portent les noms au singulier des deux entités correspondant aux tables qu'elles relient, éventuellement suivis d'un nom commun décrivant la nature de leur relation (par exemple : "artifact_attribute_associations").

Les colonnes correspondant aux clés primaires sont nommées "id", et celles correspondant aux clés étrangères sont nommées à partir du nom de l'entité correspondant à la table liée, suivi du suffixe "_id" (par exemple : "artifact_id").

Certaines tables correspondent à des listes de valeurs (hiérarchisées ou non) ; chacune d'entre elles comprend une colonne "pactols_uri" permettant de spécifier l'URI PACTOLS correspondant au terme, lorsqu'il existe dans ce référentiel. La table "materials", par exemple, contient la liste hiérarchique des valeurs de matériau. Toutes les valeurs possèdent une URI PACTOLS, et la hiérarchie des valeurs entre elles respecte autant que possible celle de ce référentiel (par exemple : mineral > rock > flint).

E. FICHIERS DE CODE

Producteur(s) des données

Caron-Laviolette, Elisa (elisa.laviolette@gmail.com) / UMR 8068 TEMPS <https://orcid.org/0000-0002-3034-1901>

Nature des données

Code source en langages R et Python

La production de données a-t-elle nécessité la réutilisation de données existantes par ailleurs ?

Différents packages R et bibliothèques Python, libres de droits, ont été utilisés en fonction des besoins. Ils sont listés dans la documentation, ainsi que la bibliographie qui leur est associée.

Comment les données ont-elles été produites ?

Code de programmation rédigé dans le cadre du projet DOM-NAT, au fur et à mesure de l'analyse. Les programmes ont été créés afin de rendre reproductible la manipulation des données présentes dans la base de données (voir produit de recherche *D. Base de données spatiale*), leur visualisation (projection des vestiges et des structures en plan, en projection verticale et en 3D ; graphiques quantitatifs ; etc.) et les analyses statistiques qui en ont été faites.

Format des données

- Fichiers de texte brut avec extensions .R pour le langage de programmation R ou .py pour le langage Python.
- Fichiers de scripts de type makefile (extension .make)

Les figures générées par ces programmes ont été enregistrées sous forme d'images vectorielles, au format PDF.

Nommage des données

Chaque fichier porte un nom unique ne dépassant pas 31 caractères, et constitué exclusivement de lettres (non accentuées), chiffres, tirets bas et traits d'union.

Les fichiers PDF générés automatiquement portent le même nom que le fichier de code source associé. Par exemple, le fichier "vestiges_coucheIV_materiau.R" génère un plan nommé "vestiges_coucheIV_materiau.pdf".

4.5.MÉTADONNÉES ASSOCIÉES

A. ARCHIVES PAPIER

Métadonnées d'origine

Les archives du fonds Jean Perrot ont fait l'objet d'un instrument de recherche rédigé selon la norme ISAD(G) / EAD et mis en ligne sur le site du service des Archives de la MSH Mondes.

A minima, les archivistes du service remplissent les 6 champs obligatoires :

1. La référence (ou cote)
2. L'intitulé (ou titre)
3. Le producteur
4. La ou les dates
5. L'importance matérielle (ou quantité)
6. Le niveau de description

À cela s'ajoutent une description plus précise, adaptée aux besoins des archéologues.

Métadonnées ajoutées

Aucune métadonnée ajoutée

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Format des métadonnées

Les fichiers image sont accompagnés d'un tableau de récolement au format CSV indexant l'ensemble des documents numérisés. Des colonnes spécifiques ont été ajoutées à ce tableau pour accueillir les métadonnées spécifiques nécessaires à l'exploitation des données dans le cadre du projet DOM-NAT (par exemple : mètre carré, niveau stratigraphique, structure, etc.).

L'ensemble des métadonnées présentes dans ce tableau est également embarqué dans les fichiers eux-même sous forme de mots-clés suivant le standard IPTC.

Quel est le circuit de production des métadonnées ?

L'organisation et le nommage des fichiers ont été prévus en amont de la numérisation, par Elisa Caron-Laviolette, en suivant les règles fixées en concertation avec Aurélie Montagne-Böras.

Au moment de la numérisation, suivant les étapes de contrôle de la qualité de la numérisation, les fichiers ont été nommés et rangés dans l'arborescence, et une partie des champs de description du fichier de récolement ont été remplis par Elisa Caron-Laviolette.

Après la numérisation et l'analyse du contenu de chaque document, de nouveaux champs descriptifs ont été remplis par Elisa Caron-Laviolette. L'ensemble des métadonnées a été transformé selon la norme IPTC et intégré aux fichiers image par Elisa Caron-Laviolette, à l'aide du logiciel xnView.

Avec quel(s) outil(s) les métadonnées sont-elles produites ?

Les métadonnées ont été ajoutées au tableau de récolement en utilisant le logiciel Microsoft Excel ; le tableau a cependant été exporté à chaque modification au format libre et ouvert CSV.

Le logiciel xnView a été utilisé pour convertir les champs de métadonnées du tableau de récolement en métadonnées embarquées au standard IPTC.

Quels sont les référentiels utilisés pour décrire les données ?

Les référentiels PACTOLS (pour la période chronologique et les sujets) et Geonames (pour le site archéologique) ont été utilisés pour les métadonnées communes à l'ensemble des documents réutilisés.

Les métadonnées spécifiques à chaque document sont :

- les campagnes de fouilles (année) ;
- des informations propres à la méthode d'enregistrement utilisée sur le terrain (mètre carré, numéro de catalogue de fouilles, numéro de structure, couche stratigraphique, etc.).

La liste des métadonnées utilisées est spécifiée dans la documentation associée aux fichiers.

Y a-t-il une documentation associée aux données ?

Sont associés aux données :

- le tableau de récolement, enrichi de champs descriptifs, au format CSV ;
- un fichier ReadMe rédigé au format Markdown (fichier de texte brut), expliquant les choix techniques et organisationnels de la numérisation, et listant les métadonnées utilisées.

Par ailleurs, un instrument de recherche archivistique est en ligne sur le site du service des Archives de la MSH Mondes pour le fonds Jean Perrot.

C. DOCUMENTS NUMÉRIQUES

Format des métadonnées

Les images géoréférencées contiennent des métadonnées embarquées : certaines géographiques (embarquées dans le format de fichier GeoTIFF), et d'autres déjà contenues dans les fichiers image à l'issue de la numérisation (sous forme de mots-clés suivant le standard IPTC ; voir produit de recherche *B. Documents d'archives numérisés*).

Les fichiers SIG au format GeoPackage contiennent les mêmes métadonnées, embarquées dans les fichiers via une table interne dédiée.

Les tableaux d'inventaires contiennent des métadonnées reprenant celles du tableau de récolement utilisé pour la numérisation (voir produit de recherche *B. Documents d'archives numérisés*), sous la forme de texte commenté précédant les données dans les fichiers CSV.

Quel est le circuit de production des métadonnées ?

Les métadonnées des images géoréférencées ont été automatiquement générées et conservées lors de la transformation des images numériques.

L'ajout des métadonnées aux fichiers GeoPackage a été fait manuellement par Elisa Caron-Lavolette, au fur et à mesure de la création des fichiers.

Les métadonnées associées aux données d'inventaire ont été saisies au fur et à mesure de la création des tableaux numériques.

Avec quel(s) outil(s) les métadonnées sont-elles produites ?

Les métadonnées géographiques sont générées lors du géoréférencement des images, via le logiciel QGIS.

L'ajout des métadonnées aux fichiers GeoPackage se fait avec le logiciel QGIS.

La saisie des métadonnées associées aux données d'inventaire a été faite dans un éditeur de texte.

Quels sont les référentiels utilisés pour décrire les données ?

Les métadonnées issues de la numérisation sont conservées (voir produit de recherche *B. Documents d'archives numérisés*). Les référentiels PACTOLS (pour la période chronologique et les sujets) et Geonames (pour le site archéologique) sont utilisés pour les métadonnées communes à l'ensemble des documents.

Y a-t-il une documentation associée aux données ?

Un fichier ReadMe rédigé au format Markdown (fichier de texte brut) est associé aux données. Y sont explicités les choix techniques opérés lors du géoréférencement (algorithme de transformation, paramètres), de la vectorisation et de la transcription.

D. BASE DE DONNÉES SPATIALE

Format des métadonnées

Un fichier de métadonnées au format YAML (texte brut) accompagne la base de données.

Quel est le circuit de production des métadonnées ?

Les métadonnées ont été produites au moment de la conception de la base.

Avec quel(s) outil(s) les métadonnées sont-elles produites ?

Les métadonnées ont été rédigées en texte brut (format YAML) dans un éditeur de texte.

Quels sont les référentiels utilisés pour décrire les données ?

Le thésaurus PACTOLS a été utilisé pour les thématiques et pour les périodes chronologiques. Le référentiel Geonames a été utilisé pour les lieux (ici le site archéologique d'Eynan-Mallaha).

Y a-t-il une documentation associée aux données ?

La base de données est accompagnée d'un fichier ReadMe rédigé au format Markdown (texte brut), d'images de diagrammes UML explicitant les tables, les colonnes et les relations de la base et de fichiers de code en langage PlantUML permettant de générer ces diagrammes UML au format image.

Dans la base de données PostgreSQL elle-même, des commentaires ont été ajoutés au code SQL des différents objets la constituant (tables, colonnes, vues, fonctions).

E. FICHIERS DE CODE

Format des métadonnées

Chaque fichier de code possède des métadonnées au format YAML (texte brut), qui forment l'en-tête du document.

Quel est le circuit de production des métadonnées ?

Les métadonnées ont été créées au fur et à mesure de la rédaction des programmes..

Avec quel(s) outil(s) les métadonnées sont-elles produites ?

Les métadonnées ont été rédigées en texte brut (format YAML) dans un éditeur de texte

Quels sont les référentiels utilisés pour décrire les données ?

Le thésaurus PACTOLS est utilisé pour les thématiques et pour les périodes chronologiques. Le référentiel Geonames est utilisé pour les lieux (ici le site archéologique d'Eynan-Mallaha).

Y a-t-il une documentation associée aux données ?

Les fichiers sont accompagnés d'un fichier ReadMe rédigé au format Markdown (texte brut), explicitant notamment l'utilisation des fichiers makefile pour générer les figures.

Par ailleurs, de nombreux commentaires ponctuent le code source lui-même.

4.6.STOCKAGE DES DONNÉES

A. ARCHIVES PAPIER

Pas de données numériques.

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Volumétrie prévisionnelle

4 Go

Infrastructure de stockage des données numériques durant le projet

Les données et la documentation associée sont partagées sur l'espace ShareDocs dédié à la fouille et l'étude du site d'Eynan-Mallaha, sur les serveurs de la TGIR Huma-Num.

En parallèle, l'ensemble est versé sur une forge regroupant l'ensemble des données liées au projet DOM-NAT, sur l'instance Gitlab proposée par la TGIR Huma-Num.

C. DOCUMENTS NUMÉRIQUES

Volumétrie prévisionnelle

8 Go

Infrastructure de stockage des données numériques durant le projet

Les données et la documentation associée sont partagées sur l'espace ShareDocs dédié à la fouille et l'étude du site d'Eynan-Mallaha, sur les serveurs de la TGIR Huma-Num.

En parallèle, l'ensemble est versé sur une forge regroupant l'ensemble des données liées au projet DOM-NAT, sur l'instance Gitlab proposée par la TGIR Huma-Num.

D. BASE DE DONNÉES SPATIALE

Volumétrie prévisionnelle

1 Go

Infrastructure de stockage des données numériques durant le projet

Les fichiers dump permettant de recréer la base, les métadonnées et la documentation associées sont partagés sur l'espace ShareDocs dédié à la fouille et l'étude du site d'Eynan-Mallaha, sur les serveurs de la TGIR Huma-Num.

En parallèle, les mêmes fichiers sont versés sur une forge regroupant l'ensemble des données liées au projet DOM-NAT, sur l'instance Gitlab proposée par la TGIR Huma-Num.

E. FICHIERS DE CODE

Volumétrie prévisionnelle

1 Go

Infrastructure de stockage des données numériques durant le projet

Les fichiers de code source, les scripts (makefiles), la documentation associée et les fichiers PDF générés par les programmes sont partagés sur l'espace ShareDocs dédié à la fouille et l'étude du site d'Eynan-Mallaha, sur les serveurs de la TGIR Huma-Num.

En parallèle, l'ensemble, à l'exception des fichiers PDF, est versé sur une forge regroupant l'ensemble des données liées au projet DOM-NAT, sur l'instance Gitlab proposée par la TGIR Huma-Num.

4.7.SÉCURITÉ DES DONNÉES

A. ARCHIVES PAPIER

Quels sont les risques ou les menaces qui pèsent sur les données ?

Les risques qui pèsent sur ces documents sont les mêmes que pour tout centre d'archives en Île-de-France.

Garantie de confidentialité des données

Les documents sont des documents d'archives publiques.

Garantie d'intégrité et de traçabilité

Le classement et la conservation des documents sont faits dans le respect des normes archivistiques.

La consultation des documents a été faite dans le respect du classement archivistique.

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Quels sont les risques ou les menaces qui pèsent sur les données ?

L'accès aux données, via ShareDocs ou via l'instance Gitlab proposée par la TGIR Huma-Num, est protégé car soumis à une inscription via HumanID.

La totalité des données étant hébergée sur les serveurs de la TGIR, celles-ci sont sauvegardées quotidiennement. Par ailleurs, ces données sont synchronisées localement sur l'ordinateur d'Elisa Caron-Laviolette, et sauvegardées régulièrement sur un disque dur physique conservé dans les locaux de la MSH Mondes.

Garantie de confidentialité des données

Aucune donnée à caractère personnel ou confidentiel

Garantie d'intégrité et de traçabilité

La numérisation a été effectuée à partir d'un ensemble d'archives décrites selon les normes ISAT(G) / EAD / ISARCPF, puis versées au service des Archives de la MSH Mondes.

L'utilisation d'une forge Gitlab permet de retracer les modifications apportées aux données, notamment au tableau de récolement.

C. DOCUMENTS NUMÉRIQUES

Quels sont les risques ou les menaces qui pèsent sur les données ?

L'accès aux données, via ShareDocs ou via l'instance Gitlab proposée par la TGIR Huma-Num, est protégé car soumis à une inscription via HumanID.

La totalité des données étant hébergée sur les serveurs de la TGIR, celles-ci sont sauvegardées quotidiennement. Par ailleurs, ces données sont synchronisées localement sur l'ordinateur d'Elisa Caron-Laviolette, et sauvegardées régulièrement sur un disque dur physique conservé dans les locaux de la MSH Mondes.

Garantie de confidentialité des données

Aucune donnée à caractère personnel ou confidentiel

Garantie d'intégrité et de traçabilité

L'utilisation d'une forge Gitlab permet de retracer les modifications apportées aux données et aux métadonnées.

D. BASE DE DONNÉES SPATIALE

Quels sont les risques ou les menaces qui pèsent sur les données ?

L'accès aux données, via ShareDocs, via le serveur Heurist de la TGIR Huma-Num ou via l'instance Gitlab proposée par la TGIR Huma-Num, est protégé car soumis à une inscription via HumanID.

La totalité des données étant hébergée sur les serveurs de la TGIR, celles-ci sont sauvegardées quotidiennement. Par ailleurs, ces données sont synchronisées localement sur l'ordinateur d'Elisa Caron-Laviolette et sauvegardées régulièrement sur un disque dur physique conservé dans les locaux de la MSH Mondes.

Garantie de confidentialité des données

Aucune donnée à caractère personnel ou confidentiel

Garantie d'intégrité et de traçabilité

L'utilisation d'une forge Gitlab permettra de retracer les modifications apportées aux données et aux métadonnées.

E. FICHIERS DE CODE

Quels sont les risques ou les menaces qui pèsent sur les données ?

L'accès aux données, via ShareDocs ou via l'instance Gitlab proposée par la TGIR Huma-Num, est protégé car soumis à une inscription via HumanID.

La totalité des données étant hébergée sur les serveurs de la TGIR, celles-ci sont sauvegardées quotidiennement. Par ailleurs, ces données sont synchronisées localement sur l'ordinateur d'Elisa Caron-Laviolette, et sauvegardées régulièrement sur un disque dur physique conservé dans les locaux de la MSH Mondes.

Garantie de confidentialité des données

Aucune donnée à caractère personnel ou confidentiel

Garantie d'intégrité et de traçabilité

L'utilisation d'une forge Gitlab permet de retracer les modifications apportées au code source, à la documentation et aux métadonnées.

4.8.ACCÈS AUX DONNÉES

A. ARCHIVES PAPIER

Lecture des données

Les documents sont mis à disposition par le service des Archives de la MSH Mondes.

Il s'agit de documents manuscrits, rédigés en langue française.

Comment la disponibilité des données est-elle garantie ?

Les documents physiques sont mis à disposition par le service des Archives de la MSH Mondes.

Les documents réutilisés dans le cadre du projet sont numérisés (voir produit de recherche *B. Documents d'archives numérisés*).

Échange et partage

Voir paragraphe précédent

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Lecture des données

Aucun logiciel spécifique n'est nécessaire à la lecture des données.

Comment la disponibilité des données est-elle garantie ?

Les données sont accessibles :

- sur l'espace ShareDocs dédié à la fouille et l'étude du site d'Eynan-Mallaha, sur les serveurs de la TGIR Huma-Num, via un accès authentifié par HumanID
- sur une forge Gitlab hébergée par la TGIR Huma-Num, via un accès authentifié par HumanID

Échange et partage

Les données sont partagées :

- avec l'ensemble de l'équipe scientifique du site d'Eynan-Mallaha, via l'espace ShareDocs dédié
- avec l'équipe projet via une forge Gitlab hébergée par la TGIR Huma-Num

C. DOCUMENTS NUMÉRIQUES

Lecture des données

Aucun logiciel propriétaire n'est nécessaire à la lecture des données. Les fichiers SIG sont notamment lisibles à l'aide du logiciel libre QGIS. Les tableaux d'inventaire et les documents graphiques vectoriels ont été enregistrés dans des formats standards (PDF, CSV) et lisibles par différents logiciels libres ou propriétaires.

Comment la disponibilité des données est-elle garantie ?

Les données sont accessibles :

- sur l'espace ShareDocs dédié à la fouille et l'étude du site d'Eynan-Mallaha, sur les serveurs de la TGIR Huma-Num, via un accès authentifié par HumanID
- sur une forge Gitlab hébergée par la TGIR Huma-Num, via un accès authentifié par HumanID

Échange et partage

Les données sont partagées :

- avec l'ensemble de l'équipe scientifique du site d'Eynan-Mallaha, via l'espace ShareDocs dédié
- avec l'équipe projet via une forge Gitlab hébergée par la TGIR Huma-Num

D. BASE DE DONNÉES SPATIALE

Lecture des données

L'accès aux données ne nécessite pas de logiciel propriétaire, et peut se faire soit en utilisant directement le langage SQL, soit en passant par un logiciel libre comme QGIS en le connectant à la base.

Comment la disponibilité des données est-elle garantie ?

Les données sont accessibles :

- sur l'espace ShareDocs dédié à la fouille et l'étude du site d'Eynan-Mallaha, sur les serveurs de la TGIR Huma-Num, via un accès authentifié par HumanID ;
- sur une forge Gitlab hébergée par la TGIR Huma-Num, via un accès authentifié par HumanID.

Échange et partage

Les données sont partagées :

- avec l'ensemble de l'équipe scientifique du site d'Eynan-Mallaha, via l'espace ShareDocs dédié
- avec l'équipe projet, via une forge Gitlab hébergée par la TGIR Huma-Num

E. FICHIERS DE CODE

Lecture des données

Aucun logiciel propriétaire n'est nécessaire à la lecture des données. À l'exception des fichiers PDF générés par les programmes informatiques, tous les fichiers peuvent être lus dans un éditeur de texte.

Comment la disponibilité des données est-elle garantie ?

Les données sont accessibles :

- sur l'espace ShareDocs dédié à la fouille et l'étude du site d'Eynan-Mallaha, sur les serveurs de la TGIR Huma-Num, via un accès authentifié par HumanID
- sur une forge Gitlab hébergée par la TGIR Huma-Num, via un accès authentifié par HumanID, à l'exception des figures elles-mêmes, qui peuvent être générées en suivant les instructions figurant dans la documentation.

Échange et partage

Les données sont partagées :

- avec l'ensemble de l'équipe scientifique du site d'Eynan-Mallaha, via l'espace ShareDocs dédié
- avec l'équipe projet via une forge Gitlab hébergée par la TGIR Huma-Num, à l'exception des figures elles-mêmes, qui peuvent être générées en suivant les instructions figurant dans la documentation.

4.9.PROTECTION DES DONNÉES SENSIBLES AU COURS DU PROJET

Y a-t-il des données nécessitant une protection juridique particulière et préciser pourquoi ?

Aucune protection juridique spécifique n'est nécessaire

Des mesures de protection ont-elles été prises pour protéger les données sensibles ?

Les documents ne contiennent pas de données sensibles

4.10.CADRE JURIDIQUE

A. ARCHIVES PAPIER

Cadre juridique d'exploitation, de réutilisation et de préservation des données

La réutilisation se fait selon le code du patrimoine (livre II Archives) et le code de la propriété intellectuelle.

B. DOCUMENTS D'ARCHIVES NUMÉRISÉS

Cadre juridique d'exploitation, de réutilisation et de préservation des données

La réutilisation se fait selon le code du patrimoine (livre II Archives) et le code de la propriété intellectuelle.

C. DOCUMENTS NUMÉRIQUES

Cadre juridique d'exploitation, de réutilisation et de préservation des données

La politique de réutilisation des données produites dans le cadre du projet DOM-NAT a fait l'objet d'une charte, suivant les recommandations du DIM PAMIR, financeur du projet.

D. BASE DE DONNÉES SPATIALE

Cadre juridique d'exploitation, de réutilisation et de préservation des données

La politique de réutilisation des données produites dans le cadre du projet DOM-NAT a fait l'objet d'une charte, suivant les recommandations du DIM PAMIR, financeur du projet.

E. FICHIERS DE CODE

Cadre juridique d'exploitation, de réutilisation et de préservation des données

La politique de réutilisation des données produites dans le cadre du projet DOM-NAT a fait l'objet d'une charte, suivant les recommandations du DIM PAMIR, financeur du projet.

5. À L'ISSUE DU PROJET

5.1.POLITIQUE D'ARCHIVAGE DES DONNÉES PRODUITES

Durée de conservation des données produites

Durée indéterminée (archives définitives)

Volumétrie finale des données issues du projet

Environ 140 Go

Modalités de conservation des données à l'issue du projet

À l'issue du projet, les fichiers sont versés au service des archives de la MSH Mondes, sous le contrôle scientifique et technique des Archives de France. À l'issue de l'exploitation scientifique, ces archives seront versées aux archives nationales.

En outre, l'ensemble des données sera déposé sur Zenodo à l'issue du projet.

Modalités d'accès aux données à l'issue du projet

Les données seront consultables au service des Archives selon la loi sur les archives publiques et le droit d'auteur. Leur mise en ligne est à l'étude.

5.1.DISSÉMINATION DES DONNÉES

Politique de dissémination des données à l'issue du projet

La mise en ligne des fichiers numérisés associés à l'instrument de recherche en XML EAD (solution full web) est à l'étude.

L'exploitation de ces jeux de données a-t-elle donné lieu à des publications ?

Bocquentin F., Caron-Laviolette E., Fourchet N., Davin, L., Withford B., Heccan L., Le Gueut E., Bessenay-Prolonge J., Montagne Bôrras A., Weissbrod, L. Hunter-Gatherer-Builders: Past and ongoing research at the Natufian hamlet of Eynan-Mallaha, Archaeological Research in Asia (article soumis)

Certaines données sont-elles sous embargo ?

Aucune donnée sous embargo

Potentiel de réutilisation des données (publics cibles, moyens nécessaires...).

Type de public : chercheurs en archéologie travaillant sur l'Épipaléolithique du Proche-Orient.

Éléments permettant la réutilisation : documentation et métadonnées associées

Les données produites dans le cadre du projet seront en premier lieu réutilisées dans le cadre de la fouille et de l'étude du site d'Eynan-Mallaha dans son ensemble (Direction F. Bocquentin et L. Weissbrod).

Conditions de réutilisation

La réutilisation des documents d'archives se fait selon la loi sur les archives publiques et le droit d'auteur.

La politique de réutilisation des données produites dans le cadre du projet DOM-NAT fait l'objet d'une charte, suivant les recommandations du DIM PAMIR, financeur du projet.