

## MASA Consortium: MASA - Modèle de PGD (français) - Contexte

### Identification du projet

#### Acronyme

*Exemple de réponse:*

VERSARC

*Recommandations:*

Information optionnelle. Si le projet possède un acronyme ou un code usuel, l'indiquer ici.

Le développé titre du projet doit être renseigné dans l'onglet *Renseignement sur le projet*.

#### Affiliation du responsable scientifique

*Recommandations:*

Indiquer l'affiliation complète du responsable scientifique. Il est fortement recommandé de lui associer également, si possible, un identifiant chercheur (**ORCID**, ISNI, IdRef, IdHal...).

Les identifiants chercheurs -- appelés aussi "identifiant auteur" -- sont des codes numériques permanents permettant de lier sans équivoque les personnes aux documents dans lesquelles elles ont une responsabilité intellectuelle (article scientifique, chapitre d'ouvrage, brevet, préface, etc.).

Le Research Organization Registry peut être utilisé pour identifier les institutions (exemples du CNRS : <https://ror.org/02feahw73> et de l'INRAP : <https://ror.org/04andmq85> )

*Exemple de réponse:*

Directrice de recherche, CNRS, UMR 1000 (irenee.david@cnrs.fr). <https://orcid.org/0000-0001-9422-1000>

#### Institution porteuse du projet, coordinateur personne morale bénéficiaire (nom, pays, Identifiant).

*Recommandations:*

Le Research Organization Registry peut être utilisé pour identifier les institutions (exemples du CNRS : <https://ror.org/02feahw73> et de l'INRAP : <https://ror.org/04andmq85> )

*Exemple de réponse:*

CNRS, France : <https://ror.org/02feahw73>

#### Partenaires du projet

*Recommandations:*

Indiquer ici les différents partenaires du projet (équipe de recherche, infrastructure, réseau, etc.).

#### Type de financement

- Autre (préciser)
- Projet de recherche doctoral
- Projet collectif de recherche
- Projet financé par la Commission européenne
- Projet financé par l'ANR
- Opération d'archéologie programmée (Etranger)
- Opération d'archéologie programmée (France)
- Opération d'archéologie préventive

#### Type de projet

- Valorisation
- Médiation
- Numérisation
- Autre
- Publication : catalogue
- Publication : édition de sources littéraires
- Publication : édition de sources épigraphiques
- Etude
- Enquête
- Inventaire archéologique ou prospection
- Fouille archéologique ou prospection

*Exemple de réponse:*

Exemple : Edition épigraphique numérique basée sur un encodage TEI/EpiDoc.

*Recommandations:*

Utiliser la zone "informations supplémentaires" si nécessaire pour qualifier plus précisément le projet en indiquant la

discipline, la thématique ou le type d'intervention.

Par exemple : archéozoologie, archéologie expérimentale, archéologie préventive, intervention sous-marine, inventaire du patrimoine, etc.

## Descriptif du projet et mots clés

Inscription du projet dans une programmation scientifique financée

### *Recommandations:*

Si le projet s'inscrit dans une programmation scientifique financée, préciser les axes scientifiques associés à la programmation. Par exemple :

- [Programmation nationale du conseil national de la recherche archéologique](#)
- Axes scientifique d'un Labex
- Programme de financement d'un projet ANR, H2020
- Axe ou programme scientifique d'une structure de recherche liée au porteur ou à l'équipe projet...

Chronologie

### *Exemple de réponse:*

Paléolithique supérieur (<https://ark.frantiq.fr/ark:/26678/pct9U8BH9pVRu>)

Paléolithique supérieur (<https://pactols.frantiq.fr/opentheso/3158>)

### *Recommandations:*

Indiquez ici le ou les mot(s) clé(s) décrivant la période chronologique associée au projet. Si le référentiel est en ligne et dispose d'identifiants pérennes (Uniform Resource Identifier ou URI) ou "permalien", lier le descripteur à cet identifiant.

Utiliser si possible un vocabulaire contrôlé. Le domaine "Chronologie" du [thésaurus Pactols](#) est recommandé.

Lieux

### *Exemple de réponse:*

Chypre (<https://pactols.frantiq.fr/opentheso/5365>)

Salamine (Chypre) (<https://pactols.frantiq.fr/opentheso/10110>)

### *Recommandations:*

Utiliser si possible des vocabulaires contrôlés. Privilégier les vocabulaires de référence pour le thème de recherche ou la communauté dans laquelle le projet s'inscrit. Si le référentiel est en ligne et dispose d'identifiants pérennes (Uniform Resource Identifier ou URI) ou "permalien", lier le descripteur à cet identifiant.

Exemple de référentiels géographiques utiles pour les projets liés à l'archéologie ou l'étude des mondes anciens :

- Domaine "Lieux" du thésaurus [Pactols](#) ;
- Index en ligne [GeoNames](#) ;
- Index géo-historique des lieux anciens [Pleiades](#) ;
- etc.

S'il s'agit d'une opération archéologique, renseigner les coordonnées selon la **projection Lambert93** (projection officielle pour les cartes de France métropolitaine, liée au système géodésique RGF93).v Indiquer le **niveau de précision utilisé**, et le justifier le cas échéant (par exemple utilisation du baricentre de la commune pour protéger le site).

Sujet ou thème

### *Exemple de réponse:*

cabane  
fossé parcellaire  
trou de poteau  
structure de combustion  
torchis  
palissade  
quartier spécial  
groupes sociaux  
céramique moderne  
ossements animaux  
objet métallique  
verre  
terre cuite architecturale

### *Recommandations:*

Utiliser si possible des vocabulaires contrôlés. Privilégier les vocabulaires de référence pour le thème de recherche ou la communauté dans laquelle le projet s'inscrit. Si le référentiel est en ligne et dispose d'identifiants pérennes (Uniform Resource Identifier ou URI) ou "permalien", lier le descripteur à cet identifiant.

L'utilisation du thésaurus [Pactols](#) est recommandé.

## Identification des opérations archéologiques associées (optionnel)

Identifiant de l'opération ou des opérations(s)

*Recommandations:*

Cette rubrique n'est à renseigner que si ce numéro existe. Pour les fouilles en France par exemple, indiquer ici l'identifiant de l'opération dans le système national de référence (code "OA").

Intitulé de l'opération ou des opérations

*Recommandations:*

Intitulé(s) de l'opération ou des opérations archéologiques tel qu'il(s) figure(nt) dans le système national de référence pour les opérations d'archéologie en France ou tel qu'il(s) a/ont été attribué(s) par le ministère des Affaires étrangères pour les missions françaises à l'étranger.

Statut administratif de l'opération

*Exemple de réponse:*

Fouille préventive, fouille programmée, fouille Ministère des affaires étrangères, etc.

Nature de(s) l'intervention(s)

*Exemple de réponse:*

Diagnostic, fouilles, prospection, surveillance de travaux, fouilles subaquatiques, etc.

Responsable(s) d'opération

*Recommandations:*

Plusieurs systèmes d'identifiants de personnes peuvent être utilisés. Si l'identifiant de personne n'existe pas dans ORCID, d'autres registres peuvent être utilisés (ex: ISNI, IDREF, VIAF, wikidata...)

*Exemple de réponse:*

Perrot, Jean / CNRS.

Coordonnées géographiques de l'opération

*Recommandations:*

Indiquer ici les coordonnées en Lambert 93.

Indiquer le niveau de précision utilisé, et le justifier le cas échéant (par exemple utilisation du baricentre de la commune pour protéger le site).

## Historique du document

Date de création

*Exemple de réponse:*

17/09/2019

*Recommandations:*

Indiquer la date sous la forme JJ/MM/AAAA

Version du document

*Exemple de réponse:*

initiale / intermédiaire / finale

*Recommandations:*

Indiquer le numéro de version du plan de gestion de données. Cela peut prendre la forme d'un libellé d'étape ("initiale" vs "intermédiaire" vs "finale") ou d'une numérotation (v.2.1.1) ou d'une version datée (version du 12/12/1019).

Nombre de versions prévues

*Exemple de réponse:*

Trois versions seront réalisées : une dans les 6 mois après le démarrage du projet, une à mi-parcours et la dernière à la fin du projet.

*Recommandations:*

Vérifiez si le financeur impose un nombre de versions ou si les modalités du versionnage sont libres. Expliquez la logique retenue.

Rédacteur(s) du plan de gestion de données

*Exemple de réponse:*

Hugo, Saturnin (saturnin.hugo@cncs.fr) / <http://orcid.org/0000-0003-4317-1800>

*Recommandations:*

Le rédacteur du plan de gestion de données n'est pas forcément le responsable scientifique du projet. Le contenu de ce champ peut être différent du contenu du champ "Créateur du PGD" de l'onglet "Renseignement sur le projet" qui désigne la personne en charge de la validation du plan de gestion de données. Dans le modèle ANR, il s'agit du coordinateur du plan de gestion de données.

Intégrer, quand il existe, un identifiant de personne (ORCID, ISNI, IdRef, IdHal).

## MASA Consortium: MASA - Modèle de PGD (français) - Gestion des données

### Responsabilités

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

*Exemple de réponse:*

- Responsable des données CNRS : Hugo, Saturnin (saturnin.hugo@cns.fr) /USR 1000 du CNRS <http://orcid.org/0000-0003-4317-1800>
- Responsables des données l'INRAP : Edmond, Jeanne (jeanne.edmond@inrap.fr)/ INRAP. <http://orcid.org/0000-0003-4317-1700>

*Recommandations:*

Le porteur du projet désignera un gestionnaire de données qualifié pour gérer les données collectées, traitées et analysées.

Le responsable de la gestion de données doit être impliqué dans le pilotage du projet (data stewardship), de sa création à son aboutissement. Il est responsable de la collecte, l'organisation et la conservation des données. Il peut être différent du responsable scientifique.

Dans le cas d'un projet associant des partenaires sur d'autres sites ou dans d'autres institutions, les responsabilités entre les partenaires doivent être précisées ici.

Autres intervenants dans la gestion des données

*Recommandations:*

Décrire les rôles et les responsabilités de l'ensemble des personnes intervenant dans la gestion des données : saisie des données, modélisation et production des métadonnées, gestion de la qualité des données et des fichiers produits, stockage et sauvegarde, archivage et partage des données. Nommer la(es) personne(s) responsable(s) impliquée(s) dans la mesure du possible.

Pour les projets menés en collaboration, expliquer comment s'effectue la coordination des responsabilités de gestion de données entre partenaires.

### Ressources

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

*Recommandations:*

Expliquer comment les ressources nécessaires (par exemple, le temps) à la préparation des données pour le partage, la préservation (curation) des données ont été chiffrées.

Examiner et justifier soigneusement toutes les ressources nécessaires pour s'assurer que les données seront *FAIR* (Facile à trouver, Accessibles, Interopérables, Réutilisables).

#### Définition des principes FAIR

Définition de la curation des données (site "Libre accès à l'information scientifique et technique" de l'Inist) :

On désigne par curation l'ensemble des activités et opérations nécessaires à une gestion active des données de recherche numériques, tout au long de leur cycle de vie. L'objectif est de les rendre accessibles, partageables et réutilisables de façon pérenne. Trois intervenants peuvent être identifiés dans le cycle de vie de données : les créateurs, le plus souvent les chercheurs, les « curateurs » et les utilisateurs.

<http://openaccess.inist.fr/?+-Curation-+>

Source : [INRA](#)

### Contrôle qualité

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

*Recommandations:*

Indiquez ici si un dispositif de contrôle de la qualité des données et des fichiers a été mis en place à l'échelle du projet ou éventuellement à l'échelle d'un ou des jeux de données.

Vous pouvez vous reporter au guide "[Traçabilité des activités de recherche et gestion des connaissances : guide pratique de mise en place](#)" élaboré par la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS (2018), disponible sur Doranum.

Pour vérifier que les formats de fichiers sont éligibles à l'archivage par le Centre national informatique de l'enseignement supérieur (CINES), il est possible d'utiliser l'outil de validation des formats de fichiers Facile : <https://facile.cines.fr/>

# MASA Consortium: MASA - Modèle de PGD (français) - Généralités sur les données

## Constitution du corpus

Origine des données

*Exemple de réponse:*

Collecte de nouvelles données, conversion, transformation de données existantes, partage ou échange de données, achat de données.

## Organisation

Arborescence de classement des fichiers

*Exemple de réponse:*

Exemple d'arborescence pour un projet lié directement à une ou plusieurs opérations archéologiques (INRAP) :

- Documentation administrative (correspondance, contrat...)
- Données (données brutes produites lors de la phase terrain)
- Données spatiales (toutes les données géoréférencées)
- Etude (données et documentation issues du traitement des données brutes stockées dans Donnée ; études spécialisées)
- Illustration (fichiers intermédiaires nécessaires à la réalisation des figures produites dans le cadre du projet pour des actions de valorisation, des posters, des publications etc.)

*Recommandations:*

Indiquer ici l'arborescence de classement des fichiers utilisés et produits au cours du projet. Celle-ci doit être mise en place au démarrage du projet et utilisée tout au long de celui-ci. Cette arborescence peut aider à identifier les jeux de données.

Vous pouvez vous reporter au guide "[Traçabilité des activités de recherche et gestion des connaissances : guide pratique de mise en place](#)" élaboré par la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS (2018), disponible sur Doranum, pour des conseils sur l'élaboration d'un plan de classement.

Nature des jeux de données décrits

*Recommandations:*

Lister ici les jeux de données et décrire brièvement leur contenu.

Si votre jeu de données correspond à un logiciel de la recherche, nous vous recommandons de consulter le modèle de plan de gestion PRESOFT. Ce modèle est disponible sur OPIDoR et documenté sur <http://www.france-grilles.fr/presoft/>

## Volume

Poids attendu

*Recommandations:*

Préciser l'unité utilisée (Go, Mo, To...etc.)

Vous trouverez des exemples dans le guide "[Traçabilité des activités de recherche et gestion des connaissances : guide pratique de mise en place](#)" élaboré par la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS (2018), disponible sur Doranum.

Poids final

*Recommandations:*

Préciser l'unité utilisée (Go, Mo, To...etc.)

Nombre final de fichiers

## Cadre juridique

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

*Recommandations:*

Donner ici toutes les informations juridiques pertinentes permettant de comprendre les conditions d'exploitation, de réutilisation et de préservation des données.

- À qui appartiennent les données produites et/ou réutilisées ?
- Quels droits régissent les bases de données et leurs contenus ?
- Comment se répartit la propriété intellectuelle sur les données ? etc.

En cas de partenariat, la propriété des données doit être fixée par l'accord définissant le rôle de chacun des partenaires du projet. La négociation des termes de l'accord en amont du projet et du PGD est essentielle car elle conditionne en partie les modalités de partage et de dissémination ultérieures des données.

cf. <https://www.ouvrirlascience.fr/ouverture-des-donnees-de-recherche-guide-danalyse-du-cadre-juridique-en-france-v2/>  
Dans le cas de développement logiciels, pensez à vous référer au droit spécifique aux logiciels.

## MASA Consortium: MASA - Modèle de PGD (français) - Jeux de données

### Jeux de données

Nom du jeux de données

*Recommandations:*

- Appliquer les règles de nommage définies dans le plan de nommage du projet, s'il existe
- En l'absence de consignes particulières, le lot peut être nommé de la manière suivante : *acronyme projet\_nature du lot\_nombre de fichiers*

Identifiant du jeu de données

*Recommandations:*

Indiquer ici si un identifiant unique type **DOI, Ark**, etc..., est associé au jeu de données

Statut des données

- Produite
- Réutilisée

*Recommandations:*

S'agit-il de données réutilisées ou de données produites au cours du projet ?

On entend ici par "données réutilisées", les données numériques *produites à d'autres fins que celles du projet en courset réutilisées telles quelles*. Il est nécessaire de s'être assuré avant des conditions de leur réutilisation.

**Important !**

Les questions marquées d'un astérisque sont spécifiques aux **jeux de données réutilisées**.

### Description des données

Producteur(s) des données

*Exemple de réponse:*

Dupont, Michel/ Inrap, topographe

*Recommandations:*

Si les données composant le jeu de données sont produites par un collectif, trouver une signature générique type "équipe projet XXX". Si elles ne sont produites que par une personne, alors c'est cette personne qui est mentionnée comme producteur du jeu de données.

Préciser : Nom, Prénom/ Affiliation, fonction au sein du projet. Identifiant

Nature des données

*Exemple de réponse:*

Mesures (expérimentales, d'observation), cartes et plans, données quantitatives, modèle 3D, données traitées, photographie, inventaires, minutes, données spatiales, étude spécialisée, échantillons, collections physiques, logiciels, supports pédagogiques, bibliographie, SIG, données administratives, modèle logique de BDD...

*Recommandations:*

Décrire brièvement les données qui constituent le jeu de données.

Si vous utilisez un référentiel pour qualifier la nature des données, précisez-le.

Les formats de fichiers, sont décrits plus loin (question n°5).

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

*Exemple de réponse:*

- producteur des données
- propriétaire des données
- lieu de stockage des données
- cadre juridique de la réutilisation

*Recommandations:*

Si la production du jeu de données a nécessité la réutilisation de données produites à d'autres fins que celles du projet pour lequel le PGD est rédigé, indiquer lesquelles et les décrire sommairement (producteur, propriétaire, lieu de stockage, conditions de réutilisation). L'onglet "*description jeu de données réutilisées*" est à utiliser pour décrire les jeux de données réutilisés tels quels dans le projet faisant l'objet du PGD.

Expliquer, le cas échéant, les raisons pour lesquelles l'utilisation de sources de données existantes a été envisagée mais écartée.

Comment les données ont-elles été produites ?



### Exemple de réponse:

Levers topographiques- Canne GPS LEICA GS08 plus

### Recommandations:

En fonction de la nature du lot, préciser la ou les méthodes d'acquisition des données : type de l'appareil photographique, relevé, scan, vectorisation etc.

Le cas échéant préciser quels sont les réglages, unités de mesure utilisés ainsi que le degré de précision ou tout autre indicateur ou information permettant de mesurer l'impact de l'outil sur la qualité des données produites.

Pour les jeux de données issus de bases de données, documenter le modèle de données et préciser la version du logiciel utilisé.

Préciser, lorsque celui-ci a une influence sur la qualité des données, leur contexte de production (conditions climatiques par ex.)

### Format des données

#### Recommandations:

- Indiquer le format natif des fichiers constituant le jeu de données décrit (jpeg, ai, cord, xls, ai; dwg, etc.).
- Privilégier les formats :  
intéropérables standard  
ouverts et non propriétaires  
validés par le CINES (testés avec l'outil en ligne Facile )
- Consulter les recommandations du Consortium 3D figurant dans le Livre blanc du consortium 3D SHS (2017).
- Vous pouvez également vous référer au registre de l'Internet Assigned Numbers Authority (IANA) ou au guide des bonnes pratiques d'Archaeology Data Service (ADS) pour les formats de fichiers non recensés par le CINES.

Justifier les choix techniques (usage répandu dans la communauté, fruit d'une expertise, recommandé pour le stockage dans un entrepôt de données...etc.)

### Exemple de réponse:

- 250 fichiers .jpeg
- 12 fichiers .mp4
- 1 fichier .csv
- 5 fichiers .PLY (ASCII 1.0)
- 2 fichiers PDF (1.4)

Dans cet exemple, le fichier csv et l'un des fichiers PDF contiennent les mêmes informations : métadonnées descriptives relatives aux images et vidéos constituant le jeu de données. Le fichier csv est destiné à faciliter la réutilisation des données, le fichier PDF 1.4 (PDF/A) vise à faciliter l'archivage à long terme des données et leur accessibilité.

### Nommage des données

#### Recommandations:

Indiquer ici les règles de nommage adoptées pour les données produites au cours du projet.

Vous pouvez vous référer aux consignes émises par le service ADS pour l'archivage pérenne des données archéologiques. Celles-ci doivent respecter les principes garantissant l'accessibilité des fichiers sur le long terme pour ce faire :

le nom d'un fichier doit

- être succinct : il ne doit pas dépasser 31 caractères, sans compter l'extension (/!\ Un maximum de 255 caractères est autorisé pour le nom complet : chemin d'accès au fichier et nom du fichier),

- être précis : selon les cas il peut ou doit contenir le sujet du document, son type, sa date de création ou encore la version,

- être unique pour éviter toute perte d'information et rendre cohérent le fonds de données et documents numériques créés lors du projet. Si le format du fichier doit être modifié pour son archivage, le nom n'est pas modifié.

Par ailleurs, le nom du fichier doit respecter les caractères et pratiques suivants :

- les 26 lettres de l'alphabet latin,

- Les 10 chiffres arabes,

- l'underscore « \_ » (tiret bas de la touche 8), que l'on préférera au trait d'union (tiret de la touche 6) et qui remplacera les espaces,

- les majuscules au début de certains termes, permettant d'éviter « » en collant au terme précédent.

Et éviter les caractères et pratiques ci-dessous :

- les signes diacritiques : à, â, ä, é, è, ê, ë, ì, î, ô, ö, ù, û, ü, ÿ, ç

- les caractères spéciaux comme , ; : ! ? % & ( ) # / \*

- les mots vides : le, la, un, une, des, et

- les dénominations vagues de types « divers », « autres », « à classer »

- le nom de l'agent créateur ou gestionnaire du fichier

le nom du fichier peut être signifiant ie. décrire le contenu du fichier, ou pas, comme lorsque l'on utilise, pour une image, son numéro d'identification dans la base de données qui la référence.

le nom du fichier doit tenir des contraintes propres à l'institution de production (référentiels institutionnels, contraintes réglementaires etc)

### Traitement des fichiers\*

### Recommandations:

- Indiquez ici si vous avez renommé les fichiers pour les utiliser dans le cadre du projet. Si c'est le cas indiquez les règles de nommage adoptées.

Celles-ci doivent respecter les principes garantissant l'accessibilité des fichiers sur le long terme, pour ce faire le nom d'un fichier doit :

- être succinct : il ne doit pas dépasser 31 caractères, sans compter l'extension (Un maximum de 255 caractères est autorisé pour le nom complet : chemin d'accès au fichier et nom du fichier),
  - être précis : selon les cas il peut ou doit contenir le sujet du document, son type, sa date de création ou encore la version,
  - être unique pour éviter toute perte d'information et rendre cohérent le fonds de données et documents numériques créés lors du projet. Si le format du fichier doit être modifié pour son archivage, le nom n'est pas modifié.
- Par ailleurs, le nom du fichier doit respecter les caractères et pratiques suivants :
- les 26 lettres de l'alphabet latin,
  - les 10 chiffres arabes,
  - l'underscore « \_ » (tiret bas de la touche 8), que l'on préférera au trait d'union (tiret de la touche 6) et qui remplacera les espaces,
  - les majuscules au début de certains termes, permettant d'éviter « » en collant au terme précédant.

Et éviter les caractères et pratiques ci-dessous :

- les signes diacritiques : à, â, ä, é, è, ê, ë, î, î, ô, ö, ù, û, ü, ÿ, ç
- les caractères spéciaux comme , ; : ! ? % & ( ) # / \*
- les mots vides : le, la, un, une, des, et
- les dénominations vagues de types « divers », « autres », « à classer »
- le nom de l'agent créateur ou gestionnaire du fichier

Vous pouvez vous référer aux consignes émises par le service ADS pour l'archivage pérenne des données archéologiques. Le nom du fichier peut être signifiant, ie. décrire le contenu du fichier, ou pas, comme lorsque l'on utilise, pour une image, son numéro d'identification dans la base de données qui la référence.

Le nom du fichier doit tenir compte des contraintes propres à l'institution de production (référentiels institutionnels, contraintes réglementaires etc)

- Mentionner ici le format des fichiers constituant le jeu de données décrit (jpeg, ai, cord, xls, ai; dwg, etc.). Indiquer si vous avez converti les fichiers sources.
- Indiquez ici le ou les système d'indexations (vocabulaires contrôlés, ontologie, etc) Utilisez les référentiels du domaine et en prévoir, le cas échéant, l'enrichissement à l'occasion du projet. Indiquez si vous avez généré une nouvelle indexation pour l'utilisation des données dans le cadre du projet.

Exemples de référentiels utilisés dans la communauté archéologique :

- Personnes : [ISNI](#) et [ORCID](#)
- Lieux : [Geonames](#)
- Sujets du thesaurus [PACTOLS](#)
- Pour les institutions : [Research Organization Registry](#)

### Conditions de réutilisation\*

#### Recommandations:

Indiquez ici dans quel cadre se fait la réutilisation (licence ouverte, convention etc...)

Le régime de propriété auquel sont soumises les données composant le jeu de données est décrit dans la section 1 de l'onglet *Données : généralités*.

## Métadonnées associées

### Format des métadonnées

#### Recommandations:

Indiquer ici le format des métadonnées utilisé pour décrire le(s) fichier(s) composant le jeu de données

L'utilisation de standards est recommandée : [Dublin Core](#), [Inspire](#), [IPTC](#) ...

### Quel est le circuit de production des métadonnées ?

#### Recommandations:

Les métadonnées sont-elles produites dès la création des données, par le producteur de données lui-même, ou a posteriori ? Si c'est le cas, par qui et à quel moment ?

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

#### Recommandations:

Indiquer dans/avec quel(s) logiciel(s) les métadonnées sont renseignées.

### Quels sont les référentiels utilisés pour décrire les données ?

*Exemple de réponse:*

- Personnes : [ISNI](#) et [ORCID](#)
- Lieux : [Geonames](#)
- Sujets du thesaurus : [PACTOLS](#)
- Ontologie : CIDOC-CRM

*Recommandations:*

- Indiquez ici le ou les systèmes d'indexation (vocabulaires contrôlés, ontologie, *etc.*)
- Utiliser les référentiels du domaine et en prévoir, le cas échéant, l'enrichissement à l'occasion du projet.

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

*Recommandations:*

Préciser ici tout ce qui facilitera l'accès aux données sur le long terme : base de données de référencement, inventaire, *etc.*

Métadonnées d'origine \*

*Recommandations:*

Lorsqu'elles existent, décrire ici les métadonnées d'origine (format).

Métadonnées ajoutées \*

*Recommandations:*

Si de nouvelles métadonnées ont été ajoutées dans le cadre du projet, indiquer ici quel est leur format, comment elles ont été produites, qui était responsable de leur production.

## Stockage des données

Documentation papier produite au cours du projet

*Recommandations:*

Décrire le(s) type(s) de document produits exclusivement sur support papier en complément des données numériques. Indiquer aussi le conditionnement et le(s) lieu(x) de stockage des données autre que numérique.

Volumétrie prévisionnelle.

*Recommandations:*

Indiquer ici la volumétrie attendue des données produites au cours du projet, exprimée en mesure de stockage (Go, Mo, To, *etc.*). Cette estimation est utile pour prévoir l'espace de stockage nécessaire.

Infrastructure de stockage des données numériques durant le projet

*Exemple de réponse:*

- Données stockées sur la Huma-Num Box de la MSH de Nanterre mise à disposition par la TGIR Huma-Num, serveurs gérés à l'IN2P3 (CNRS).
- Modèles 3D stockés sur la plateforme Archéogrid d'Archéovision ([https://www.archeogrid.fr/home\\_conservatoire3d](https://www.archeogrid.fr/home_conservatoire3d)), mise à disposition par la TGIR Huma-Num, serveurs gérés à l'IN2P3 (CNRS).

*Recommandations:*

Décrire le matériel et le lieu de stockage des données numériques.

S'assurer que l'infrastructure de stockage est compatible avec la réglementation de votre institution.

Expliquer comment les données seront récupérées en cas d'incident.

Vous pouvez préciser les dispositifs de sauvegarde (à distinguer de l'archivage) cf guide ["Traçabilité des activités de recherche et gestion des connaissances : guide pratique de mise en place"](#) élaboré par la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS (2018), disponible sur Doranum.

## Sécurité des données

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

*Recommandations:*

Accès sécurisé ou pas ;

sauvegarde automatique ou pas

Garantie de confidentialité des données

*Recommandations:*

Préciser les mesures prises pour garantir :

- la protection des données à caractère personnel ;

- la sécurité des sites.

Garantie d'intégrité et de traçabilité

*Recommandations:*

Décrire ici les processus visant à garantir l'intégrité et la traçabilité des données. Certains aspects, le contrôle qualité, la gestion des accès, sont traités par ailleurs dans le plan de gestion de données et ne sont pas à répéter ici. Insister sur les procédures de sauvegarde, de dédoublement, de gestion des versions.

Vous pouvez vous reporter au guide "[Traçabilité des activités de recherche et gestion des connaissances : guide pratique de mise en place](#)" élaboré par la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS (2018), disponible sur Doranum.

## Accès aux données

Lecture des données

*Recommandations:*

Précisez ici si pour accéder à certaines données, l'utilisation de logiciels propriétaires est requise.

Comment la disponibilité des données est-elle garantie ?

*Recommandations:*

Indiquer si les données sont accessibles sur un serveur local, un intranet, sur internet, en libre accès ou avec un accès authentifié. Préciser, le cas échéant, les différents niveaux d'habilitations et les règles qui les régissent.

*Ce champ est à mettre à jour au fur et à mesure du déroulé du projet.*

Échange et partage

*Recommandations:*

Expliquer comment, au cours du projet, les données seront partagées et pourront être retrouvées (par exemple, par le dépôt dans un entrepôt de données de confiance, l'indexation dans un catalogue, par l'utilisation d'un service de données sécurisé, par le traitement direct des demandes de données, ou l'utilisation de tout autre mécanisme).

## Protection des données sensibles au cours du projet

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

*Recommandations:*

Listez les données sensibles et précisez ici la nature du risque ou de la contrainte pesant sur elles, susceptible de justifier des restrictions d'accès (ex : protection des personnes, des sites, des droits d'auteurs).

Vérifier la conformité de votre politique au règlement général sur la protection des données (RGPD).

Quelques ressources :

- [kit CNIL](#)
- [Atelier RGPD de la CNIL](#) (MOOC)

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

*Recommandations:*

Lister ici les mesures prises pour assurer la protection des données (élargir à l'ensemble des données, pas uniquement les données sensibles), sécurité des serveurs, des postes de travail, du réseau interne, gestion des incidents, etc..

## MASA Consortium: MASA - Modèle de PGD (français) - À l'issue du projet

### Politique d'archivage des données produites

Durée de conservation des données produites

*Recommandations:*

Se référer aux services dédiés de votre institution (exemple : Pôle National de Conservation des Données et Documents du CNRS, archivistes de votre MSH, de votre université...).

La politique d'archivage doit respecter le cadre réglementaire (Livre II du code du Patrimoine notamment).

Pour les unités de recherche et de service CNRS, noter l'existence de l'instruction n°DAF/DPACI/RES/2007/002 du 15 janvier 2007 : [Traitements et conservation des archives des délégations du Centre national de la recherche scientifique \(CNRS\) et des archives des unités de recherche et de service](#)

Vous pouvez également vous reporter au guide "[Traçabilité des activités de recherche et gestion des connaissances : guide pratique de mise en place](#)" élaboré par la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS (2018), disponible sur Doranum.

Volumétrie finale des données issues du projet

*Recommandations:*

Préciser la volumétrie globale, et par sous ensemble (ex: par jeu de données, par producteur, par site...).

Modalités de conservation des données à l'issue du projet

*Recommandations:*

Préciser les moyens techniques, procédures, lieux de conservation, moyens humains et institutionnels dédiés à la conservation des données.

Veillez vous reporter au guide "[Traçabilité des activités de recherche et gestion des connaissances : guide pratique de mise en place](#)" élaboré par la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS (2018), disponible sur Doranum.

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

*Recommandations:*

Préciser comment il sera possible d'accéder aux données.

### Dissémination des données

Politique de dissémination des données à l'issue du projet

*Exemple de réponse:*

Dissémination vers la communauté des archéologues français :

- Intégration de nouveaux concepts dans les thesaurus PACTOLS, en lien avec le GDS Frantiq.
- Intégration des jeux de données XX et YY dans la plateforme Open Archeo : <https://masa.hypotheses.org/openarchaeo> (triplestore dédié à l'archéologie développé dans le cadre du consortium MASA : il est doté d'une interface d'interrogation intuitive inspiré du moteur de recherche *Research Space* du British Museum)

Dissémination vers la communauté de recherche internationale :

- Publications d'articles avec un lien vers les données concernées, data papers. Dépôts dans HAL.
- Dépôt de jeux de données dans l'entrepôt certifié du Cines
- Données moissonnées par le moteur de recherche SHS Isidore (<https://isidore.science/>)

Dissémination vers les conservateurs du patrimoine et les médiateurs culturels:

- Participation à la conférence X sur la médiation culturelle et l'art rupestre

*Recommandations:*

En fonction de la politique de dissémination des données retenue, décrire ici les règles de diffusion s'appliquant aux différentes données ou jeux de données produits au cours du projet, en lien notamment avec les publics cibles. Indiquer où sont déposées et référencées les données à l'issue du projet.

L'ANR recommande "d'expliquer comment les données pourront être retrouvées et partagées (par exemple, par le dépôt dans un entrepôt de données de confiance, l'indexation dans un catalogue, par l'utilisation d'un service de données sécurisé, par le traitement direct des demandes de données, ou l'utilisation de tout autre mécanisme)."

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

*Exemple de réponse:*

- Data paper "jeu de données "peintures rupestres" du projet XXX". DOI

- Jean-Michel Chazine, Jean-Georges Ferrie. Recent archaeological discoveries in East Kalimantan, Indonesia BIPPA - Bulletin of the Indo-Pacific Prehistory Association , The Indo-Pacific Prehistory Association, 2008, 28, pp.16-22. <https://hal.archives-ouvertes.fr/hal-00342551v1>

*Recommandations:*

- Monographie, article scientifique, communication dans des congrès, data paper...
- Donner au moins : l'auteur, l'année de publication, le titre, l'éditeur et le lieu de publication. Préciser comment identifier et accéder à la publication et aux données associées
- Si une collection Zotero publique est associée au projet, la signaler.

L'ANR recommande d'"expliquer à quel moment les données seront rendues disponibles", d'"indiquer les délais de publication prévus", d'"expliquer si une utilisation exclusive des données est revendiquée et, dans l'affirmative, pour quelle raison et pour combien de temps".

Certaines données sont-elles sous embargo?

*Recommandations:*

*Embargo = Protection temporaire permettant de reporter la libre diffusion et/ou la réutilisation.*

Préciser lesquelles, pour quelle durée, pour quelle raison.

L'ANR recommande d'"indiquer si le partage des données sera différé ou limité, par exemple pour des raisons de publication, pour protéger la propriété intellectuelle, ou le dépôt de brevets".

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

*Exemple de réponse:*

Jeu de données relatif à l'iconographie des peintures rupestres (cf rubrique type de licence).

Fichier documentation\_peintures.txt associé au jeu de données.

Types de publics potentiels :

- conservateurs du patrimoine (musées)
- médiateurs culturels
- doctorants en préhistoire
- enseignants en (pré)histoire

*Recommandations:*

Préciser :

- le type de public ou les domaines d'application potentiels
- quels éléments permettent la bonne réutilisation des données : documentation associée, contextualisation, qualification des données (degrés de précision, de certitude...), DOIs...

L'ANR recommande d'"indiquer qui pourra utiliser les données. S'il s'avère nécessaire de restreindre l'accès pour certaines communautés ou d'imposer un accord pour le partage de données, expliquer comment et pourquoi. Expliquer les mesures qui seront prises pour dépasser ou minimiser ces restrictions."

Vous pouvez renvoyer, le cas échéant, à d'autres rubriques déjà renseignées du PGD

Conditions de réutilisation

*Exemple de réponse:*

Licence ouverte (Etalab). Citer les données sous la forme : Auteur (Année de publication) : Titre du jeu de données. Editeur. Identifiant du jeu de données (DOI)

*Recommandations:*

**1. Préciser les licences qui s'appliquent à l'ensemble des données du projet ou pour chaque jeu de données et en explicitent les modalités de réutilisation.**

Le droit français préconise l'utilisation de la licence ouverte (Etalab) :

<https://www.legifrance.gouv.fr/eli/decret/2017/4/27/2017-638/jo/texte>

Dans l'éventualité où une licence creative commons serait utilisée, la licence CC BY est préconisée :

<https://creativecommons.org/licenses/?lang=fr-FR>

**Éléments de documentation :**

Intervention de Jean-Charles Bédague, du service interministériel des Archives de France, le 19 septembre 2019 : *La réutilisation des informations publiques : un droit enfin stabilisé* <https://www.nakala.fr/data/11280/8cf441a2>

**Les licences ouvertes :**

- [Licence Etalab](#)
- [Licences Creative Commons](#)
- [Licences GNU](#)

**2. Vérifier l'existence des éléments nécessaires à la citation des données et donnez un exemple de la manière de citer les données que vous avez produites.** Datacite préconise de vérifier l'existence d'a minima ces 5

éléments : Auteur (Année de publication) : Titre. Editeur. Identifiant / *Creator (PublicationYear): Title. Publisher. Identifier*  
*cf recommandations du CIRAD (en français) : <https://coop-ist.cirad.fr/gerer-des-donnees/citer-un-jeu-de-donnees/2-comment-libeller-la-referance-bibliographique-d-un-jeu-de-donnees> :*

**3. Si des logiciels spécifiques et des compétences techniques et/ou scientifiques sont nécessaires pour pouvoir réutiliser les données ou jeux de données, donner ici les informations utiles.**