

# MODULE DOCTORAL - ATCHA (ACCOMPAGNER L'ADAPTATION DE L'AGRICULTURE IRRIGUÉE AU CHANGEMENT CLIMATIQUE) \_ V3

*Plan de gestion de données créé à l'aide de DMP OPIDoR*

**Créateurs du PGD :** Filatre Annabelle, Isabelle NAULT, Hanka Hensens

**Affiliation du créateur principal :** Other Organisation

**Modèle du PGD :** ANR - Modèle de PGD (français)

**Dernière modification du PGD :** 15/04/2021

**Financeur :** ANR

**Numéro de subvention :** ANR-16-CE03-0006

## **Résumé du projet :**

L'objectif du projet ATCHA est de combiner un modèle intégré et une approche participative pour accompagner l'adaptation agricoles au changement climatique dans un réseau de bassins versants expérimentaux dans le sud de l'Inde.

Grâce à une approche transdisciplinaire, associant hydrologues, géochimistes, pédologues, agronomes, géographes, économistes et sociologues et avec la participation de scientifiques et acteurs du développement agricole indiens, nous visons à démontrer la capacité des scientifiques et acteurs à partager leurs savoirs pour co-construire et évaluer des scénarios de développement durable.

Le projet ATCHA est composé de 3 parties impliquant chacune des défis scientifiques :

- i) le développement de nouvelles méthodologies pour obtenir des informations sur les sols et l'usage des terres à hautes résolutions spatiales et temporelles en associant mesures de terrain et données satellitaires
- ii) l'amélioration du réalisme du modèle par la calibration d'un grand nombre de cultures tropicales et l'analyse du cycle des nutriments dans les agrosystèmes irrigués tropicaux et
- iii) le développement d'une approche participative pour co-construire et évaluer des scénarios d'adaptation au changement climatique.

**Chercheur Principal :** RUIZ Laurent (UMR SAS - INRA - RENNES)

**Identifiant ORCID :** 0000-0001-5043-282X

**Contact pour les Données :** RUIZ Laurent (UMR SAS - INRA - RENNES)

## **Produits de recherche :**

1. Carte SIG Karnataka : Carte sur la typologie et l'utilisation des sols dans la région de Karnataka au sud de l'Inde ( Ressource interactive )
2. Modélisation ATCHA : Modélisation compartimentale dynamique des zones d'étude géographique ( Modèle )
3. Enquête ATCHA : Enquête qualitative auprès des populations clés ( Son )
4. BDD ATCHA : Base de données sur le fonctionnement des sols irrigués en zone tropicale ( Collection )
5. Données Satellite : Données satellite sur les types de sol et l'utilisation du sol du Sud de l'Inde ( Image )

# MODULE DOCTORAL - ATCHA (ACCOMPAGNER L'ADAPTATION DE L'AGRICULTURE IRRIGUÉE AU CHANGEMENT CLIMATIQUE) \_ V3

## CARTE SIG KARNATAKA : CARTE SUR LA TYPOLOGIE ET L'UTILISATION DES SOLS DANS LA RÉGION DE KARNATAKA AU SUD DE L'INDE

### 1. DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES

#### 1A. COMMENT DE NOUVELLES DONNÉES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNÉES PRÉEXISTANTES SERONT-ELLES RÉUTILISÉES ?

Les données seront issues de différents systèmes :

Données de terrain collectées sur les types de sols et l'utilisation du sol :

- méthodologie : observation / relevés d'échantillon / analyse biologique

Données collectées via le système ARGIS mis à disposition par le laboratoire CERCI (Toulouse - France)

- méthodologie : récupération des données via le système ARGIS

Le rassemblement de ces données donnera lieu à l'établissement de cartes (SIG).

#### 1B. QUELLES DONNÉES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTÉES OU PRODUITES ?

• Données de terrain collectées sur les types de sols et l'utilisation du sol donnant lieu à l'établissement de cartes (SIG) : format matriciel et vectoriel avec représentation en 3D de la surface du sol (Format TIF et projetée en L93/CC43 - Taille du pixel 20x20m)

• Données satellite sur les types de sol et l'utilisation du sol du Sud de l'Inde à acquérir : les messages codés des satellites au format DS, données de localisation au format TX et les paramètres de calcul au format DIAG (prévoit environ 5 pétaoctet)

### 2. DOCUMENTATION ET QUALITÉ DES DONNÉES

#### 2A. QUELLES MÉTADONNÉES ET QUELLE DOCUMENTATION (PAR EXEMPLE MÉTHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNÉES) ACCOMPAGNERONT LES DONNÉES ?

##### Métadonnées :

Métadonnées géographiques (cartes et images satellites) [FGDC/CSDGM- Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata](#)

##### Nommage et organisation des fichiers :

Les fichiers seront nommés selon le principe suivant :

- ACHA(1-5) \_type\_thème et/ou lieu\_version majeure\_annee-mois-jour.ext
- Exemple : ATCHA5\_CR\_berambadi7\_v3\_2017-08-05.odt
- Ils seront classés dans des répertoires et sous répertoires aux noms explicites de 25 caractères maximum contenant au plus 25 fichiers.
- Les versions majeures (décision prise en réunion d'équipe) donneront lieu à la destruction de toutes les versions intermédiaires les séparant de la version majeure précédente. A défaut d'une telle décision, une version mensuelle seulement sera conservée le 1er jour ouvré du mois suivant.

#### 2B. QUELLES MESURES DE CONTRÔLE DE LA QUALITÉ DES DONNÉES SERONT MISES EN ŒUVRE ?

Le suivi des procédures INRAE (documentation précise des différentes étapes d'acquisition et de traitement des données, note de tous les problèmes rencontrés, cahiers de recherche) assurera qualité et traçabilité des données.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### ***3A. COMMENT LES DONNÉES ET LES MÉTADONNÉES SERONT-ELLES STOCKÉES ET SAUVEGARDÉES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?***

Les données collectées sur place sont engrangées sur tablettes ou ordinateurs portables. Elles sont transférées chaque soir sur Disque Dur externe et, dès que le réseau le permet, du Disque Dur sur le Cloud par FTP.

3 espaces de stockage prévus :

- Disque Durs externes
- Cloud
- Serveur sur réseau interne INRA

Les données seront transférées et sauvegardées automatiquement chaque nuit. Les versions quotidiennes seront conservées 7 jours, les versions dominicales seront conservées 4 semaines, et les versions du 1er de chaque mois seront conservées jusqu'à la fin du projet, afin de permettre une éventuelle récupération.

L'ingénieur Données sera responsable de la mise en place du système de transfert, stockage et sauvegarde des données durant toute la durée du projet. A l'issue du projet l'entrepôt Zenodo, d'une part et le CINES, d'autre part prendront le relai pour la sauvegarde à moyen et long terme.

Durant le projet, les ressortissants des UMR (AGIR, EMMAH, GET SAS TSE-R) et organismes (IFP et Univ Rennes 2) partenaires du projet et les organisations indiennes partenaires des projets antérieurs (CEFIRSE, ORE BVET, Watershed Department du Karnataka) habilités auront accès à tout ou partie des données produites ou acquises sur login/mot de passe attribués par les responsables thématiques des données, Mme Y et MM X et Z, dans le respect des recommandations de leurs organismes de tutelle.

Volumétrie prévue : données 1 : xxx To, Données 2 : yyy To, etc...

NB : cette estimation sera précisée dans les versions ultérieures de ce plan

#### ***3B. COMMENT LA SÉCURITÉ DES DONNÉES ET LA PROTECTION DES DONNÉES SENSIBLES SERONT-ELLES ASSURÉES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?***

Les procédures de sécurité informatique de l'INRA (réseau) et du CNRS (cloud) seront appliquées

### **4. EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE**

#### ***4A. SI DES DONNÉES À CARACTÈRE PERSONNEL SONT TRAITÉES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LÉGISLATION SUR LES DONNÉES À CARACTÈRE PERSONNEL ET SUR LA SÉCURITÉ DES DONNÉES SERA-T-IL ASSURÉ ?***

Il n'y pas de traitement de données à caractère personnel dans ce jeu de données

#### ***4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITÉ OU LES DROITS DE PROPRIÉTÉ INTELLECTUELLE SUR LES DONNÉES, SÉRONT-ELLES ABORDÉES ? QUELLE EST LA LÉGISLATION APPLICABLE EN LA MATIÈRE ?***

Les données de départ étant issues de différents organismes, des conventions sont en cours de relecture par les services juridiques des partenaires avant signature par leurs dirigeants.

Elles autoriseront l'utilisation et la modification des données agronomiques, hydrologiques et pédologiques fournies par l'ensemble des partenaires pendant le projet et leur partage au plus grand nombre 24 mois après la fin du projet . Ces données étant en grande partie issues des savoirs paysans, elles doivent rester réutilisables par les paysans eux même et ne doivent donc pas pouvoir être brevetables. Elles seront donc publiées sous licence CC BY NC

#### ***4C. COMMENT LES ÉVENTUELLES QUESTIONS ÉTHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DÉONTOLOGIQUES RESPECTÉS ?***

Toutes les données seront analysées par le juriste de l'INRA permettant ainsi de vérifier le bon traitement des questions éthiques selon la norme ISO 7498-2 (ISO90).

Un comité d'éthique composé des membres du projet, du directeur du LISAH et de la directrice adjointe de l'INRA sera monté pour assurer le respect des codes déontologiques

## **5. PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME**

### **5A. COMMENT ET QUAND LES DONNÉES SERONT-ELLES PARTAGÉES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNÉES OU DES RAISONS DE DÉFINIR UN EMBARGO ?**

Les données déposées sur Zenodo disposeront chacune d'un DOI (Digital Object Identifier) permettant leur identification pérenne au cours du temps. Ces DOI seront notamment utilisés dans les publications scientifiques générées à partir de ces données. Pendant la durée du projet, les fichiers de données seront déposés en accès à la demande, le temps que les équipes de recherche publient les résultats de leurs analyses. Une fois le projet terminé et les articles correspondants publiés, les fichiers anonymisés de données quantitatives seront en accès libre et distribués sous licence Creative Commons - Attribution - Partage dans les Mêmes Conditions 4.0 International (<https://creativecommons.org/licenses/by-sa/4.0/>).

### **5B. COMMENT LES DONNÉES À CONSERVER SERONT-ELLES SÉLECTIONNÉES ET OÙ SERONT-ELLES PRÉSERVÉES SUR LE LONG TERME (PAR EX. UN ENTREPÔT DE DONNÉES OU UNE ARCHIVE) ?**

Les données personnelles seront détruites à la fin du projet, lors de l'anonymisation qui fera disparaître tout élément permettant de reconnaître les participants. Toutes les autres données seront conservées à long terme au CINES (convention en cours de rédaction)

### **5C. QUELLES MÉTHODES OU QUELS OUTILS LOGICIELS SERONT NÉCESSAIRES POUR ACCÉDER ET UTILISER LES DONNÉES ?**

Tout le matériel ainsi que les méthodologies et les logiciels seront fournis avec les jeux de données et déposés dans Zenodo.

Les liens vers des outils à accès ouvert pouvant permettre la réutilisation des données seront également disponibles.

### **5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PÉRENNE (COMME LE DOI) SERA-T-ELLE ASSURÉE POUR CHAQUE JEU DE DONNÉES ?**

Pour chaque jeu de données, les DOI seront attribués par le service INRA.

## **6. RESPONSABILITÉS ET RESSOURCES EN MATIÈRE DE GESTION DES DONNÉES**

### **6A. QUI (PAR EXEMPLE RÔLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNÉES (C'EST-À-DIRE LE GESTIONNAIRE DES DONNÉES) ?**

Chaque work package est coordonné par un responsable scientifique (voir protocole) qui est en charge de s'assurer de la bonne conduite des activités de son work package. L'investigateur principal s'assurera de la bonne exécution globale du plan de gestion de données. Un organigramme détaillé des rôles et responsabilités sera développé en début de projet. Les ressources humaines, matérielles et logicielles requises pour la mise en œuvre du plan de gestion des données ont été prévues et incluses dans le budget du projet.

### **6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUÉS) DÉDIÉES À LA GESTION DES DONNÉES PERMETTANT DE S'ASSURER QUE LES DONNÉES SERONT FAIR (FACILE À TROUVER, ACCESSIBLE, INTEROPÉRABLE, RÉUTILISABLE) ?**

#### **RH :**

- Recrutement Ingénieur de Recherche responsable des données en France : 48 mois (coût mensuel charges comprises à préciser)
- Recrutement 15 techniciens en CDD de 24 mois en Inde pour recueil, la saisie et le transfert des données de terrain sur les sols et leur utilisation, d'une part (données 1), la traduction et la prise de note, d'autre part (données 5) : 15 x 24 mois x coût mensuel charges comprises
- Formation des responsables sectoriels des données en France sur les logiciels choisis, et sur les protocoles de qualité (y compris métadonnées), de sauvegarde et de transfert des données :
  - Données 1 et 2 : M. X, IRD
  - Données 3 et 4 : Mme Y, Inrae

- Données 5 : M. T, CNRS
- Frais de transport et de mission de ces responsables pour la formation de ces techniciens sur place à la saisie et la prise de note sur tablette et le transfert sécurisé via internet

**Logiciels :**

Professionnels, standards, et libres si possibles : SIG, BDD, texte mining, encodage des données avant transfert, ...(reste à préciser)

**Images satellites :**

Fournisseur, satellite, résolution, filtre, ...

**Matériel :**

- X PC boostés, X PC portables, X tablettes
- X Disques Durs externes
- Espace cloud sur serveur public français
- Baie de serveur

## **MODÉLISATION ATCHA : MODÉLISATION COMPARTIMENTALE DYNAMIQUE DES ZONES D'ÉTUDE GÉOGRAPHIQUE**

### **1. DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES**

#### **1A. COMMENT DE NOUVELLES DONNÉES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNÉES PRÉEXISTANTES SERONT-ELLES RÉUTILISÉES ?**

La modélisation se fera à partir des données réunies dans les deux bases de données précédemment citées à savoir :

- Données satellite sur les types de sol et l'utilisation du sol du Sud de l'Inde à acquérir auprès de GéoSud (Montpellier). La demande est en cours ainsi que le devis pour l'acquisition de ces données.
- Données agronomiques sur diverses cultures tropicales issues de la littérature scientifique et des recherches des partenaires (IRD, INRA, CNRS, Services agronomiques des Etats du Sud de l'Inde,...) rassemblées et homogénéisées dans une base de données créée pour alimenter notre modèle intégré

#### **1B. QUELLES DONNÉES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTÉES OU PRODUITES ?**

Les données collectées seront issues des deux bases de données et de 3 natures :

- Données type texte (format .txt)
- Données de type radar (TerraSAR-X et CosmoSkyMed)
- format .jpg ou en haute définition (format .pdf)

### **2. DOCUMENTATION ET QUALITÉ DES DONNÉES**

#### **2A. QUELLES MÉTADONNÉES ET QUELLE DOCUMENTATION (PAR EXEMPLE MÉTHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNÉES) ACCOMPAGNERONT LES DONNÉES ?**

Le choix de l'outil de modélisation et les différents procédés expérimentaux seront explicités dans un document partagé en ligne et sécurisé de type "carnet de recherche".

#### **2B. QUELLES MESURES DE CONTRÔLE DE LA QUALITÉ DES DONNÉES SERONT MISES EN ŒUVRE ?**

Le suivi des procédures INRA (documentation précise des différentes étapes d'acquisition et de traitement des données, note de tous les problèmes rencontrés, cahiers de recherche) assurera qualité et traçabilité des données.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### **3A. COMMENT LES DONNÉES ET LES MÉTADONNÉES SERONT-ELLES STOCKÉES ET SAUVEGARDÉES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les données collectées sur place sont engrangées sur tablettes ou ordinateurs portables. Elles sont transférées chaque soir sur Disque Dur externe et, dès que le réseau le permet, du Disque Dur sur le Cloud par FTP.

3 espaces de stockage prévus :

- Disque Durs externes
- Cloud
- Serveur sur réseau interne INRA

Les données seront transférées et sauvegardées automatiquement chaque nuit. Les versions quotidiennes seront conservées 7 jours, les versions dominicales seront conservées 4 semaines, et les versions du 1er de chaque mois seront conservées jusqu'à la fin du projet, afin de permettre une éventuelle récupération.

L'ingénieur Données sera responsable de la mise en place du système de transfert, stockage et sauvegarde des données durant toute la durée du projet. A l'issue du projet l'entrepôt Zenodo, d'une part et le CINES, d'autre part prendront le relais pour la sauvegarde à moyen et long terme.

Durant le projet, les ressortissants des UMR (AGIR, EMMAH, GET SAS TSE-R) et organismes (IFP et Univ Rennes 2) partenaires du projet et les organisations indiennes partenaires des projets antérieurs (CEFIRSE, ORE BVET, Watershed Department du Karnataka) habilités auront accès à tout ou partie des données produites ou acquises sur login/mot de passe attribués par les responsables thématiques des données, Mme Y et MM X et Z, dans le respect des recommandations de leurs organismes de tutelle.

Volumétrie prévue : données 1 : xxx To, Données 2 : yyy To, etc...

NB : cette estimation sera précisée dans les versions ultérieures de ce plan

### **3B. COMMENT LA SÉCURITÉ DES DONNÉES ET LA PROTECTION DES DONNÉES SENSIBLES SERONT-ELLES ASSURÉES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les procédures de sécurité informatique de l'INRA (réseau) et du CNRS (cloud) seront appliquées

## **4. EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE**

### **4A. SI DES DONNÉES À CARACTÈRE PERSONNEL SONT TRAITÉES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LÉGISLATION SUR LES DONNÉES À CARACTÈRE PERSONNEL ET SUR LA SÉCURITÉ DES DONNÉES SERA-T-IL ASSURÉ ?**

Il n'y a pas de traitement de données à caractère personnel dans ce jeu de données

### **4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITÉ OU LES DROITS DE PROPRIÉTÉ INTELLECTUELLE SUR LES DONNÉES, SERONT-ELLES ABORDÉES ? QUELLE EST LA LÉGISLATION APPLICABLE EN LA MATIÈRE ?**

Les données de départ étant issues de différents organismes, des conventions sont en cours de relecture par les services juridiques des partenaires avant signature par leurs dirigeants.

Elles autoriseront l'utilisation et la modification des données agronomiques, hydrologiques et pédologiques fournies par l'ensemble des partenaires pendant le projet et leur partage au plus grand nombre 24 mois après la fin du projet. Ces données étant en grande partie issues des savoirs paysans, elles doivent rester réutilisables par les paysans eux même et ne doivent donc pas pouvoir être brevetables. Elles seront donc publiées sous licence CC BY NC

### **4C. COMMENT LES ÉVENTUELLES QUESTIONS ÉTHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DÉONTOLOGIQUES RESPECTÉS ?**

Toutes les données seront analysées par le juriste de l'INRA permettant ainsi de vérifier le bon traitement des questions éthiques selon la norme ISO 7498-2 (ISO90).

Un comité d'éthique composé des membres du projet, du directeur du LISAH et de la directrice adjointe de l'INRA sera monté pour assurer le respect des codes déontologiques

## **5. PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME**

### **5A. COMMENT ET QUAND LES DONNÉES SERONT-ELLES PARTAGÉES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNÉES OU DES RAISONS DE DÉFINIR UN EMBARGO ?**

- Achetées sous licence, elles ne pourront être partagées à d'autres personnes que les membres des institutions partenaires du projet (sur login/mot de passe sur le serveur de la Maison de la Télédétection).

### **5B. COMMENT LES DONNÉES À CONSERVER SERONT-ELLES SÉLECTIONNÉES ET OÙ SERONT-ELLES PRÉSERVÉES SUR LE LONG TERME (PAR EX. UN ENTREPÔT DE DONNÉES OU UNE ARCHIVE) ?**

Les données personnelles seront détruites à la fin du projet, lors de l'anonymisation qui fera disparaître tout élément permettant de reconnaître les participants. Toutes les autres données seront conservées à long terme au CINES (convention en cours de rédaction)

**5C. QUELLES MÉTHODES OU QUELS OUTILS LOGICIELS SERONT NÉCESSAIRES POUR ACCÉDER ET UTILISER LES DONNÉES ?**

Tout le matériel ainsi que les méthodologies et les logiciels seront fournis avec les jeux de données et déposés dans Zenodo.

Les liens vers des outils à accès ouvert pouvant permettre la réutilisation des données seront également disponibles.

**5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PÉRENNE (COMME LE DOI) SERA-T-ELLE ASSURÉE POUR CHAQUE JEU DE DONNÉES ?**

Pour chaque jeu de données, les DOI seront attribués par le service INRA.

**6. RESPONSABILITÉS ET RESSOURCES EN MATIÈRE DE GESTION DES DONNÉES**

**6A. QUI (PAR EXEMPLE RÔLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNÉES (C'EST-À-DIRE LE GESTIONNAIRE DES DONNÉES) ?**

Chaque work package est coordonné par un responsable scientifique (voir protocole) qui est en charge de s'assurer de la bonne conduite des activités de son work package. L'investigateur principal s'assurera de la bonne exécution globale du plan de gestion de données. Un organigramme détaillé des rôles et responsabilités sera développé en début de projet. Les ressources humaines, matérielles et logicielles requises pour la mise en œuvre du plan de gestion des données ont été prévues et incluses dans le budget du projet.

**6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUÉS) DÉDIÉES À LA GESTION DES DONNÉES PERMETTANT DE S'ASSURER QUE LES DONNÉES SERONT FAIR (FACILE À TROUVER, ACCESSIBLE, INTEROPÉRABLE, RÉUTILISABLE) ?**

**RH :**

- Recrutement Ingénieur de Recherche responsable des données en France : 48 mois (coût mensuel charges comprises à préciser)
- Recrutement 15 techniciens en CDD de 24 mois en Inde pour recueil, la saisie et le transfert des données de terrain sur les sols et leur utilisation, d'une part (données 1), la traduction et la prise de note, d'autre part (données 5) : 15 x 24 mois x coût mensuel charges comprises
- Formation des responsables sectoriels des données en France sur les logiciels choisis, et sur les protocoles de qualité (yc métadonnées), de sauvegarde et de transfert des données :
  - Données 1 et 2 : M. X, IRD
  - Données 3 et 4 : Mme Y, Inrae
  - Données 5 : M. T, CNRS
- Frais de transport et de mission de ces responsables pour la formation de ces techniciens sur place à la saisie et la prise de note sur tablette et le transfert sécurisé via internet

**Logiciels :**

Professionnels, standards, et libres si possibles : SIG, BDD, texte mining, encodage des données avant transfert, ...(reste à préciser)

**Images satellites :**

Fournisseur, satellite, résolution, filtre, ...

**Matériel :**

- X PC boostés, X PC portables, X tablettes
- X Disques Durs externes
- Espace cloud sur serveur public français
- Baie de serveur

## **ENQUÊTE ATCHA : ENQUÊTE QUALITATIVE AUPRÈS DES POPULATIONS CLÉS**

### **1. DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES**

#### **1A. COMMENT DE NOUVELLES DONNÉES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNÉES PRÉEXISTANTES SERONT-ELLES RÉUTILISÉES ?**

Il s'agit d'entretiens individuels et de groupe. Chaque entretien sera numéroté en fonction de l'enquête ou de la sous-enquête à laquelle il correspond et du pays de collecte.

#### **1B. QUELLES DONNÉES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTÉES OU PRODUITES ?**

Les entretiens seront enregistrés (audio, format MP3 ou équivalent) avec le consentement des participants. Pour les entretiens individuels, si l'enquêté·e refuse l'enregistrement audio, l'enquêteur·rice procédera à une prise de notes manuelles.

### **2. DOCUMENTATION ET QUALITÉ DES DONNÉES**

#### **2A. QUELLES MÉTADONNÉES ET QUELLE DOCUMENTATION (PAR EXEMPLE MÉTHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNÉES) ACCOMPAGNERONT LES DONNÉES ?**

Les enregistrements audio et les prises de note feront l'objet d'une retranscription (fichier texte avec mise en forme) en français (avec traduction le cas échéant pour les entretiens menés dans une autre langue). Chaque retranscription d'entretien sera accompagnée d'un rapport d'entretien incluant : le contexte de passation de l'entretien, les caractéristiques sociodémographiques et une présentation de la ou des personnes interrogées et d'un résumé.

#### **2B. QUELLES MESURES DE CONTRÔLE DE LA QUALITÉ DES DONNÉES SERONT MISES EN ŒUVRE ?**

Le numéro d'entretien sera reporté sur le formulaire de consentement. Les entretiens seront menés par des anthropologues seniors avec l'appui d'assistant·e·s de recherche formées aux sciences sociales et à la conduite d'entretien.

L'équipe qualitative se réunira à intervalles réguliers pour coordonner les activités de terrain et vérifier la qualité des entretiens.

La contrôle qualité de la collecte des données sera assuré par la mise en place de contrôles de cohérence au moment de la saisie, la double-écoute par les superviseurs d'entretiens conduits par les enquêteurs et des vérifications sur la base analytique en cours d'enquête. L'enquête coupons étant organisée en 3 vagues, une analyse poussée des données de la première vague sera réalisée avant la conduite des vagues 2 et 3 et les procédures opérationnelles seront ajustés en fonction.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### **3A. COMMENT LES DONNÉES ET LES MÉTADONNÉES SERONT-ELLES STOCKÉES ET SAUVEGARDÉES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les enregistrements audio des entretiens seront conservés sur l'ordinateur personnel du/de la chercheur·e chargé(e) de la conduite de l'entretien et protégés par mot de passe.

Les entretiens retranscrits et "pseudonymisés" seront déposés sur la plateforme dédiée à l'analyse collaborative de données qualitatives et notamment à la codification thématique d'entretiens. Cette plateforme est un produit développé et commercialisé par la société Scientific Software Development GmbH spécialisée dans les logiciels d'analyses qualitatives scientifiques. La plateforme est sécurisée et les accès individuels et privés. Selon la politique de confidentialité du service, la plateforme respecte le règlement général sur la protection des données (RGPD) de l'Union Européenne. Les communications seront chiffrées (https), les accès individuels (compte utilisateur) avec gestion des droits.

### **3B. COMMENT LA SÉCURITÉ DES DONNÉES ET LA PROTECTION DES DONNÉES SENSIBLES SERONT-ELLES ASSURÉES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les données personnelles (données nominatives ou données permettant l'identification des personnes par recoupement d'information) seront conservées au maximum pendant 5 ans après la fin du projet, soit jusque fin 2026 au plus. De même, les formulaires de consentement seront conservés pendant la même période afin de permettre aux personnes enquêtées d'exercer leur droit d'accès, de rectification et/ou d'opposition (voir section 3.1). Seules des données anonymisées générées à partir des données brutes feront l'objet d'un archivage.

## **4. EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE**

### **4A. SI DES DONNÉES À CARACTÈRE PERSONNEL SONT TRAITÉES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LÉGISLATION SUR LES DONNÉES À CARACTÈRE PERSONNEL ET SUR LA SÉCURITÉ DES DONNÉES SERA-T-IL ASSURÉ ?**

Une procédure d'anonymisation des retranscriptions pseudonymisées des entretiens sera mise en place. Les rapports d'entretiens seront repris un à un et tous les passages contenant des informations susceptibles d'identifier la personne interrogée ou celles qu'elle mentionne seront supprimés (avec indication de la suppression). En cas de doute, la suppression sera favorisée. Les rapports d'entretiens anonymisés seront validés collégalement par les équipes de recherche qualitatives.

### **4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITÉ OU LES DROITS DE PROPRIÉTÉ INTELLECTUELLE SUR LES DONNÉES, SERONT-ELLES ABORDÉES ? QUELLE EST LA LÉGISLATION APPLICABLE EN LA MATIÈRE ?**

Les droits de propriété intellectuelle sur les données du projets et leurs dérivés restent acquises par les différentes parties ayant créé ou produit ces données. Cependant, les différents membres du consortium Atlas se sont engagés à partager les données produites entre eux. Toute publication ou communication scientifique réalisée dans le cadre du projet ATLAS et portant sur les données du projet devra être discutée en amont par le comité de pilotage scientifique du projet et s'inscrire dans le plan de valorisation du projet. Les données de recherche anonymisée (voir section 5. Sélection et Préservation) qui feront l'objet d'un archivage en libre accès seront distribuées sous licence Creative Commons - Attribution - Partage dans les Mêmes Conditions 4.0 International (<https://creativecommons.org/licenses/by-sa/4.0/>). Les institutions de recherche du projet se sont par ailleurs toute engagées contractuellement à fournir une sous-licence non-exclusive, irrévocable, mondiale, libre de droits et sous-licence à l'OMS pour le compte d'UNITAID pour utiliser les données anonymisées (voir section 5. Sélection et Préservation) du projet et les publications scientifiques produites à des fins non commerciales de santé publique, d'éducation et de recherche.

### **4C. COMMENT LES ÉVENTUELLES QUESTIONS ÉTHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DÉONTOLOGIQUES RESPECTÉS ?**

Lors de la retranscription, les entretiens seront codés (changement des noms et prénoms par exemple) afin de les pseudonymiser.

## **5. PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME**

### **5A. COMMENT ET QUAND LES DONNÉES SERONT-ELLES PARTAGÉES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNÉES OU DES RAISONS DE DÉFINIR UN EMBARGO ?**

Concernant les données qualitatives, considérant que le processus d'anonymisation décrit précédemment pourrait être source d'erreur et que l'identification de certaines des personnes enquêtées pourrait toujours être théoriquement possible par croisement d'informations, l'accès aux données archivées sera restreint. L'accès, via la plateforme Zenodo, se fera sur demande, par des équipes académiques et sera soumis à la signature d'un engagement de confidentialité.

### **5B. COMMENT LES DONNÉES À CONSERVER SERONT-ELLES SÉLECTIONNÉES ET OÙ SERONT-ELLES PRÉSERVÉES SUR LE LONG TERME (PAR EX. UN ENTREPÔT DE DONNÉES OU UNE ARCHIVE) ?**

Les données personnelles seront détruites à la fin du projet, lors de l'anonymisation qui fera disparaître tout élément permettant de reconnaître les participants. Toutes les autres données seront conservées à long terme au CINES (convention en cours de rédaction)

**5C. QUELLES MÉTHODES OU QUELS OUTILS LOGICIELS SERONT NÉCESSAIRES POUR ACCÉDER ET UTILISER LES DONNÉES ?**

Tout le matériel ainsi que les méthodologies et les logiciels seront fournis avec les jeux de données et déposer dans Zenodo.

Les liens vers des outils en accès ouvert pouvant permettre la réutilisation des données seront également disponibles.

**5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PÉRENNE (COMME LE DOI) SERA-T-ELLE ASSURÉE POUR CHAQUE JEU DE DONNÉES ?**

Pour chaque jeu de données, les DOI seront attribués par le service INRA.

**6. RESPONSABILITÉS ET RESSOURCES EN MATIÈRE DE GESTION DES DONNÉES**

**6A. QUI (PAR EXEMPLE RÔLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNÉES (C'EST-À-DIRE LE GESTIONNAIRE DES DONNÉES) ?**

Chaque work package est coordonné par un responsable scientifique (voir protocole) qui est en charge de s'assurer de la bonne conduite des activités de son work package. L'investigateur principal s'assurera de la bonne exécution globale du plan de gestion de données. Un organigramme détaillé des rôles et responsabilités sera développé en début de projet. Les ressources humaines, matérielles et logicielles requises pour la mise en œuvre du plan de gestion des données ont été prévues et incluses dans le budget du projet.

**6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUÉS) DÉDIÉES À LA GESTION DES DONNÉES PERMETTANT DE S'ASSURER QUE LES DONNÉES SERONT FAIR (FACILE À TROUVER, ACCESSIBLE, INTEROPÉRABLE, RÉUTILISABLE) ?**

**RH :**

- Recrutement Ingénieur de Recherche responsable des données en France : 48 mois (coût mensuel charges comprises à préciser)
- Recrutement 15 techniciens en CDD de 24 mois en Inde pour recueil, la saisie et le transfert des données de terrain sur les sols et leur utilisation, d'une part (données 1), la traduction et la prise de note, d'autre part (données 5) : 15 x 24 mois x coût mensuel charges comprises
- Formation des responsables sectoriels des données en France sur les logiciels choisis, et sur les protocoles de qualité (yc métadonnées), de sauvegarde et de transfert des données :
  - Données 1 et 2 : M. X, IRD
  - Données 3 et 4 : Mme Y, Inrae
  - Données 5 : M. T, CNRS
- Frais de transport et de mission de ces responsables pour la formation de ces techniciens sur place à la saisie et la prise de note sur tablette et le transfert sécurisé via internet

**Logiciels :**

Professionnels, standards, et libres si possibles : SIG, BDD, texte mining, encodage des données avant transfert, ...(reste à préciser)

**Images satellites :**

Fournisseur, satellite, résolution, filtre, ...

**Matériel :**

- X PC boostés, X PC portables, X tablettes
- X Disques Durs externes
- Espace cloud sur serveur public français
- Baie de serveur

## **BDD ATCHA : BASE DE DONNÉES SUR LE FONCTIONNEMENT DES SOLS IRRIGUÉS EN ZONE TROPICALE**

### **1. DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES**

#### **1A. COMMENT DE NOUVELLES DONNÉES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNÉES PRÉEXISTANTES SERONT-ELLES RÉUTILISÉES ?**

Ces données sur le fonctionnement des sols irrigués en zone tropicale : cycle des nutriments, fertilité,... seront extraites de la littérature scientifique et des recherches des partenaires (IRD, INRA, CNRS, Services agronomiques des Etats du Sud de l'Inde,...) par deux des doctorants travaillant sur le projet, en liaison étroite avec leur responsable, C. Aubron. Le but poursuivi est d'enrichir la réflexion théorique et de pouvoir déduire des constantes et modéliser le fonctionnement des sols irrigués en fonction des inputs (eau, température, travail du sol, type de culture...).

Le corpus des documents sources sera établi de façon transparente, objective et rigoureuse sous la direction experte de B. Livoreil, selon la méthode des cartographies systématiques. La base bibliographique sera constituée sous Zotero durant l'année 1 du projet et un document de suivi méthodologique sera complété au fur et à mesure des opérations (Protocole, équation de recherche, sources, sélection,...).

Les données extraites seront rassemblées et homogénéisées dans une base de données créée et maintenue sous MySQL par l'ingénieur informaticien Mme Y. de l'UMR EMMAH, en liaison avec l'équipe BASMATI (BASE de données, Modèles, Applications et Techniques Informatiques) de l'UMR SAS (Sol Agro et hydrosystème Spatialisation), partenaires du projet.

#### **1B. QUELLES DONNÉES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTÉES OU PRODUITES ?**

Les données rassemblées seront de 3 types :

- base bibliographique Zotero partagée online, exportée au format .csv en fin de projet. Volume inférieur à 1 Go.
- fichiers textes explicitant la constitution du corpus bibliographique. Volume : quelques Mo
- base de données MySQL (fichiers .SQL, .MYD et .FRM) comporteront une 50taine de tables et ne dépasseront pas 1 To. Ils seront partagés en fin de projet en format Open Document (.odb)

### **2. DOCUMENTATION ET QUALITÉ DES DONNÉES**

#### **2A. QUELLES MÉTADONNÉES ET QUELLE DOCUMENTATION (PAR EXEMPLE MÉTHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNÉES) ACCOMPAGNERONT LES DONNÉES ?**

Les données issues des 2 bases de données seront accompagnées :

- Pour la base de données bibliographique, de fichiers texte explicitant la procédure de constitution du corpus (cf. 1a)
- Pour la base de données SQL, de fichiers textes dictionnaires répertoriant les abréviations, les unités de mesure, etc. utilisés dans les différentes tables et d'un fichier explicitant l'organisation de la base de données

Par ailleurs, ces 2 bases de données seront décrites à l'aide des Métadonnées thématiques agronomiques et pédologiques standard [AgMES - Agricultural Metadata Element Set](#)

#### **2B. QUELLES MESURES DE CONTRÔLE DE LA QUALITÉ DES DONNÉES SERONT MISES EN ŒUVRE ?**

Toutes les étapes de la collecte, de la saisie, du contrôle, du traitement et de l'analyse des données seront décrites à l'aide de vocabulaires standards par les ingénieurs et chercheurs du projet dans les cahiers de Laboratoire repris dans les fichiers lisez-moi déposés dans l'entrepôt avec les jeux de données.

La politique de sauvegarde et de stockage décrite au §3 minimisera les risques de perte ou de corruption des fichiers.

L'ensemble de la recherche respectera la politique qualité de l'INRA (lien vers document), pour le respect de la traçabilité et de la fiabilité des résultats.

## Nommage et organisation des fichiers :

- Les fichiers sauvegardés seront nommés selon le principe suivant :  
ATCHA(1-5)\_type\_thème et/ou lieu\_version majeure\_annee-mois-jour.ext  
Exemple : ATCHA4\_BDB\_v3\_2017-08-05.odt
- Ils seront classés dans des répertoires et sous répertoires aux noms explicites de 25 caractères maximum contenant au plus 25 fichiers.
- Les versions majeures (décision prise en réunion d'équipe) donneront lieu à la destruction de toutes les versions intermédiaires les séparant de la version majeure précédente. A défaut d'une telle décision, une version mensuelle seulement sera conservée le 1er jour ouvré du mois suivant.

## 3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

### 3A. COMMENT LES DONNÉES ET LES MÉTADONNÉES SERONT-ELLES STOCKÉES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Les fichiers et bases de données seront hébergés et/ou sauvegardés sur des serveurs de l'INRA (espace réservé au projet) et dans le cloud (pour la base Zotero, exportées toutes les semaines sur serveur INRA).

Les données des serveurs INRA sont sauvegardées automatiquement chaque nuit par l'hébergeur.

Les versions quotidiennes seront conservées 7 jours, les versions dominicales seront conservées 6 semaines, et les versions du 1er de chaque mois seront conservées jusqu'à la fin du projet, afin de permettre une éventuelle récupération.

L'ingénieur Données sera responsable du stockage et de la sauvegarde des données durant toute la durée du projet. A l'issue du projet l'entrepôt Dataverse INRA, d'une part et le CINES, d'autre part prendront le relai pour la sauvegarde à moyen et long terme cf. §5).

### 3B. COMMENT LA SÉCURITÉ DES DONNÉES ET LA PROTECTION DES DONNÉES SENSIBLES SERONT-ELLES ASSURÉES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Les procédures de sécurité informatique de l'INRA (lien) seront appliquées, en particulier pour la récupération des données en cas d'incident.

Durant le projet, les ressortissants des UMR (AGIR, EMMAH, GET SAS TSE-R) et organismes (IFP et Univ Rennes 2) partenaires du projet et les organisations indiennes partenaires des projets antérieurs (CEFIRSE, ORE BVET, Watershed Department du Karnataka) habilités auront accès aux données sur login/mot de passe attribués par la responsable des données, Mme Y, dans le respect des recommandations de l'INRA.

## 4. EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE

### 4A. SI DES DONNÉES À CARACTÈRE PERSONNEL SONT TRAITÉES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LÉGISLATION SUR LES DONNÉES À CARACTÈRE PERSONNEL ET SUR LA SÉCURITÉ DES DONNÉES SERA-T-IL ASSURÉ ?

Il n'y a pas de traitement de données à caractère personnel dans ce jeu de données

### 4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITÉ OU LES DROITS DE PROPRIÉTÉ INTELLECTUELLE SUR LES DONNÉES, SERONT-ELLES ABORDÉES ? QUELLE EST LA LÉGISLATION APPLICABLE EN LA MATIÈRE ?

L'ensemble des partenaires est copropriétaire des données produites par le projet, et délègue leur gestion durant le projet aux UMR SAS et EMMAH. Une partie de l'accord de consortium signé par l'ensemble des organismes partenaires détaille les droits d'accès, d'utilisation, et de partage des données.

Les données extraites des publications et rapports sont en principe libres de droit (vérification sera faite en cas de doute, et les données ne seront pas intégrées si ce n'était pas le cas).

Par principe, ces données étant issues de la recherche publique française et indienne, les partenaires sont d'accord pour qu'elles soient partagées librement dans l'entrepôt Dataverse de l'INRAE après la fin du projet de recherche et la publication des résultats (cf. détail §5).

#### **4C. COMMENT LES ÉVENTUELLES QUESTIONS ÉTHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DÉONTOLOGIQUES RESPECTÉS ?**

Le code de déontologie et d'éthique de l'INRAE (lien) sera le document de référence pour régler les éventuels problèmes d'éthique. En cas de problème déontologique non répertorié, le Comité de Déontologie et d'éthique de l'INRA sera saisi.

### **5. PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME**

#### **5A. COMMENT ET QUAND LES DONNÉES SERONT-ELLES PARTAGÉES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNÉES OU DES RAISONS DE DÉFINIR UN EMBARGO ?**

Les données seront accessibles, après un embargo de 24 mois après la fin du projet pour publications :

- Les données et leurs métadonnées détaillées (cf 2.a) seront déposées sur le Dataverse INRA.
- Les jeux de données seront publiés sous licence CC BY
- Elles intéresseront particulièrement les agronomes, agriculteurs, pédologues, aménageurs, développeurs,... travaillant dans les zone tropicales.

#### **5B. COMMENT LES DONNÉES À CONSERVER SERONT-ELLES SÉLECTIONNÉES ET OÙ SERONT-ELLES PRÉSERVÉES SUR LE LONG TERME (PAR EX. UN ENTREPÔT DE DONNÉES OU UNE ARCHIVE) ?**

Vu le faible volume des fichiers, tout sera conservé.

#### **5C. QUELLES MÉTHODES OU QUELS OUTILS LOGICIELS SERONT NÉCESSAIRES POUR ACCÉDER ET UTILISER LES DONNÉES ?**

Les données seront partagées sous des formats libres et ouverts, elles ne nécessitent pas de logiciels spécifiques pour y accéder ou les utiliser. Elles seront accessibles directement via l'entrepôt Dataverse INRA.

#### **5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PÉRENNE (COMME LE DOI) SERA-T-ELLE ASSURÉE POUR CHAQUE JEU DE DONNÉES ?**

Le Dataverse INRAE attribuera un DOI à chaque jeu de données déposé.

### **6. RESPONSABILITÉS ET RESSOURCES EN MATIÈRE DE GESTION DES DONNÉES**

#### **6A. QUI (PAR EXEMPLE RÔLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNÉES (C'EST-À-DIRE LE GESTIONNAIRE DES DONNÉES) ?**

Chaque work package est coordonné par un responsable scientifique (voir protocole) qui est en charge de s'assurer de la bonne conduite des activités de son work package. L'investigateur principal s'assurera de la bonne exécution globale du plan de gestion de données. Un organigramme détaillé des rôles et responsabilité sera développé en début de projet. Les ressources humaines, matérielles et logicielles requises pour la mise en œuvre du plan de gestion des données ont été prévues et incluses dans le budget du projet.

#### **6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUÉS) DÉDIÉES À LA GESTION DES DONNÉES PERMETTANT DE S'ASSURER QUE LES DONNÉES SERONT FAIR (FACILE À TROUVER, ACCESSIBLE, INTEROPÉRABLE, RÉUTILISABLE) ?**

**RH :**

- Recrutement Ingénieur de Recherche responsable des données en France : 48 mois (coût mensuel charges comprises à préciser)
- Recrutement 15 techniciens en CDD de 24 mois en Inde pour recueil, la saisie et le transfert des données de terrain sur les sols et leur utilisation, d'une part (données 1), la traduction et la prise de note, d'autre part (données 5) : 15 x 24 mois x coût mensuel charges comprises
- Formation des responsables sectoriels des données en France sur les logiciels choisis, et sur les protocoles de qualité (yc métadonnées), de sauvegarde et de transfert des données :

- Données 1 et 2 : M. X, IRD
- Données 3 et 4 : Mme Y, Inrae
- Données 5 : M. T, CNRS
- Frais de transport et de mission de ces responsables pour la formation de ces techniciens sur place à la saisie et la prise de note sur tablette et le transfert sécurisé via internet

**Logiciels :**

Professionnels, standards, et libres si possibles : SIG, BDD, texte mining, encodage des données avant transfert, ...(reste à préciser)

**Images satellites :**

Fournisseur, satellite, résolution, filtre, ...

**Matériel :**

- X PC boostés, X PC portables, X tablettes
- X Disques Durs externes
- Espace cloud sur serveur public français
- Baie de serveur

## **DONNÉES SATELLITE : DONNÉES SATELLITE SUR LES TYPES DE SOL ET L'UTILISATION DU SOL DU SUD DE L'INDE**

### **1. DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES**

#### **1A. COMMENT DE NOUVELLES DONNÉES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNÉES PRÉEXISTANTES SERONT-ELLES RÉUTILISÉES ?**

Ces images satellite seront à acquérir auprès de SPOT Image. Elles seront conservées dans la base d'image satellite de l'UMR GET, partenaire du projet. Elles seront traitées à la demande des différents membres du projet par M. X, ingénieur spécialiste en télédétection.

#### **1B. QUELLES DONNÉES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTÉES OU PRODUITES ?**

Ses images très volumineuses (x To) seront conservées au format .BMP. Elles nécessitent un hébergement et des logiciels spécifiques pour la lecture et le traitement. Elles sont de plus protégées par le contrat qui liera le projet à SPOT Image lors de l'acquisition.

### **2. DOCUMENTATION ET QUALITÉ DES DONNÉES**

#### **2A. QUELLES MÉTADONNÉES ET QUELLE DOCUMENTATION (PAR EXEMPLE MÉTHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNÉES) ACCOMPAGNERONT LES DONNÉES ?**

Les images seront indexées à l'aide des métadonnées géographiques standards [FGDC/CSDGM- Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata](#)

Les différents traitements appliqués sur les images initiales seront décrits dans le cahier de laboratoire de l'UMR GET par M. X, tout au long du projet. Les éventuels logiciels ou codes utilisés seront énumérés. Ils seront de préférences choisis libres, ouverts et gratuits

#### **2B. QUELLES MESURES DE CONTRÔLE DE LA QUALITÉ DES DONNÉES SERONT MISES EN ŒUVRE ?**

Toutes les étapes de la collecte, de la saisie, du contrôle, du traitement et de l'analyse des données seront décrites à l'aide de vocabulaires standards par les ingénieurs et chercheurs du projet dans les cahiers de Laboratoire repris dans les fichiers lisez-moi déposés dans l'entrepôt avec les jeux de données.

La politique de sauvegarde et de stockage décrite au §3 minimisera les risques de perte ou de corruption des fichiers.

L'ensemble de la recherche respectera la politique qualité de l'INRA (lien vers document), pour le respect de la traçabilité et de la fiabilité des résultats.

#### **Nommage et organisation des fichiers :**

- Les images sauvegardées seront nommés selon le principe suivant :  
ATCHA(1-5)\_type\_thème et/ou lieu\_version majeure\_annee-mois-jour.ext  
Exemple : ATCHA5\_IMG\_berambadi7\_v2\_2017-08-05.odt
- Ils seront classés dans des répertoires et sous répertoires aux noms explicites de 25 caractères maximum contenant au plus 25 fichiers.
- Les versions majeures (décision prise en réunion d'équipe) donneront lieu à la destruction de toutes les versions intermédiaires les séparant de la version majeure précédente. A défaut d'une telle décision, une version mensuelle seulement sera conservée le 1er jour ouvré du mois suivant.

### **3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE**

#### **3A. COMMENT LES DONNÉES ET LES MÉTADONNÉES SERONT-ELLES STOCKÉES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les données et images satellite seront hébergés et sauvegardés sur des serveurs de l'UMR GET.

Les données de ces serveurs sont sauvegardées automatiquement chaque nuit par l'hébergeur.

Les versions quotidiennes seront conservées 7 jours, les versions dominicales seront conservées 6 semaines, et les versions du 1er de chaque mois seront conservées jusqu'à la fin du projet, afin de permettre une éventuelle récupération.

L'ingénieur Données sera responsable du stockage et de la sauvegarde des données durant toute la durée du projet.

**3B. COMMENT LA SÉCURITÉ DES DONNÉES ET LA PROTECTION DES DONNÉES SENSIBLES SERONT-ELLES ASSURÉES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?**

Les procédures de sécurité informatique de l'INRA (lien) seront appliquées, en particulier pour la récupération des données en cas d'incident.

Durant le projet, les ressortissants des UMR (AGIR, EMMAH, GET SAS TSE-R) et organismes (IFP et Univ Rennes 2) partenaires du projet et les organisations indiennes partenaires des projets antérieurs (CEFIRSE, ORE BVET, Watershed Department du Karnataka) habilités auront accès images à la demande auprès de M X qui pourra également leur appliquer les traitements nécessaires à leur interprétation.

**4. EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE**

**4A. SI DES DONNÉES À CARACTÈRE PERSONNEL SONT TRAITÉES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LÉGISLATION SUR LES DONNÉES À CARACTÈRE PERSONNEL ET SUR LA SÉCURITÉ DES DONNÉES SERA-T-IL ASSURÉ ?**

Il n'y pas de traitement de données à caractère personnel dans ce jeu de données

**4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITÉ OU LES DROITS DE PROPRIÉTÉ INTELLECTUELLE SUR LES DONNÉES, SERONT-ELLES ABORDÉES ? QUELLE EST LA LÉGISLATION APPLICABLE EN LA MATIÈRE ?**

L'ensemble des partenaire est copropriétaire des données produites par le projet, et délègue leur gestion durant le projet à l'UMR GET. Une partie de l'accord de consortium signé par l'ensemble des organismes partenaires détaille les droits d'accès, d'utilisation, et de partage des données.

Les images satellite, achetées sous licence, ne pourront être partagées à d'autres personnes que les membres des institutions partenaires du projet (sur demande auprès de M. X).

**4C. COMMENT LES ÉVENTUELLES QUESTIONS ÉTHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DÉONTOLOGIQUES RESPECTÉS ?**

Le code de déontologie et d'éthique de l'INRA (lien) sera le document de référence pour régler les éventuels problèmes d'éthique. En cas de problème déontologique non répertorié, le Comité de Déontologie et d'éthique de l'INRA sera saisi.

**5. PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME**

**5A. COMMENT ET QUAND LES DONNÉES SERONT-ELLES PARTAGÉES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNÉES OU DES RAISONS DE DÉFINIR UN EMBARGO ?**

Les images satellite sont achetées sous licence et ne pourront donc être partagées.

**5B. COMMENT LES DONNÉES À CONSERVER SERONT-ELLES SÉLECTIONNÉES ET OÙ SERONT-ELLES PRÉSERVÉES SUR LE LONG TERME (PAR EX. UN ENTREPÔT DE DONNÉES OU UNE ARCHIVE) ?**

Les images satellite, achetées sous licence, ne seront pas conservées sur un autre serveur que celui de l'UMR GET.

**5C. QUELLES MÉTHODES OU QUELS OUTILS LOGICIELS SERONT NÉCESSAIRES POUR ACCÉDER ET UTILISER LES DONNÉES ?**

Les images satellites ne seront ni partagées ni conservées à l'issue du projet.

**5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PÉRENNE (COMME LE DOI) SERA-T-ELLE ASSURÉE POUR CHAQUE JEU DE DONNÉES ?**

Les images satellites ne seront ni partagées ni conservées à l'issue du projet.

## 6. RESPONSABILITÉS ET RESSOURCES EN MATIÈRE DE GESTION DES DONNÉES

6A. QUI (PAR EXEMPLE RÔLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNÉES (C'EST-À-DIRE LE GESTIONNAIRE DES DONNÉES) ?

Chaque work package est coordonné par un responsable scientifique (voir protocole) qui est en charge de s'assurer de la bonne conduite des activités de son work package. L'investigateur principal s'assurera de la bonne exécution globale du plan de gestion de données. Un organigramme détaillé des rôles et responsabilités sera développé en début de projet. Les ressources humaines, matérielles et logicielles requises pour la mise en œuvre du plan de gestion des données ont été prévues et incluses dans le budget du projet.

6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUÉS) DÉDIÉES À LA GESTION DES DONNÉES PERMETTANT DE S'ASSURER QUE LES DONNÉES SERONT FAIR (FACILE À TROUVER, ACCESSIBLE, INTEROPÉRABLE, RÉUTILISABLE) ?

**RH :**

- Recrutement Ingénieur de Recherche responsable des données en France : 48 mois (coût mensuel charges comprises à préciser)
- Recrutement 15 techniciens en CDD de 24 mois en Inde pour recueil, la saisie et le transfert des données de terrain sur les sols et leur utilisation, d'une part (données 1), la traduction et la prise de note, d'autre part (données 5) : 15 x 24 mois x coût mensuel charges comprises
- Formation des responsables sectoriels des données en France sur les logiciels choisis, et sur les protocoles de qualité (yc métadonnées), de sauvegarde et de transfert des données :
  - Données 1 et 2 : M. X, IRD
  - Données 3 et 4 : Mme Y, Inrae
  - Données 5 : M. T, CNRS
- Frais de transport et de mission de ces responsables pour la formation de ces techniciens sur place à la saisie et la prise de note sur tablette et le transfert sécurisé via internet

**Logiciels :**

Professionnels, standards, et libres si possibles : SIG, BDD, texte mining, encodage des données avant transfert, ...(reste à préciser)

**Images satellites :**

Fournisseur, satellite, résolution, filtre, ...

**Matériel :**

- X PC boostés, X PC portables, X tablettes
- X Disques Durs externes
- Espace cloud sur serveur public français
- Baie de serveur