



Vers l'interopérabilité des bases de données  
pour être plus FAIR

## Plan de Gestion des Données – version finale

2022

Julien Demenois, Fabrice Bénédet



N° ANR-19-DATA-0005-01



---

# DMP du projet "DATA4C+ - Interopérabilité des bases de données sur le carbone du sol"

*Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Horizon 2020 FAIR DMP (anglais)" fourni par Commission européenne.*

## Plan Details

|  |  |
|--|--|
| <b>Plan title</b>  | DMP du projet "DATA4C+ - Interopérabilité des bases de données sur le carbone du sol"  |
| <b>Version</b>   | Final version  |
| <b>Fields of science and technology (from OECD classification)</b> |  |
| <b>Language</b>  | fra  |
| <b>Creation date</b>   | 2019-12-13   |
| <b>Last modification date</b>                                      | 2020-03-25   |
| <b>Identifier</b>  | DATA4C+  |
| <b>Identifier type</b>   | URL  |
| <b>License</b>   | Creative Commons Attribution 4.0 International   |
| <b>Management plans related to the project</b>                     | <ul style="list-style-type: none"><li>Version initiale du PGD - Projet ANR DATA4C+ - Interopérabilité des bases de données sur le carbone du sol :<br/><a href="https://agritrop.cirad.fr/596385/">https://agritrop.cirad.fr/596385/</a></li></ul> |

## Project Details

**Project title** DATA4C+ - Interopérabilité des bases de données sur le carbone du sol

**Abstract**

DATA4C+ a comme objectifs de :

- Définir des règles de bonnes pratiques pour décrire les données contenues dans les bases de données sur le carbone du sol et les renseigner de manière homogène
- Analyser les verrous informatiques et juridiques à l'interopérabilité entre bases de données sur le carbone du sol et les modes de gestion des sols afin de proposer des solutions pour les lever
- Expérimenter la mise en œuvre des solutions méthodologiques formulées par le projet dans un département d'outre-mer français ; Echanger et diffuser à l'échelle internationale sur le projet.

DATA4C+ se focalise notamment sur la structuration des bases de données sur le carbone du sol, leur ergonomie pour faciliter les recherches d'informations, la gestion des accès ou encore le format des données pour les rendre potentiellement échangeables et réutilisables dans la durée.

**Funding**

- ANR : N° ANR-19-DATA-0005-01

**Start date** 2020-10-01

**End date** 2022-03-31

**Partners**

- Institut National de la Recherche Agronomique ()
- Institut de recherche pour le développement ()

#### Produits de recherche :

1. DATA4C+ - A thesaurus to define land management practices in agriculture and forestry for soil carbon storage (Dataset)
2. Documents de projet (Text)
3. Prototype DATA4C+ (Software)
4. Bases de données utilisées pour le prototype DATA4C+ (Dataset)

#### Contributeurs

| Nom             | Affiliation | Rôles   |
|-----------------|-------------|---|
| Julien DEMENOIS | CIRAD       | <ul style="list-style-type: none"><li>• DMP manager</li><li>• Personne contact pour les données (Thesaurus, Documents de projet, Prototype DATA4C+, Bases de données)</li><li>• Project coordinator</li></ul> |

# DMP du projet "DATA4C+ - Interopérabilité des bases de données sur le carbone du sol"

## 1. Data summary

### Documents de projet

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

Par documents de projet, on entend :

- la proposition initiale acceptée par l'ANR pour financement (a)
- la convention de financement de l'ANR et accord de consortium (b)
- les compte-rendus de comité de pilotage (c)
- les compte-rendus d'ateliers (d)
- le rapport d'avancement intermédiaire (e)
- le rapport final de projet (f)
- les présentations du projet (g)
- le Plan de Gestion des Données (h)
- les livrables et les articles scientifiques (i)

- State the purpose of the data collection/generation

La production de ces documents de projet répond aux nécessités de coordination interne des activités du projet et à la communication externe.

- Explain the relation to the objectives of the project

DATA4C+ a 4 objectifs spécifiques :

1. (OS1) Définir des règles de bonnes pratiques pour décrire les données contenues dans les bases de données sur le carbone du sol et les renseigner de manière homogène
2. (OS2) Analyser les verrous informatiques et juridiques à l'interopérabilité entre bases de données sur le carbone du sol et les modes de gestion des sols afin de proposer des solutions pour les lever
3. (OS3) Expérimenter la mise en oeuvre des solutions méthodologiques formulées par le projet dans un département d'outre-mer français
4. (OS4) Echanger et diffuser à l'échelle internationale sur le projet.

Les documents de projet répondent prioritairement à l'OS4.

- Specify the types and formats of data generated/collected

Les documents de projets générées dans DATA4C+ consistent en des documents au format texte.

- Specify if existing data is being re-used (if any)

Non applicable

- Specify the origin of the data

Les documents de projet ont pour origine la mise en oeuvre du projet.

- State the expected size of the data (if known)

Quelques centaines de Mo.

- Outline the data utility: to whom will it be useful

Les documents a à f sont destinés aux partenaires du projet (Cirad, INRAE, IRD), aux personnes issues de ces partenaires impliquées dans le projet et à l'ANR.

Les documents g, h et i sont destinés à la communauté des chercheurs sur le carbone du sol.

## DATA4C+ - A thesaurus to define land management practices in agriculture and forestry for soil carbon storage

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

Par thésaurus, on entend la classification et la définition de vocabulaire.

- State the purpose of the data collection/generation

Le thésaurus se concentre sur les pratiques de gestion des terres identifiées dans la littérature scientifique comme facteurs de changement du carbone organique du sol (SOC). Le but du thésaurus est de combler le manque d'un thésaurus complet pour les pratiques de gestion des terres qui affectent directement ou indirectement la dynamique du SOC.

- Explain the relation to the objectives of the project

DATA4C+ a 4 objectifs spécifiques :

1. (OS1) Définir des règles de bonnes pratiques pour décrire les données contenues dans les bases de données sur le carbone du sol et les renseigner de manière homogène
2. (OS2) Analyser les verrous informatiques et juridiques à l'interopérabilité entre bases de données sur le carbone du sol et les modes de gestion des sols afin de proposer des solutions pour les lever
3. (OS3) Expérimenter la mise en oeuvre des solutions méthodologiques formulées par le projet dans un département d'outre-mer français
4. (OS4) Echanger et diffuser à l'échelle internationale sur le projet.

Le thésaurus est le résultat de l'OS1. Il répond également à l'OS4.

- Specify the types and formats of data generated/collected

Le thésaurus est disponible dans un format texte ouvert CSV (Comma-separated values) et représenté selon un modèle de données SKOS (Simple Knowledge Organization System).

- Specify if existing data is being re-used (if any)

Le thésaurus mobilise des définitions disponibles dans de nombreuses sources bibliographiques. Celles-ci sont mentionnées dans le thésaurus.

- Specify the origin of the data

Le thésaurus est l'un des résultats des activités du projet.

- State the expected size of the data (if known)

Quelques centaines de Ko.

- Outline the data utility: to whom will it be useful

Le thésaurus est orienté pour être utilisé par les scientifiques en agronomie, foresterie et sciences du sol dans le but d'uniformiser la description des pratiques influençant le SOC dans leurs recherches originales. Le thésaurus peut également être utilisé par les gestionnaires de terres ou par les laboratoires pour décrire les échantillons de sol analysés.

## Prototype DATA4C+

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

Par prototype DATA4C+, on entend l'outil informatique développé dans le projet qui permet de rendre interopérables les bases de données sur le carbone du sol en Guyane.

- State the purpose of the data collection/generation

La production du prototype vise à démontrer la faisabilité de l'interopérabilité entre bases de données sur le carbone du sol des partenaires du projet (Cirad, INRAE, IRD).

- Explain the relation to the objectives of the project

DATA4C+ a 4 objectifs spécifiques :

1. (OS1) Définir des règles de bonnes pratiques pour décrire les données contenues dans les bases de données sur le carbone du sol et les renseigner de manière homogène
2. (OS2) Analyser les verrous informatiques et juridiques à l'interopérabilité entre bases de données sur le carbone du sol et les modes de gestion des sols afin de proposer des solutions pour les lever
3. (OS3) Expérimenter la mise en œuvre des solutions méthodologiques formulées par le projet dans un département d'outre-mer français
4. (OS4) Echanger et diffuser à l'échelle internationale sur le projet.

Le prototype répond à l'OS3.

- Specify the types and formats of data generated/collected

Le prototype est développé à partir d'outils libres : le système de gestion de base de données PostgreSQL et son extension PostGIS, Lizmap, Symphony, Leaflet, API SensorThings.

- Specify if existing data is being re-used (if any)

Non applicable

- Specify the origin of the data

Le prototype a pour origine la mise en œuvre du projet.

- State the expected size of the data (if known)

Quelques centaines de Mo.

- Outline the data utility: to whom will it be useful

Le prototype est destiné aux chercheurs des partenaires du projet (Cirad, INRAE, IRD) et à tout partenaire scientifique souhaitant développer des solutions d'interoperabilité entre bases de données sur les sols.

## Bases de données utilisées pour le prototype DATA4C+

**Provide a summary of the data addressing the following issues:**

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

Le prototype DATA4C+ développé dans le projet est basé sur des jeux de données existants issues (i) des bases de données sur le carbone du sol des partenaires du projet (Cirad, INRAE, IRD) et (ii) des ressources librement accessibles en ligne.

| <b>Acronyme/Dénomination</b> | <b>Description</b>   | <b>Institut</b>   |
|------------------------------|--|---|
| DoneSol                      | Bases de données contenant les données des programmes IGCS (Inventaire, Gestion et Conservation des Sols) et RMQS (Réseau de Mesure de la Qualité des Sols)  | INRAE   |
| BD DATA4C - Guyane           | Base de données provenant du laboratoire d'analyse des sols du Cirad (US Analyse) (LIMS - Laboratory Information Management System)  | Cirad   |
| ValSol - Guyane              | Base de données regroupant des données ponctuelles et surfaciques issues de programme de cartographie des sols de l'ORSTOM puis IRD, ainsi que de projets de recherche   | IRD   |
| Parcelles ONF                | Ressource cartographique présentant le parcellaire forestier du Domaine Forestier Permanent de Guyane  | ONF   |
| Occupation du sol            | Ressource cartographique présentant un état des lieux des modes d'occupation du sol sur la bande littorale guyanaise en 2015 par l'analyse d'images satellite, de photographies aériennes.                         | ONF   |
| Suivi déforestation          | Ressource cartographique présentant les données de suivi des déforestations par imagerie RADAR (Sentinel 1, 10m de résolution). Elle permet d'observer les déforestations sur toute la Guyane durant l'année 2020. | ONF   |
| Atlas des paysages           | Ressource cartographique présentant le découpage en unités, sous-unités et séquences paysagères de l'ensemble du territoire guyanais.  | AUDeG - Agence d'Urbanisme et de Développement de la Guyane |
| GSOCMap                      | La carte mondiale du carbone du sol se compose de cartes nationales du carbone organique du sol, élaborées sous forme de grilles pédologiques de 1 km, couvrant une profondeur de 0 à 30 cm.                       | FAO   |
| Soil Grid deep (0-5 cm)      | Stock de carbone organique du sol en tonnes par ha pour l'intervalle de profondeur 0-5 cm prédits à l'aide de la compilation mondiale des observations du sol au sol.  | ISRIC   |

## 2. FAIR data

### Documents de projet

#### 2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Par documents de projet, on entend :

- la proposition initiale acceptée par l'ANR pour financement (a)
- la convention de financement de l'ANR et accord de consortium (b)
- les compte-rendus de comité de pilotage (c)
- les compte-rendus d'ateliers (d)
- le rapport d'avancement intermédiaire (e)
- le rapport final de projet (f)
- les présentations du projet (g)
- le Plan de Gestion des Données (h)
- les livrables et les articles scientifiques (i)
- Outline the discoverability of data (metadata provision)

Les documents a, b, c, d, e, f sont accessibles aux personnes impliquées dans le projet via un nom d'utilisateur et un mot de passe sur un entrepôt privé du Cirad ([Alfresco](#)).

Les documents g, h et i sont accessibles sur le [site web du projet](#).

Le document h est accessible dans les archives ouvertes du Cirad ([Agritrop](#)), dans sa [version initiale](#) et dans sa [version finale](#).

Les documents i seront accessibles dans Agritrop dès publication. Les articles scientifiques seront référencés et disponibles en open access sur les sites des revues : [Etude et Gestion des Sols](#) (article soumis) et [SOIL](#) (article soumis).

- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**

Les documents déposés dans les archives ouvertes du Cirad ([Agritrop](#)) sont associés à une URL unique.

Les documents seront associés à un DOI dès publication.

Les autres documents ne font pas l'objet d'une identification unique. Le nom des fichiers et leur description permettent en revanche leur identification.

- **Outline naming conventions used**

Pour les documents ne faisant pas l'objet d'une identification unique, la règle de nommage est : objet du document \_année \_mois \_jour.

- **Outline the approach towards search keyword**

Les mots-clés utilisés s'appuient sur ceux des archives ouvertes du Cirad ([Agritrop](#)).

- **Outline the approach for clear versioning**

Les éventuelles différentes versions des documents sont gérées à travers la règle de nommage incluant la date du document.

- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

Voir la section "Outline naming conventions used"

## 2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**

Les documents g, h et i sont/seront accessibles. Les documents versés dans les archives ouvertes du Cirad (Agritrop) sont accessibles sous licence Creative Commons CC-BY-NC-ND.

Les documents i ont été soumis dans des revues en Open Access.

Les documents internes au projet (convention de financement, compte-rendus, ...) constituent des éléments contractuels entre les partenaires. A ce titre, ils ne sont pas accessibles de manière ouverte.

- **Specify how the data will be made available**

Les documents versés dans les archives ouvertes du Cirad (Agritrop) sont accessibles sous licence Creative Commons CC-BY-NC-ND.

Les documents i ont été soumis dans des revues en Open Access.

- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**

L'accès à l'ensemble des documents de projet nécessite un lecteur de fichiers au format PDF.

- **Specify where the data and associated metadata, documentation and code are deposited**

La description des documents de projet est disponible :

- dans l'entrepot privé du Cirad ([Alfresco](#)) pour les documents a, b, c, d, e, f
- dans les archives ouvertes du Cirad (Agritrop) pour les documents g, h et i

- **Specify how access will be provided in case there are any restrictions**

Non applicable

---

### **2.3 Making data interoperable:**

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**

L'interopérabilité des documents de projet est assurée par l'utilisation du format PDF, la description des documents par des métadonnées et les règles de nommage des documents.

- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

Pour les documents versés dans les archives ouvertes du Cirad (Agritrop), le vocabulaire utilisé pour les métadonnées est celui d'Agritrop.

Pour les documents i, le vocabulaire utilisé pour les métadonnées est celui des revues scientifiques.

---

### **2.4 Increase data re-use (through clarifying licenses):**

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

- **Specify how the data will be licenced to permit the widest reuse possible**

Les documents versés dans les archives ouvertes du Cirad (Agritrop) sont accessibles sous licence Creative Commons CC-BY-NC-ND.

Les documents i ont été soumis dans des revues scientifiques en Open Access. Les articles de la revue SOIL sont sous licence Creative Commons CC-BY-4.0.

Les autres documents du projet ne sont pas accessibles de manière ouverte. Leur ré-utilisation doit se faire conformément à l'accord de consortium conclut entre les partenaires du projet.

- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**

Les données produites par DATA4C+ sont disponibles au fur et à mesure de leur production ou acceptation pour publication par les revues scientifiques.

- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**

En dehors des documents de projet disponibles sous licence Creative Commons, la ré-utilisation des autres documents du projet doit se faire conformément à l'accord de consortium conclut entre les partenaires du projet.

- **Describe data quality assurance processes**

Les documents a à h ont été produits après contrôle qualité par les responsables des OS et le coordinateur du projet (Julien Demenois).

Les documents i ont été produits après contrôle qualité des co-auteurs, des relecteurs désignés par les revues scientifiques et de son comité éditorial.

- **Specify the length of time for which the data will remain re-usable**

Pas de durée spécifique.

## **DATA4C+ - A thesaurus to define land management practices in agriculture and forestry for soil carbon storage**

### **2.1 Making data findable, including provisions for metadata:**

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

- Outline the discoverability of data (metadata provision)

Le thésaurus est déposé dans :

- le [dataverse du Cirad](#)
- [Agroportal](#)

Il est aussi accessible sur Internet ici : <http://data4c-plus.net/admin/thesaurus/index>

- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

Le thésaurus est identifié avec le DOI : [10.18167/DVN1/HMCPMF](https://doi.org/10.18167/DVN1/HMCPMF)

- Outline naming conventions used

Le fichier CSV est nommé de la manière suivante : Thesaurus\_DATA4C+\_année-mois-jour

Le fichier SKOS est nommé de la manière suivante : D4c\_SKOS\_numéro de version conformément à Agroportal

- Outline the approach towards search keyword

Les mots-clés utilisés s'appuient sur ceux des archives ouvertes du Cirad ([Agritrop](#)).

- Outline the approach for clear versioning

Les éventuelles différentes versions des documents sont gérées à travers les règles de nommage : voir la section "Outline naming conventions used"

- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Les standards des métadonnées suivent ceux d'Agroportal et du dataverse du Cirad.

---

### **2.2 Making data openly accessible:**

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Le thésaurus est accessible de manière ouverte sous licence Creative Commons CC-BY-4.0.

- Specify how the data will be made available

Le thésaurus est déposé dans :

- le [dataverse du Cirad](#)
- [Agroportal](#)

Il est aussi accessible sur Internet ici : <http://data4c-plus.net/admin/thesaurus/index>

Il est également décrit dans un article scientifique soumis à la revue SOIL (open access).

- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**

Le thésaurus nécessite un lecteur de fichiers CSV ou SKOS ou HTML.

- **Specify where the data and associated metadata, documentation and code are deposited**

La description du thésaurus est disponible :

- le [dataverse du Cirad](#)
- [Agroportal](#)

L'article scientifique soumis à la revue SOIL vient compléter cette description.

- **Specify how access will be provided in case there are any restrictions**

Non applicable

---

### 2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**
  
- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**

L'interopérabilité du thésaurus est assurée par l'utilisation du format SKOS et la description des documents par des métadonnées.

- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

Le thésaurus est partiellement aligné avec AGROVOC. Il mobilise également des définitions disponibles dans de nombreuses sources bibliographiques. Celles-ci sont mentionnées dans le thésaurus.

---

### 2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**
  
- **Specify how the data will be licenced to permit the widest reuse possible**

Le thésaurus est sous licence Creative Commons CC-BY-4.0.

- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**

Le thésaurus est disponible sur le dataverse du Cirad et Agroportal depuis février 2022.

- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Le thésaurus est sous licence Creative Commons CC-BY-4.0. Il est ré-utilisable par une tierce partie selon les termes de cette licence.

- **Describe data quality assurance processes**

Le thésaurus a fait l'objet d'un contrôle qualité par le responsable de l'OS1 et le coordinateur du projet. Lors de son dépôt dans Agroportal, il a fait l'objet d'un contrôle qualité externe par les gestionnaires d'Agroportal. Lors de son dépôt dans le dataverse du Cirad il a fait l'objet d'un contrôle qualité externe par les gestionnaires du dataverse du Cirad.

Le thésaurus fait enfin l'objet d'une revue externe dans le cadre du processus de soumission à la revue scientifique SOIL.

- **Specify the length of time for which the data will remain re-usable**

Pas de durée spécifique.

## Prototype DATA4C+

### 2.1 Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

- Outline the discoverability of data (metadata provision)

Le prototype est déposé dans le [GitLab du Cirad](#)

Il est aussi accessible sur Internet ici : <http://data4c-plus.net/>. Le site Internet contient une description des fonctionnalités du prototype et permet un accès réservé aux partenaires du projet (via nom d'utilisateur et mot de passe) au prototype développé.

- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

Le prototype est identifié via la diffusion du modèle de données et du dictionnaire des données

- Outline naming conventions used

Non applicable

- Outline the approach towards search keyword

Non applicable

- Outline the approach for clear versioning

Le versionning est géré par l'outil GitLab du Cirad qui permet le développement collaboratif

- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Non applicable

### 2.2 Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**

Le prototype est accessible uniquement via un identifiant et un mot de passe. L'accès est réservé aux personnes identifiées sur le site. L'accès est réservé aux partenaires du projet.

Les codes du prototype sont disponibles en open access.

- **Specify how the data will be made available**

Le prototype est accessible sur Internet à l'adresse suivante : <http://data4c-plus.net/>

Les codes du prototype sont disponibles sur le [GitLab du Cirad](#)

- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**

Le prototype est développé à partir d'outils libres : le système de gestion de base de données PostgreSQL et son extension PostGIS, Lizmap, Symfony, Leaflet, API SensorThings.

La documentation du prototype est disponible sur le [GitLab du Cirad](#)

- **Specify where the data and associated metadata, documentation and code are deposited**

Les codes du prototype et sa documentation sont disponibles sur le [GitLab du Cirad](#)

- **Specify how access will be provided in case there are any restrictions**

Le prototype n'est pas accessible, en dehors des participants au projet.

Les codes du prototype sont accessibles en open access sous licence CC-BY-4.0.

---

### 2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow interdisciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**

Le prototype suit les standards de l'Open Geospatial Consortium (OGC) d'échange de données par interopérabilité. Le développement de services Web permettent de connecter des bases de données spatiales distantes au serveur QGis Serveur utilisé dans Data4C+

- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow interdisciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

Non applicable

---

### 2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

- **Specify how the data will be licenced to permit the widest reuse possible**

Les codes du prototype et sa documentation sont sous licence Creative Commons CC-BY-4.0.

- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**

Les codes du prototype et sa documentation sont d'ores et déjà disponibles sur le [GitLab du Cirad](#)

- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**

Les codes du prototype et sa documentation sont sous licence Creative Commons CC-BY-4.0. Ils sont ré-utilisables par une tierce partie selon les termes de cette licence.

- **Describe data quality assurance processes**

Les codes du prototype et sa documentation ont fait l'objet d'un contrôle qualité par le responsable informatique de l'OS2 et le coordinateur du projet

- **Specify the length of time for which the data will remain re-usable**

Pas de durée spécifique.

## Bases de données utilisées pour le prototype DATA4C+

### 2.1 Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

Les jeux de données issues des bases de données des partenaires sont connus et documentés par les équipes qui les maintiennent. Ils suivent des standards de métadonnées propres à chaque organisme et les règles de bonnes pratiques de l'OS1 pour décrire les données permettent d'extraire de façon harmonisée et standardisée les variables d'intérêt.

La cartographie des correspondances entre variables d'intérêt dans les bases de données Cirad, INRAE et IRD permet de documenter la méthodologie d'élaboration de la base de données.

Les ressources de données externes sont documentés par chaque fournisseur et suivent les normes de métadonnées utilisées par la communauté.

### 2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

Les jeux de données issues des bases de données des partenaires restent privées et accessibles uniquement aux membres du projets au travers du prototype mis en oeuvre.

Les ressources de données externes sont toutes disponibles sur Internet (les liens sont fournis dans le tableau ci-dessous). Les services Web ont été principalement utilisés pour y accéder et les manipuler au sein du prototype.

| Dénomination Lien   | Licence   |
|---|---|
| Parcelles ONF <a href="#">Suivre le lien...</a>   | Utilisation libre sous réserve de mentionner la source (NOM DU SERVICE) et la date de sa dernière mise à jour (ANNEE DE MAJ)  |
| Occupation du sol <a href="#">Suivre le lien...</a>   | Utilisation libre sous réserve de mentionner la source (a minima le nom du producteur) et la date de sa dernière mise à jour  |
| Suivi déforestation <a href="#">Suivre le lien...</a>   | Utilisation libre sous réserve de mentionner la source (ONF)  |
| Atlas des paysages <a href="#">Suivre le lien...</a>  | Utilisation libre sous réserve de mentionner la source (a minima le nom du producteur) et la date de sa dernière mise à jour. |
| GSOCMap <a href="http://54.229.242.119/geoserver/GSOC/wms?">http://54.229.242.119/geoserver/GSOC/wms?</a>   | Creative commons CC BY 4.0 license  |
| Soil grid deep 0-5 cm <a href="https://data.isric.org/geonetwork/srv/api/records/98062ae9-911d-4e04-80a9-e4b480f87799">https://data.isric.org/geonetwork/srv/api/records/98062ae9-911d-4e04-80a9-e4b480f87799</a> |   |

### 2.3 Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Les variables des jeux de données issues des bases de données des partenaires utilisées dans le prototype ont été unifiées et interconnectées en utilisant les règles de correspondance définies dans l'OS1 et le cadre de développement ouvert "OGC SensorThings API" (norme de l'Open Geospatial Consortium).

### 2.4 Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

Les jeux de données issues des bases de données des partenaires ne sont ré-utilisables que par les partenaires du projet au travers du prototype mis en oeuvre.

Les jeux de données accessibles librement sur Internet peuvent être réutilisés sans restriction particulière. Les licences et les conditions d'accès sont spécifiques à chaque produit de données et indiquées sur les sites d'accès.

## 3. Allocation of resources

### Documents de projet

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project

- **Describe costs and potential value of long term preservation**

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**

Non estimés

- **Clearly identify responsibilities for data management in your project**

La responsabilité de la gestion des données est faite conformément à l'accord de consortium conclut entre les partenaires du projet.

- **Describe costs and potential value of long term preservation**

Non estimés

## **DATA4C+ - A thesaurus to define land management practices in agriculture and forestry for soil carbon storage**

**Explain the allocation of resources, addressing the following issues:**

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**

Non estimés

- **Clearly identify responsibilities for data management in your project**

La responsabilité de la gestion des données est faite conformément à l'accord de consortium conclut entre les partenaires du projet.

- **Describe costs and potential value of long term preservation**

Non estimés

## **Prototype DATA4C+**

**Explain the allocation of resources, addressing the following issues:**

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**

Le coût du développement du prototype est de 40 k€, hors temps de chercheurs des partenaires du projet.

- **Clearly identify responsibilities for data management in your project**

La responsabilité de la gestion des données est faite conformément à l'accord de consortium conclut entre les partenaires du projet.

- **Describe costs and potential value of long term preservation**

Pas de coût associé au dépôt et à son archivage dans le Gitlab du Cirad.

## Bases de données utilisées pour le prototype DATA4C+

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

La cartographie des correspondances entre bases de données Cirad, INRAE et IRD a coûté environ 15 k€ correspondant à du temps de CDD, hors temps de personnels permanents des partenaires du projet.

- Clearly identify responsibilities for data management in your project

La responsabilité de la gestion de la base de données est faite conformément à l'accord de consortium conclut entre les partenaires du projet.

- Describe costs and potential value of long term preservation

Pas de coût associé au maintien dans le temps.

## 4. Data security

### Documents de projet

Address data recovery as well as secure storage and transfer of sensitive data

Les documents a, b, c, d, e, f sont stockées dans l'entrepôt privé du Cirad ([Alfresco](#)). Ils sont accessibles aux personnes impliquées dans le projet via un nom d'utilisateur et un mot de passe.

Le [site web du projet](#) sera archivé pour garantir son accès dans la durée.

Le document h est accessible dans les archives ouvertes du Cirad ([Agritrop](#))

Les documents i seront accessibles dans Agritrop dès publication et sur les sites des revues : [Etude et Gestion des Sols](#) et [SOIL](#).

### DATA4C+ - A thesaurus to define land management practices in agriculture and forestry for soil carbon storage

Address data recovery as well as secure storage and transfer of sensitive data

Le thésaurus est stocké dans le dataverse du Cirad, dans Agroportal et dans l'entrepôt privé du Cirad (Alfresco) (<https://collaboratif.cirad.fr/share/page/site/ANRDATA4C/dashboard>).

### Prototype DATA4C+

#### **Address data recovery as well as secure storage and transfer of sensitive data**

Le prototype est hébergé sur un serveur du Cirad et chez un prestataire externe ([OVHcloud](#)) . Les codes du prototype sont versés dans la forge logiciel du Cirad ([GitLab](#)).

#### **Bases de données utilisées pour le prototype DATA4C+**

#### **Address data recovery as well as secure storage and transfer of sensitive data**

Les bases de données Cirad, INRAE et IRD sont gérées par chacun des organismes. Les modalités de stockage et de sécurité sont leurs responsabilités respectives.

## **5. Ethical aspects**

**To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former**

Non applicable

---

## **6. Other**

**Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)**

Non applicable