

ECOLE NATIONALE SUPERIEURE des SCIENCES AGRONOMIQUES de  
BORDEAUX AQUITAINE  
Bordeaux Sciences Agro - 1, cours du Général de Gaulle - CS 40201 – 33175 GRADIGNAN Cedex

**Mémoire de stage de 2<sup>ème</sup> année**

# Utilisation d'outils numériques pour la valorisation de données pédologiques

**Paul VILVANDRE**

---

Pré-spécialisation : NumAg Numérique pour l'Agriculture

Etude réalisée à Bordeaux Sciences Agro

**Tutrice de stage : Stéphanie Jalabert, équipe Sols, Bordeaux Sciences Agro**

Mai 2023 à Août 2023

## Résumé

Le projet Solid'R, inscrit dans le programme IGCS du Gis Sol, vise à harmoniser les métadonnées des Référentiels Régionaux Pédologiques (RRP) et à développer des outils numériques pour la valorisation de ces bases de données. L'objectif du stage était de concevoir un script Python pour automatiser le calcul du Réservoir en eau Utilisable Maximal des sols (RUM) à partir du RRP du Lot-et-Garonne. Le script est structuré en 7 étapes, de l'importation des bibliothèques de programmation nécessaires à la cartographie des résultats, en passant par la correction de données. La durée d'exécution d'un tel script est de cinq minutes, contre plusieurs jours pour réaliser le travail manuellement. Des défis subsistent, notamment sur le format des données d'entrée, et la finalisation du script. Cet outil vise à améliorer l'usage des données sols pour répondre aux enjeux de gestion durable des sols au sein des territoires.

**Mots clés :** métadonnées, sols, outils numériques, script, Python, cartographie, réservoir en eau utilisable

## Abstract

The Solid'R project, part of the IGCS program of Gis Sol, aims to harmonize the metadata of the "Référentiels Régionaux Pédologiques" (RRP) and develop digital tools for the enhancement of these databases. The objective of the internship was to design a Python script to automate the calculation of the Maximum available water capacity (RUM) from the Lot-et-Garonne RRP. The script is structured into 7 steps, from importing the necessary programming libraries to mapping the results, including data correction. The execution time of such a script is five minutes, as opposed to several days of manual work. Challenges remain, particularly regarding the input data format and script finalization. This tool aims to improve the use of soil data to address the challenges of sustainable soil management within territories.

**Keywords:** metadata, soils, digital tools, script, Python, cartography, available water capacity

## Sommaire

Résumé .....	2
Abstract .....	2
Sommaire .....	3
Liste des abréviations .....	4
Liste des figures .....	4
Liste des tableaux .....	5
Liste des équations .....	5
Liste des annexes.....	5
Introduction .....	6
1. Contextualisation du stage .....	7
1.1. Le Gis Sol, entité pilote des programmes de cartographie en France .....	7
1.2. Le programme de cartographie multi-échelles Inventaire, Gestion et Conservation des Sols (IGCS).....	8
1.3. Etat des lieux de la diffusion des Référentiels Régionaux Pédologiques (RRP).....	9
1.4. Le projet Solid'R pour soutenir la diffusion des RRP .....	11
1.5. Les objectifs du stage .....	12
2. Développement de l'outil numérique de calcul d'indicateurs de propriétés des sols : cas du réservoir utilisable maximal en eau des sols .....	13
2.1. Estimation du Réservoir Utilisable Maximum en eau des sols .....	13
2.2. L'outil numérique : script en langage Python sous environnement Google Colab ...	19
2.3. Limites et perspectives .....	24
3. Mise en forme cartographique des résultats.....	25
3.1. Le script pour la représentation cartographique .....	25
3.2. Limites et perspectives .....	30
Conclusion.....	33
Bibliographie.....	34
Annexes.....	37

## Liste des abréviations

ADEME : Agence de la transition écologique  
 BDAT : Banque de Données des Analyses de Terre  
 BDETM : Base de Données des Eléments Traces Métalliques dans les sols  
 CPF : Connaissance Pédologique de la France  
 CPT : Classe de pédotransfert  
 EG : Éléments grossiers  
 FPT : Fonction de pédotransfert  
 Gis Sol : Groupement d'Intérêt Scientifique sur les Sols  
 HUM : Humidité utile maximale  
 IDG : Infrastructure de données géographiques (Plateforme d'informations géographiques)  
 IGCS : Inventaire, Gestion et Conservation des Sols  
 IGN : Institut national de l'information géographique et forestière  
 Inrae (ex-Inra) : Institut national de recherche pour l'agriculture, l'alimentation et l'environnement

OAD : Outil d'aide à la décision  
 OEAE : Outils d'Evaluation Agro-Environnementale  
 RMQS : Réseau de Mesures de la Qualité des Sols  
 RMT : Réseau mixte technologique  
 RMT S&T : Réseau mixte technologique Sols et Territoires  
 RRP : Référentiel Régional Pédologique  
 RU : Réservoir en eau utilisable des sols  
 RUM : Réservoir en eau utilisable maximal des sols  
 SIGAT : Master Géomatique, parcours Système d'information géographique et analyse des territoires, de l'Université Rennes 2  
 Solid'R : Soutien à la diffusion des Référentiels Régionaux Pédologiques  
 SR : Secteurs de référence  
 TF : Terres fines  
 UMR SAS : Unité mixte de recherche Sol, Agro et hydrosystème, Spatialisation

## Liste des figures

Figure 1 - Réseau de partenaires régionaux IGCS (Bispo, et al., 2020) .....	8
Figure 2 - Etat des lieux de la couverture cartographique des sols en France métropolitaine et Corse (Laroche, Richer de Forges, & Jalabert, À la découverte des sols : panorama des données, méthodes et outils existants, 2023) .....	9
Figure 3 - Encadré : RMT Sols et Territoires (RMT Sols & Territoires, 2023) .....	10
Figure 4 - Etapes du script et délimitation par parties du plan.....	12
Figure 5 - Notion du réservoir utilisable en eau des sols (Doussan, 2017).....	13
Figure 6 - Découpage des unités de sol en UCS, UTS et strates dans la BDD Donesol (Laroche, Thiry, Chenu, Moulin, & Bardy, 2014).....	17
Figure 7 - Liens entre les différentes tables de DoneSol qui caractérisent les données surfaciques (Lemercier, et al., 2017) .....	17
Figure 8 - Les étapes du script Python du calcul du RUM des sols du RRP47 avec sorties cartographiques .....	20
Figure 9 - Extrait du DataFrame après l'import et les ajouts des champs de STRATE_QUAL et STRATE_QUANT (Equipe Sols de Bordeaux Sciences Agro) .....	22
Figure 10 - Exemple de résultats du traitement de données de RU par UTS, à spatialiser par UCS (Lemercier, et al., 2017) .....	26
Figure 11 - Agrégation et spatialisation par UCS des valeurs des UTS selon la méthode de la classe majoritaire (Lemercier, et al., 2017) .....	27
Figure 12 - Agrégation et spatialisation par UCS des valeurs de RU des UTS selon la méthode de représentation du pourcentage de chaque classe de RU (Lemercier, et al., 2017) .....	27

Figure 13 - Seuils des classes et légendes utilisés pour la représentation cartographique du RUM (Script, 2023).....	28
Figure 14- Rendu final de la carte obtenue avec le script(Script de valorisation cartographique - Version du 19/09/2023) .....	30

### Liste des tableaux

Tableau 1 - Les programmes du GIS Sol (Gis Sol, 2023).....	7
Tableau 2 - Champs retenus pour chaque table importée et utilité de ces champs dans le calcul du RUM.....	18
Tableau 3 - Bibliothèques Python utilisées pour tout le script.....	20
Tableau 4 - Répartition des UCS dans chaque classe pour le RRP 47 (Equipe Sols de Bordeaux Sciences Agro) .....	28
Tableau 5 - Bibliothèques Python importées dans la seconde partie du script .....	29
Tableau 6 - Temps d'exécution des différentes étapes du script pour le RRP 47 .....	31

### Liste des équations

Équation 1- Le réservoir en eau utilisable maximal (RUM) d'un sol : $RUM_{SOL}$ .....	14
Équation 2 - Le réservoir en eau utilisable maximal d'un horizon i de sol : $RUM_i$ .....	14
Équation 3- RUM dans les terres fines.....	14
Équation 4 - RUM dans les éléments grossiers .....	15
Équation 5- HUM dans les éléments grossiers avec (Tetegan, 2011).....	15
Équation 6 - Formule générale du RUM d'un horizon i.....	15

### Liste des annexes

Annexe 1 - Recensement de fiches de métadonnées liées aux RRP recensées sur Internet (d'après le travail d'inventaire de Jalabert & Le Bas, 2022) .....	37
Annexe 2 - Lien de l'organigramme de décision de calcul des propriétés des sols.....	37

## Introduction

Le Gis Sol est une entité française dont l'objectif est de mieux comprendre les sols en France. Il coordonne plusieurs programmes de cartographie des sols, dont l'IGCS. Ce dernier se concentre sur la cartographie des sols à différentes échelles, du niveau régional tel que les Référentiels Régionaux Pédologiques (RRP) jusqu'à des échelles plus fines (e.g. Secteur de Référence). Cependant, la diffusion des données IGCS, notamment celles des RRP, pose des défis en termes de publication de métadonnées, d'accessibilité mais aussi de valorisation de ces données.

Le projet Solid'R, dans lequel s'insère ce stage, vise à résoudre ces problèmes en harmonisant les métadonnées des RRP et en développant des outils numériques pour utiliser ces données. L'équipe Sols de Bordeaux Sciences Agro pilote ce projet, avec pour objectif de faciliter l'accès et l'utilisation des données sur les sols en France.

Les sols jouent un rôle vital dans les écosystèmes en assurant de multiples fonctions. Parmi elles, celles de filtre et de stockage d'eau sont essentielles pour, par exemple, la production de biomasse. Le réservoir en eau utilisable maximal des sols est un indicateur très utile pour quantifier ce stockage d'eau dans le sol. Les sols et leurs réserves en eau sont cependant également vulnérables et menacés en partie par l'activité humaine. La connaissance des sols et de leur fonctionnement permet donc de les exploiter et de les protéger efficacement.

La production d'indicateurs pertinents, en plus de leur représentation cartographique permet d'augmenter cette connaissance des sols, surtout lorsque ces productions sont valorisées efficacement. Les outils numériques permettent de produire ces résultats de manière rapide et quasi-automatisée, en plus de permettre une reproductibilité et un partage incomparable de par leur nature virtuelle. Une problématique se forme donc :

« Comment concevoir et utiliser les outils numériques pour valoriser des données pédologiques ? En quoi ces méthodes représentent-elles un avantage ? »

Dans un premier temps, on s'intéressera à l'environnement de production et de diffusion des données pédologiques en France, ses objectifs futurs et ses défis à surmonter. Par la suite, on verra les bases techniques de l'outil numérique produit dans le cadre du stage, l'extraction et le calcul des propriétés nécessaires à l'estimation de l'indicateur choisi, ici le Réservoir en eau Utilisable Maximal des sols ou RUM. Enfin, nous terminerons par l'explication de la seconde partie de l'outil numérique qui traite de la représentation cartographique de l'indicateur.

## 1. Contextualisation du stage

### 1.1. Le Gis Sol, entité pilote des programmes de cartographie en France

Le **Gis Sol** (Groupement d'Intérêt Scientifique sur les sols) est une entité créée en 2001 qui se consacre à la constitution et à la gestion d'un système d'informations sur les sols de France. Il agit pour l'acquisition d'échantillons de sols, de données sur les propriétés des sols, et progressivement, pour une construction d'indicateurs opérationnels de la qualité des sols en appui aux politiques publiques et aux acteurs privés.

Le sol est le principal support des activités humaines. Or, c'est une ressource naturelle non renouvelable à l'échelle humaine, soumise à des menaces telles que l'érosion, les pollutions, l'artificialisation ou la perte de biodiversité. Ces nombreux phénomènes sont majoritairement liés aux activités de l'homme. L'objectif commun du Gis Sol est donc de mieux connaître et comprendre les sols, pour mieux les gérer. Il regroupe notamment les intérêts de divers acteurs comme les Ministères en charge de l'agriculture et de l'environnement, Inrae, l'IGN ou l'ADEME (Inrae, 2021). Le Gis Sol coordonne 4 grands programmes nationaux de cartographie pour répondre à ces objectifs : l'IGCS, le RMQS, la BDAT et la BDETM (Tableau 1).

Tableau 1 - Les programmes du GIS Sol (Gis Sol, 2023)

Sigle du programme	Désignation	Objectifs	Echelles de résolution / Limites d'usage des données
IGCS	Inventaire, Gestion et Conservation des Sols	Produire un système d'informations spatialisées des sols d'un territoire	Echelles multiples, du 1/250000 au 1/10000 sur tout le territoire
RMQS	Réseau de Mesures de la Qualité des Sols	Surveillance de sites pour évaluer la fertilité des sols, le stock de carbone ou la présence de certains éléments en traces	2240 sites répartis sur tout le territoire français. 1 site tous les 16 km
BDAT	Banque de Données des Analyses de Terre	Suivi spatial et temporel de paramètres des horizons de surface des sols cultivés	Moyenne échelle. Données de paramètres pédologiques agrégées au canton. Répartition non homogène sur le territoire
BDETM	Base de Données Eléments Traces Métalliques	Rassembler les analyses des teneurs en éléments traces métalliques des horizons de surface de sols cultivés, recevant des boues de station d'épuration	Sols agricoles, en situation plane. Répartition non homogène sur le territoire. Production d'indicateurs statistiques pour 79 départements et 232 régions agricoles (Gis Sol, 2023).

## 1.2. Le programme de cartographie multi-échelles Inventaire, Gestion et Conservation des Sols (IGCS)

Le programme **IGCS** est une cartographie d'inventaire de la nature des sols et de leur distribution spatiale sur un territoire, à différentes échelles. Cela est permis par la réalisation de différentes missions successives répondant aux règles de cartographie des sols (Legros, 1996): l'identification, la définition et la localisation des types de sols dans un premier temps, puis l'élaboration de documents cartographiques, la production de données de qualité et enfin leur valorisation à travers des sorties cartographiques thématiques par exemple (Laroche, Sauter, & Schnebelen, 2013). Au sein d'IGCS, trois volets de cartographie sont définis selon l'échelle de représentation :

- Le **Référentiel Régional Pédologique ou RRP** qui constitue un système d'informations sur les sols d'une région ou d'un département, à une échelle du 1/250 000,
- Le volet Connaissance Pédologique de la France ou CPF, aux échelles moyennes (1/100 000 à 1/50 000) permettant d'améliorer la connaissance de la variabilité des sols sur des territoires plus restreints (bassins versants, petite région agricole...),
- Les Secteurs de Référence ou SR, études de sols à grande échelle ( $\geq 1/10 000$ ) abordant des problématiques agricoles et environnementales à l'échelle parcellaire.

Ces trois volets IGCS reposent sur un socle commun : une démarche harmonisée et cadrée d'acquisition des données (Arrouays, et al., 2004), leur capitalisation dans la base de données nationales Donesol (Grolleau, et al., 2004), et la coordination par l'unité Info&Sols d'Inrae, d'un réseau de partenaires régionaux en charge de la constitution des cartographies (Figure 1).

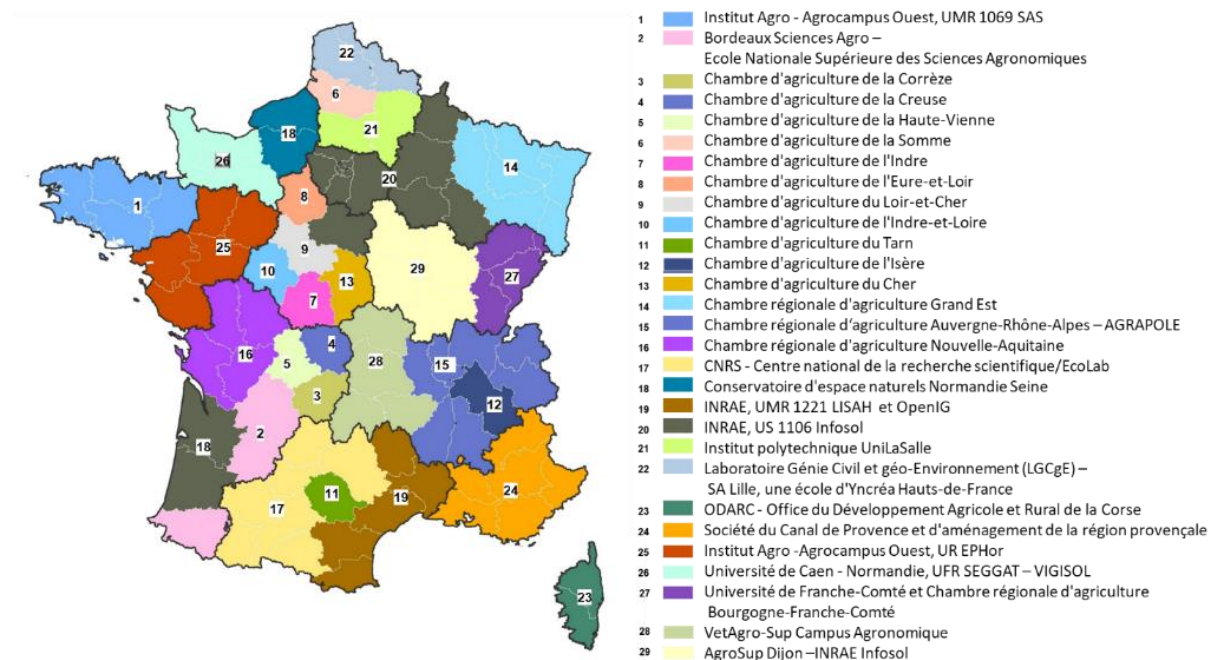


Figure 1 - Réseau de partenaires régionaux IGCS (Bispo, et al., 2020)



Les volets IGCS ne présentent pas le même état d'avancement, notamment pour les RRP à 1/250 000 qui couvrent aujourd'hui la quasi-totalité du territoire français métropolitain (Figure 2). En effet, la priorité du Gis Sol a été la réalisation du volet « petite échelle » dans l'optique d'obtenir une couverture complète du territoire français dans des délais raisonnables (Arrouays, et al., 2004).

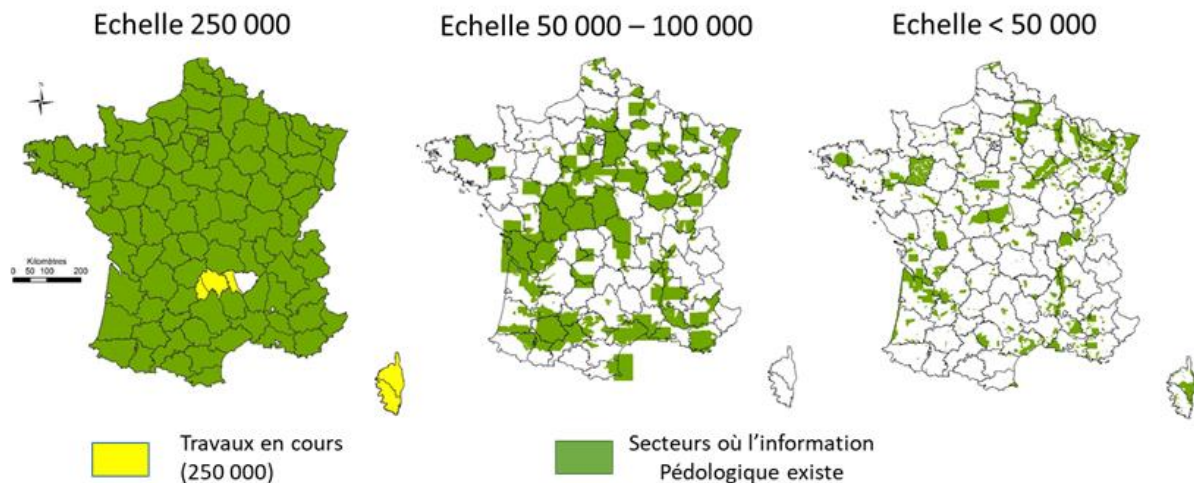


Figure 2 - Etat des lieux de la couverture cartographique des sols en France métropolitaine et Corse (Laroche, Richer de Forges, & Jalabert, À la découverte des sols : panorama des données, méthodes et outils existants, 2023)

Aujourd'hui, on compte 71 RRP finalisés en France métropolitaine. L'étape suivante est la diffusion de ces cartographies. Il convient à présent de se demander à quel stade en est cette diffusion.

### 1.3. Etat des lieux de la diffusion des Référentiels Régionaux Pédologiques (RRP)

La diffusion d'une donnée doit répondre à plusieurs objectifs : informer que la donnée existe et fournir ses caractéristiques, connaître ses modalités d'accès et mettre à disposition la donnée. La fiche de métadonnées, outil de publication de données permettant de répondre à ces objectifs, constitue la porte d'entrée de la donnée.

Les RRP constituent des données à caractère géographiques et leurs métadonnées peuvent être cataloguées au sein de plateformes d'informations géographiques ou IDG, au niveau national, régional ou plus local (Bispo, et al., 2020). La gestion de la diffusion des RRP relève à l'heure actuelle des partenaires régionaux IGCS qui ont eu en charge la réalisation de ces cartographies.

#### 1.3.1. Diffusion sur le Géoportail

Au niveau national, un produit cartographique issu des RRP, présent sur le Géoportail, a permis de booster la connaissance du programme IGCS et notamment le volet RRP : il s'agit de la carte des sols dominants en France métropolitaine. Cette production cartographique a été à l'initiative d'un groupe de travail constitué de membres du **Réseau Mixte Technologique Sols et Territoires**

Figure 3 (Figure 3) et du Gis Sol (Messant, et al., 2021).

Les métadonnées RRP ne sont toutefois pas toutes accessibles depuis le clic sur les polygones de cette carte, car l'état de publication des fiches de métadonnées RRP est hétérogène sur le territoire.

### **RMT Sols et Territoires, un réseau de plus de 40 partenaires**

La diffusion des données sols en France à ce jour est orchestrée par le Gis Sol, l'Association Française pour l'Etude des Sols et le réseau mixte technologique (RMT) Sols et Territoires. Créé en 2010 pour mieux faire connaître les données sur les sols, favoriser leurs utilisations et répondre à divers enjeux territoriaux, le RMT est aujourd'hui à sa 3<sup>ème</sup> génération (2020-2024). Les travaux proposés s'articulent autour de 3 axes thématiques (Agriculture-Agroécologie ; Forêts-Cultures pérennes et Collectivités et pouvoirs publics) et 4 axes transversaux portant sur l'animation du réseau (Axe 1), l'accès aux données sols (Axe 2), l'acquisition et la capitalisation de données (Axe 3) et la formation (Axe 4). L'axe 2 se concentre sur l'amélioration de la diffusion et de l'accessibilité des données pédologiques brutes ou traitées, avec des actions sur la diffusion sur le web, la production d'outils d'extraction et l'estimation de propriétés fonctionnelles des sols mobilisables dans des modèles (e.g. OAD, OEAE).

Figure 3 - Encadré : RMT Sols et Territoires (RMT Sols & Territoires, 2023)

### **1.3.2. Diffusion en régions**

Deux inventaires de métadonnées liées aux RRP ont été réalisés en 2021 et mis à jour en avril 2022 par Stéphanie Jalabert (Bordeaux Sciences Agro) et Christine Le Bas (Inrae Info&Sols), procédant par mots-clefs au sein des catalogues d'IDG nationales ou régionales, mais aussi sur le moteur de recherche Google®. Malgré les 71 RRP finalisés, seulement 27 fiches de métadonnées sont recensées sur des IDG et 21 RRP sont référencés sans métadonnées sur d'autres supports web (sites de chambres d'agriculture). On peut également constater une disparité importante entre les régions, anciennes ou nouvelles (Annexe 1 **Erreur ! Source du renvoi introuvable.**). La diffusion des RRP n'est pas aussi avancée que leur création.

Le format de diffusion et de valorisation de données extraites des RRP comme sur le Géoportail fait partie de dynamiques existantes de valorisations thématiques. L'une des premières valorisations de cette forme a vu le jour sur l'IDG GéoSas de l'Unité Mixte de Recherche SAS. Un portail thématique Sol, intitulé « Sols de Bretagne » (Sols de Bretagne, 2015) propose parmi d'autres outils, un visualiseur cartographique sous le code source Mviewer, permettant d'accéder à un panel de données cartographiques (cartes de propriétés de sols, sous divers formats, agrégées à la commune ou non...) (GéoSas.fr, Institut Agro Rennes & Agath.a, 2023). La métadonnée des RRP est également reliée à l'information sur la carte.

Depuis l'apparition de ce portail « Sols de Bretagne », plusieurs visualiseurs similaires ont vu le jour pour d'autres plateformes régionales, comme « [Cartographie des sols d'Alsace](#) » en production depuis l'IDG régionale DataGrandEst, ou « [Données Sols en Occitanie](#) » en production depuis OpenIG.

Trois constats peuvent être dressés de la situation des RRP et de leur diffusion : les métadonnées ne sont pas publiées pour une grande partie des RRP, la donnée est visible sur le Géoportail, ce dernier étant un portail central de sources de données géographiques en France, mais celui-ci ne renvoie pas vers l'accès aux données RRP, et enfin il existe plusieurs supports de visualisation cartographique en OpenSource sur différentes plateformes régionales. C'est de

ces constats qu'est né le projet **Solid'R** « Soutien à la diffusion des Référentiels Régionaux Pédologiques », avec pour objectif de répondre à la problématique suivante : « comment accompagner les gestionnaires RRP à valoriser leurs données sol ? »

#### 1.4. Le projet Solid'R pour soutenir la diffusion des RRP

Le projet Solid'R (avril 2022 – déc. 2023) présente deux objectifs : l'accompagnement des partenaires régionaux IGCS dans la mise en conformité et la diffusion des métadonnées des RRP (RMT Sols et Territoires, 2023) et l'aide à la valorisation des données issues des RRP. En effet, la connaissance et l'accès à la donnée sont très importants, mais la communication sur l'usage que l'on peut en faire et la présentation d'exemples concrets le sont tout autant. L'enjeu du projet Solid'R est donc également de montrer le chemin vers l'utilisation des données des RRP. Le projet Solid'R est complètement en lien avec l'axe 2 du RMT Sols et Territoires 2020-2023 sur la diffusion des données sols (RMT Sols et Territoires, 2023). Le pilote du projet est l'équipe Sols de Bordeaux Sciences Agro, à travers l'engagement de Stéphanie Jalabert, ingénieure d'études en sciences du sol, membre fondateur du RMT S&T, copilote de l'Axe 2 et responsable de l'action 2.1.

##### 1.4.1. L'équipe Sols de Bordeaux Sciences Agro, pilote du projet Solid'R

Bordeaux Sciences Agro est un établissement public d'enseignement supérieur et de recherche agronomique sous tutelle du Ministère en charge de l'agriculture. L'équipe Sols de Bordeaux Sciences Agro participe aux programmes nationaux d'études des sols que sont l'IGCS, le RMQS ou d'autres en lien avec la cartographie des sols de zones humides par exemple. Partenaire régional des programmes du Gis Sol, l'équipe Sols y intervient à différents niveaux, de l'acquisition de données de terrain à la gestion et à la valorisation des bases de données spatialisées. Pour la Nouvelle-Aquitaine (N-A), l'équipe agit en tant que gestionnaire pour les RRP des départements 24, 47 et 64, et co-gestionnaire avec la Chambre Régionale d'Agriculture pour les RRP 16, 17, 79 et 86. En parallèle, la supervision des RRP 33 et 40 est assurée par Inrae Info&Sols. Une particularité importante à noter est que l'ensemble des RRP de la région N-A sont accessibles en OpenData (données ouvertes, *i.e.* accessibles gratuitement, sous licence Etalab 2.0) par le biais de la plateforme PIGMA, infrastructure régionale de données géographiques en N-A. Cette initiative s'aligne sur une démarche de transparence et de partage qui convergent vers une gestion plus efficiente des ressources pédologiques, favorisant la prise de décision éclairée et le développement durable au sein de la région N-A.

##### 1.4.2. Avancement du projet Solid'R

La première phase du projet Solid'R était une analyse des opportunités existantes pour structurer et harmoniser la diffusion des données RRP, avec l'aide d'un groupe de master SIGAT (janv.–mai 2023). Dans ce contexte, le groupe a également travaillé sur un premier essai de développement d'outil numérique, *i.e.* un script informatique, qui permet de calculer une propriété des sols à partir de la base de données brute.

La seconde est de guider les gestionnaires RRP dans l'usage des outils numériques :

- De type script informatique, comme le travail réalisé par les étudiants SIGAT, basé sur les données des RRP ;
- De type visualiseur cartographique, tel que « Sols de Bretagne » sous application Mviewer dont les codes sources sont libres d'accès.

L'enjeu est de fournir les bonnes bases aux gestionnaires RRP pour réussir la valorisation de leurs données, de partager la prise en main de ces outils, et de pouvoir les reproduire et les adapter facilement.

### 1.5. Les objectifs du stage

Les objectifs du stage sont reliés à la seconde phase du projet Solid'R. Tout d'abord, il s'agissait de développer un outil numérique permettant de calculer des indicateurs de propriétés des sols à partir des données disponibles dans les bases de données, ou BDD, DoneSol. La forme choisie pour cet outil est un script en langage Python. L'objectif opérationnel est d'obtenir un outil quasi « en un clic » : de l'intégration de la BDD brute, à son traitement, jusqu'à la mise en forme cartographique, en passant par la vérification préalable des données (Figure 4).

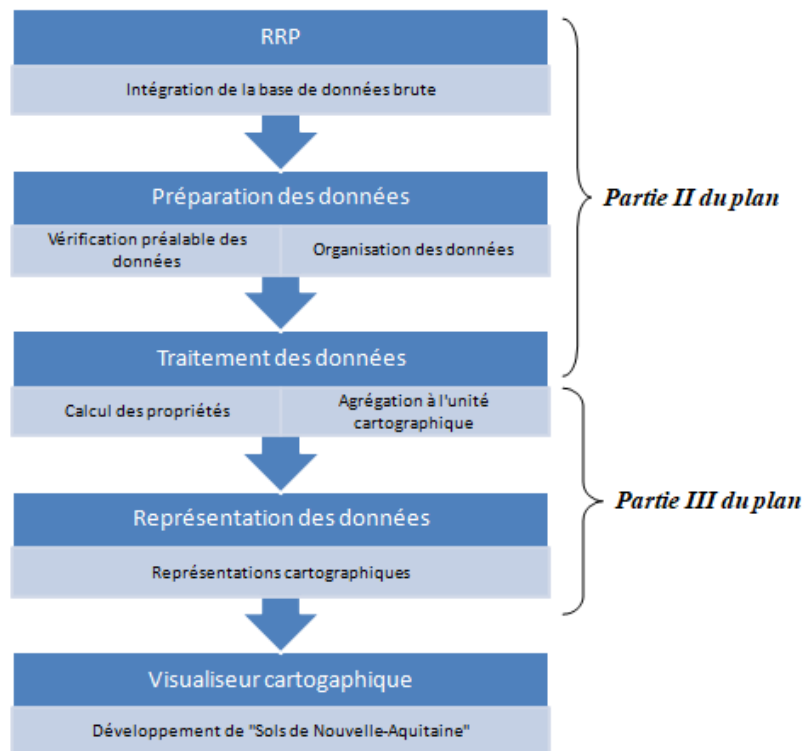


Figure 4 - Etapes du script et délimitation par parties du plan

Le script se veut dans sa conception le plus clair possible, que ce soit dans son utilisation, sa structure ou dans son code, pour permettre sa réutilisation et son adaptation à des usages similaires. Il doit donc également être accompagné à terme d'un guide d'utilisation.

La seconde mission du stage était de commencer à réfléchir et à développer un visualiseur cartographique simplifié « Sols de Nouvelle-Aquitaine », basé sur le visualiseur « Sols de Bretagne ». L'équipe Sols souhaite mettre en production une première version de ce Mviewer sur PIGMA courant année 2024.

## 2. Développement de l'outil numérique de calcul d'indicateurs de propriétés des sols : cas du réservoir utilisable maximal en eau des sols

Cette section explore en détail l'organisation de la première partie du script ainsi que les différentes étapes impliquées dans la préparation des données et le calcul des indicateurs. Le script est conçu pour traiter les données des Référentiels Régionaux Pédologiques (RRP) et générer les informations nécessaires à l'analyse des propriétés des sols, ici le réservoir en eau utilisable maximal des sols. Le RRP utilisé en exemple et pour la majeure partie du développement est le RRP du Lot-et-Garonne (RRP 47). Le RRP 47 a été choisi en particulier car il a été réalisé par l'équipe Sols de Bordeaux Sciences Agro, mais surtout car c'est le premier qui a été réalisé. Il présente le plus de valorisation, et a permis des retours rapides de la part de l'équipe Sols, entraînant de fait un contrôle plus efficace.

### 2.1. Estimation du Réservoir Utilisable Maximum en eau des sols

#### 2.1.1. Le RUM : définition et formules de calcul

La propriété étudiée ici est le **réservoir en eau utilisable maximal** des sols ou **RUM** exprimé en mm d'eau. Ce dernier représente la capacité maximale théorique de rétention en eau du sol. La disponibilité de l'eau dans un sol, pour les cultures notamment, dépend des états d'humidité dans le sol (Figure 5). Quand le sol est saturé en eau ( $\theta_{sat}$ ), l'eau dite gravitaire circule librement dans le sol - elle n'est pas retenue par les particules du sol - jusqu'à atteindre le point de ressuyage du sol ou la capacité au champ ( $\theta_{cc}$ ). A partir de cet état d'humidité, il n'y a plus d'eau gravitaire, l'eau est retenue par le sol, elle devient utilisable par les plantes. Sans apport d'eau (pluie, irrigation, *etc.*), cet état d'humidité va se dégrader jusqu'au point de flétrissement permanent, au-delà duquel l'eau est trop retenue par le sol et devient indisponible pour les racines des cultures ( $\theta_{fp}$ ). Entre la capacité au champ et le point de flétrissement permanent, l'eau du sol est utilisable : on parle alors de **réservoir utilisable en eau du sol**.

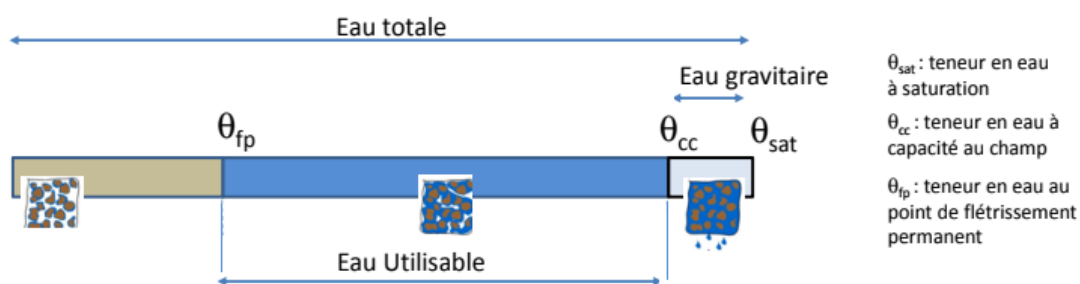


Figure 5 - Notion du réservoir utilisable en eau des sols (Doussan, 2017)

Le Réservoir en eau Utilisable Maximal des sols (RUM) est évalué en fonction des caractéristiques du sol et sur toute sa profondeur (Bouthier & al., 2022). En effet, développés à partir d'un matériau géologique, les sols sont constitués de particules minérales de taille variable : les argiles, limons et sables qui forment les **terres fines** (TF) et les **éléments grossiers** (EG), fragments d'une taille supérieure à 2mm (graviers, cailloux, pierres et blocs).

L'eau utilisable est contenue dans la porosité du sol, retenue par les terres fines. Elle peut également être logée dans la porosité de certains EG (de nature calcaire, crayeux...) (Tetegan, 2011). Le sol se différencie en une ou plusieurs couches, appelées horizons, présentant des caractéristiques physico-chimiques différentes (épaisseur, structure, texture, teneur en EG...).

Ainsi, le RUM d'un sol est obtenu en sommant les RUM de chacun des horizons du sol (Équation 1) :

$$RUM_{SOL} = \sum_{i=1}^n RUM_i$$

*Équation 1 - Le réservoir en eau utilisable maximal (RUM) d'un sol :  $RUM_{SOL}$*

Avec  $RUM_{SOL}$ , le RUM pour l'entièreté d'un sol, et  $RUM_i$ , le RUM pour un horizon  $i$  de ce sol où  $n$  est le nombre d'horizons du sol.

Un horizon  $i$  est composé de terres fines (argiles, limons et/ou sables) avec ou non présence d'éléments grossiers. Le RUM d'un horizon  $i$  est donc la somme de la RUM des terres fines et le RUM des EG, présents dans cet horizon  $i$  (Équation 2)

$$RUM_i = RUM_{TFi} + RUM_{EGi}$$

*Équation 2 - Le réservoir en eau utilisable maximal d'un horizon  $i$  de sol :  $RUM_i$*

Avec  $RUM_{TFi}$ , le RUM contenu dans les terres fines pour l'horizon  $i$ , et  $RUM_{EGi}$ , le RUM contenu dans les éléments grossiers.

L'estimation du RUM pour les terres fines et les EG suit deux méthodes distinctes que nous détaillons ci-dessous.

#### *Estimation du $RUM_{TFi}$ selon (Bruand, Duval, & Cousin, 2004)*

La formule générale du  $RUM_{TFi}$  est donnée dans l'Équation 3 :

$$RUM_{TFi} = (\Theta_{CC} - \Theta_{FP}) * MVA_i * e_i * TFi$$

*Équation 3- RUM dans les terres fines*

Avec  $\Theta_{CC}$  l'humidité pondérale de l'horizon  $i$  à la capacité au champ (en %),  $\Theta_{FP}$  l'humidité pondérale au point de flétrissement permanent (en %),  $MVA_i$ , la masse volumique de l'horizon  $i$  (sans unité),  $e_i$  l'épaisseur de l'horizon  $i$  en cm, et  $TFi$  le pourcentage de terres fines dans l'horizon  $i$  (sans unité).

L'estimation du  $RUM_{TFi}$  suppose que l'on connaisse l'humidité à la capacité au champ et l'humidité au point de flétrissement permanent pour chaque horizon du profil de sol étudié. Cela demanderait de nombreuses opérations longues et coûteuses. C'est pourquoi des classes ou fonctions de pédotransfert (respectivement CPT ou FPT) sont utilisées, elles permettent d'estimer le  $RUM_{TFi}$  à partir de données plus facilement accessibles en simplifiant une partie des formules par estimation (Bouthier & al., 2022). La classe de pédotransfert - CPT - de (Bruand, Duval, & Cousin, 2004) permet, à partir de la classe texturale de l'horizon et du type



d'horizon (horizon de surface entre 0 et 30 cm ou horizon de profondeur), d'estimer une Humidité Utile Maximale (HUM) en mm d'eau par cm de sol. C'est la classe de pédotransfert la plus utilisée en France pour les calculs de RUM des terres fines des horizons de sols. Grâce à cette CPT, on peut alors estimer la première partie de l'Équation 3, i.e.  $(\Theta_{CC} - \Theta_{FP}) * MVA$  représentant l'HUM.

#### *Estimation du RUM<sub>EG</sub> selon (Tetegan, 2011)*

Pour estimer le RUM<sub>EGi</sub>, la méthode est très similaire, on passe par une fonction de pédotransfert également, pour simplifier la formule qui reste la même :

$$RUM_{EGi} = (\Theta_{CC} - \Theta_{FP}) * MVA_{EG} * e_i * EG_i$$

#### *Équation 4 - RUM dans les éléments grossiers*

Avec  $\Theta_{CC}$  l'humidité pondérale à la capacité au champ des éléments grossiers (en %),  $\Theta_{FP}$  l'humidité pondérale à l'humidité au point de flétrissement des éléments grossiers (en %),  $MVA_{EG}$  la masse volumique apparente des éléments grossiers (sans unité),  $e_i$  l'épaisseur de l'horizon  $i$  en cm et  $EG$  le pourcentage d'éléments grossiers dans l'horizon  $i$  (sans unité)

Pour estimer l'HUM des EG de l'horizon  $i$ , i.e.  $(\Theta_{CC} - \Theta_{FP}) * MVA_{EG}$ , on utilise pour les éléments grossiers la FPT de (Tetegan, 2011) qui utilise les valeurs de masse volumique des éléments grossiers. Ainsi on obtient la formule suivante :

$$HUM_{EGi} = (\Theta_{CC} - \Theta_{FP}) * MVA_{EG} = (a_3 * \ln(MVA_{EG}) + b_3) * MVA_{EG}$$

#### *Équation 5- HUM dans les éléments grossiers avec (Tetegan, 2011)*

Avec  $HUM_{EGi}$  l'humidité utile maximale des éléments grossiers (en mm d'eau/cm),  $\Theta_{CC}$  l'humidité pondérale à la capacité au champ des éléments grossiers (en %),  $\Theta_{FP}$  l'humidité pondérale à l'humidité au point de flétrissement des éléments grossiers (en %),  $a_3$  et  $b_3$  des coefficients obtenus selon le type d'origine sédimentaire des éléments grossiers (sans unités) et  $MVA_{EG}$  la masse volumique apparente des éléments grossiers (en  $g.cm^{-3}$ ).

#### *Formule finale du RUM d'un horizon*

A partir des différentes formules présentées au préalable, on peut donc obtenir une formule générale pour le RUM d'un horizon :

$$RUM_i = [HUM_{TFi} * e_i * TF_i] + [(a_3 * \ln(MVA_{EGi}) + b_3) * MVA_{EGi} * EG_i * (\frac{e_i}{10})]$$

#### *Équation 6 - Formule générale du RUM d'un horizon $i$*

Avec  $RUM_i$  le RUM pour un horizon  $i$ ,  $HUM_{TFi}$  l'humidité utile maximale des terres fines (en mm d'eau/cm),  $e_i$  l'épaisseur de l'horizon (en cm),  $TF$  le pourcentage de terres fines dans le sol,  $a_3$  et  $b_3$  des coefficients obtenus selon le type d'origine sédimentaire (sans unités) et  $MVA_{EG}$  la masse volumique apparente des éléments grossiers (en  $g.cm^{-3}$ ).

### *Pourquoi calculer le RUM des sols ?*

Le RUM des sols rentre dans le calcul du **réservoir en eau utilisable** ou **RU**. Le RU représente également la quantité d'eau maximale que le sol peut retenir sur une épaisseur donnée, mais il est exclusif à une espèce végétale, puisqu'il est fonction de la profondeur d'enracinement de cette espèce. Le RU est un concept utilisé dans beaucoup de domaines car il constitue un paramètre essentiel au calcul de bilans hydriques, et est intégré dans de nombreux OAD. En agronomie, la connaissance du RU permet de gérer efficacement l'irrigation des cultures, d'évaluer le potentiel de rendement d'une culture en vue de déterminer ses besoins en azote, et d'estimer les pertes de production dues à la sécheresse. Il rentre également en jeu dans les diagnostics environnementaux ou en hydrologie.

Les données du RRP permettent d'estimer le RUM à l'échelle de la France entière, ce qui permettrait à n'importe qui en France d'en avoir la connaissance rapidement. La globalité des données permettrait également de mettre en place des OAD à petite échelle pour dresser des tendances sur la ressource en eau des sols sur des territoires très étendus, notamment dans un contexte de changement climatique.

#### **2.1.2. Les données des RRP utilisées et leurs caractéristiques**

Les données des RRP sont stockées sous la forme d'une base de données sémantiques, associée à une carte de pédopaysages, construite selon le modèle DoneSol, modèle de structure informatique unique (Grolleau É. &., 2004). DoneSol est la base de données nationale structurant et regroupant les données ponctuelles et surfaciques des études pédologiques. Dans le cas de ce modèle, le système de gestion de base de données utilisé ici est Microsoft® Access, les formats des fichiers étant donc en .mdb, .csv ou .txt.

La base de données surfaciques Donesol est structurée en unités hiérarchiques (Figure 6). Une strate est formée par une série d'horizons du sol cohérents, et la combinaison de plusieurs strates forme ce que l'on appelle une Unité Topologique de Sol (UTS). La succession logique d'UTS constitue une Unité Cartographique de Sol (UCS), qui est considérée comme une partie de la couverture pédologique d'un pédopaysage spécifique. Notons qu'une même UTS peut être présente dans la composition de plusieurs UCS et que contrairement aux UCS, les UTS ne correspondent pas à des unités géographiques et ne sont donc pas représentées sur la carte des pédopaysages du RRP restituée à l'échelle du 1/250 000. En fonction du nombre d'UTS dans l'UCS, on parle d'UCS simple (une seule UTS représentant 100 % de l'UCS) ou d'UCS complexes (plusieurs UTS). Les UCS seront les unités d'intégration spatiale, les unités qui constitueront la carte finale.



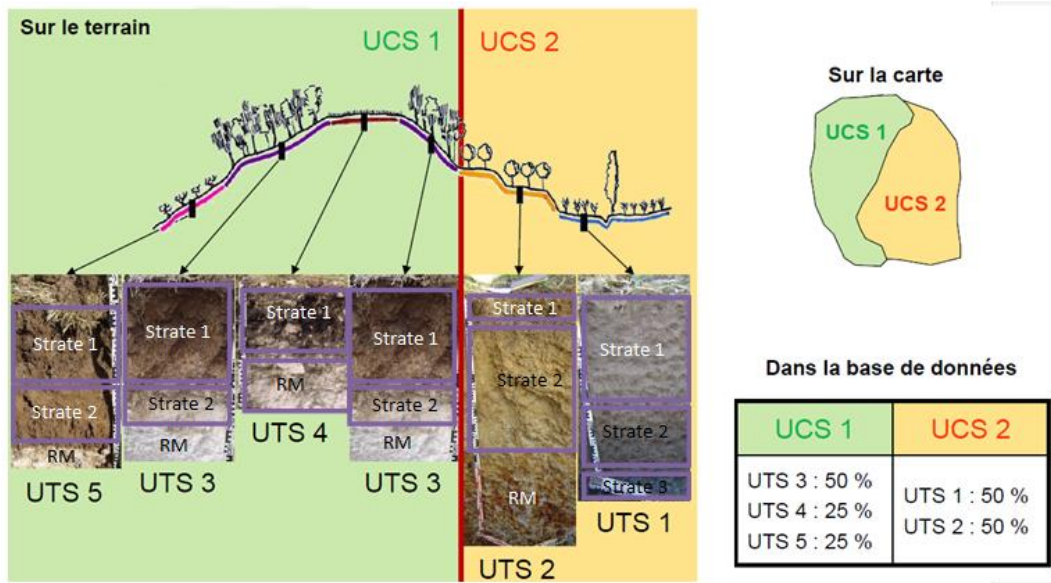


Figure 6 - Découpage des unités de sol en UCS, UTS et strates dans la BDD Donesol (Laroche, Thiry, Chenu, Moulin, & Bardy, 2014)

Pour le script, toute l’information de la base de données n’est pas nécessaire, seules certaines tables sont importées : **STRATE\_QUAL**, **STRATE\_QUANT**, **STRATE**, **L\_UCS\_UTS** et **UCS**. La table UTS n’est par exemple pas nécessaire, le lien est fait sur le champ **ID\_UTS** entre **STRATE** et **L\_UCS\_UTS**. Les tables sont reliées entre elles selon la hiérarchie présentée plus tôt (Figure 7).

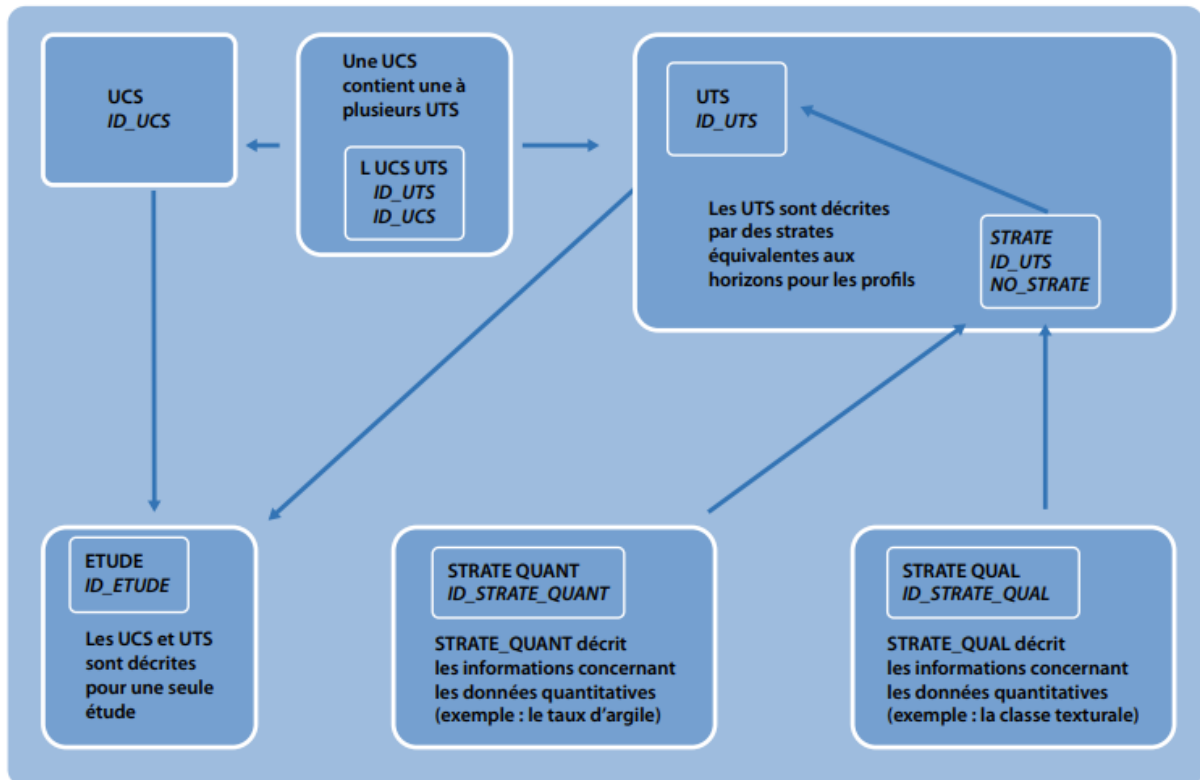


Figure 7 - Liens entre les différentes tables de DoneSol qui caractérisent les données surfaciques (Lemerrier, et al., 2017)

L'ensemble de ces tables présentent toutes les informations nécessaires pour les calculs des différentes propriétés des sols nécessaires au calcul du RUM. En reprenant la formule du RUM, on peut associer les données d'entrées à chacune des tables importées

. Parmi les variables présentées qui ne sont pas explicitement contenues dans l'Équation 6, on a la **classe texturale** (classe\_texturale\_strate) et le **type de strate** (type\_strate) qui servent à calculer  $HUM_{TFi}$ , et la **classe lithologique** (classe\_lithol) qui sert à calculer les coefficients  $a_3$  et  $b_3$ , ainsi que  $MVA_{EG}$ , ces deux calculs basés sur les fonctions de pédotransfert présentées plus tôt. Les champs pour l'**épaisseur moyenne** (epais\_strate), et **les abondances en terres fines** (ab\_tf\_strate) et en **éléments grossiers** (ab\_eg\_strate) rentrent directement dans le calcul.

De plus, en début du script, quelques statistiques descriptives sont effectuées pour pouvoir constater de la conservation de l'information dans les étapes suivantes.

Tableau 2 - Champs retenus pour chaque table importée et utilité de ces champs dans le calcul du RUM

Tables sémantiques de Donesol	Champs retenus	Désignation du champ	Objectifs	Enregistrements dans le RRP 47
<b>UCS</b>	no_ucs, id_ucs	Numéro de l'UCS. Champ de liaison entre la table UCS et la table attributaire de la couche géographique	Restitution cartographique de l'indicateur de propriété des sols	84 UCS
<b>L_UCS_UTS</b>	pourcent, id_uts, id_ucs	Pourcentage surfacique de l'UTS au sein de l'UCS	Etape préalable à l'agrégation de l'information à l'UCS	341 lignes
<b>STRATE</b>	no_strate, id_uts, epais_min, epais_moy, epais_max, rp_95_nom, rp_2008_nom, cpes_nom, prof_fin_moy, prof_appar_moy	Numéro de la strate, identifiant de l'UTS de la strate, différentes variables propres à la strate	Agrégation des strates à l'UTS. Calculs des champs créés : epais_strate, type_strate	617 strates
<b>STRATE_QUANT</b>	TAUX LIMON, TAUX ARGILE, TAUX SABLE, ABONDANCE_EG	Taux de terres fines et d'éléments grossiers	Calculs des champs créés : classe_texturale_strate, ab_tf_strate, ab_eg_strate	
<b>STRATE_QUAL</b>	TEXTURE_AISNE, TEXTURE GEPPA, NOM EG	Texture de la strate	Calculs des champs créés : classe_texturale_strate, classe_lithol	

## 2.2. L'outil numérique : script en langage Python sous environnement Google Colab

### 2.2.1. Le langage Python

Le script est programmé exclusivement dans le langage Python en version 3.11.5 (Python, s.d.), ce qui entraîne de multiples contraintes et méthodes liées au langage. Python est reconnu pour sa simplicité et sa lisibilité, ce qui facilite grandement le développement et la maintenance du code. De plus, Python est un langage populaire et fréquemment utilisé, facilitant donc les échanges autour du script.

Sa vaste bibliothèque de modules et d'outils dédiés à l'analyse de données, à la manipulation de fichiers et aux calculs scientifiques offre un environnement propice à la réalisation des traitements impliqués dans l'analyse des propriétés des sols.

### 2.2.2. L'environnement Google Colab

Le choix s'est porté sur la plateforme Google Colaboratory en vue de la mise en œuvre du script ainsi que de sa transmission. Cette sélection repose sur la simplicité d'accès et la dimension collaborative (mais pas en temps réel) qu'offre la plateforme (Google, 2023). De plus, elle assure l'hébergement temporaire des données pendant la session. Cette option présente l'avantage de mettre à disposition des machines virtuelles<sup>1</sup>, ce qui permet de profiter d'une capacité de stockage considérable et une rapidité de traitement indépendante des spécificités de son ordinateur. Les limites exactes de Colab ne sont cependant pas précisées par Google, « *entre autres parce qu'elles peuvent varier rapidement* » (FAQ Colaboratory, 2023).

Il convient de noter que travailler au sein d'un environnement local pourrait impliquer l'adoption de méthodes alternatives pour l'initialisation du contexte de travail et l'importation des données. Dans l'environnement de Colab, il suffit d'importer les données via le volet latéral gauche et d'exécuter les lignes de script dédiées à la lecture des fichiers.

Il est toutefois important de remarquer que le choix de l'environnement Colab, et dans une certaine mesure du langage Python est avant tout plutôt arbitraire, reprenant le travail et les méthodes des étudiants SIGAT ayant préalablement travaillé sur le projet Solid'R dans sa première phase.

Tous ces choix réalisés, de l'indicateur utilisé en exemple, au langage et environnement mobilisés, en passant par la forme des données exploitées, permettent alors de se lancer dans la réalisation de l'outil numérique en question.

### 2.2.3. Un script complexe structuré en 7 étapes

L'efficacité, la lisibilité et la réutilisabilité de l'outil de calcul des indicateurs de propriétés des sols dépendent fortement de l'organisation du script. Afin d'assurer la fluidité des opérations, les différentes sections du script ont été méticuleusement structurées (Figure 8).

---

<sup>1</sup> Environnement virtualisé qui fonctionne sur une machine physique, bénéficiant des mêmes équipements : CPU, mémoire RAM, disque dur...

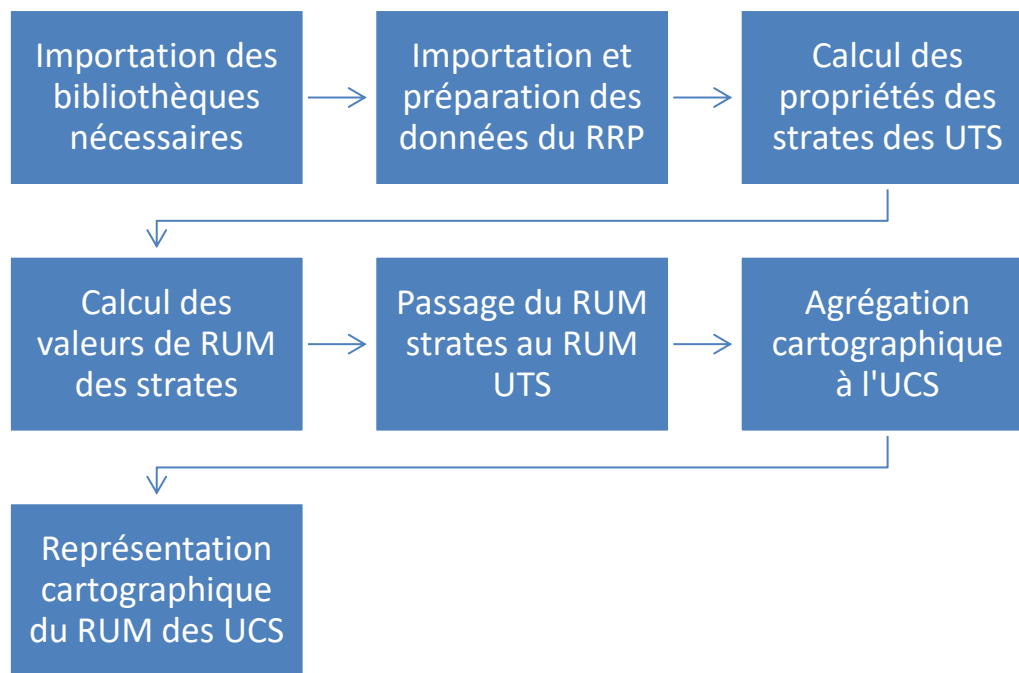


Figure 8 - Les étapes du script Python du calcul du RUM des sols du RRP47 avec sorties cartographiques

#### Etape 1 - Importation des bibliothèques nécessaires

Les bibliothèques (ou packages) sont les sources des fonctions qui sont utilisées dans le script. Un ensemble de bibliothèques Python a été sélectionné pour soutenir la mise en œuvre du processus analytique (Tableau 3). Ces bibliothèques ont été intégrées afin d'accomplir des tâches spécifiques et complémentaires tout au long de l'analyse des données.

Tableau 3 - Bibliothèques Python utilisées pour tout le script

Bibliothèque	Description
<b>Importlib</b>	Importée en début de script pour gérer les imports des autres bibliothèques.
<b>Pandas</b>	Importée pour la gestion efficace des données tabulaires, permettant manipulation, transformation et agrégation des informations de manière structurée.
<b>Numpy</b>	Utilisée pour la manipulation performante des tableaux et des opérations mathématiques sur les données numériques.
<b>Access_parser</b>	Fondamentale pour l'extraction des données depuis les fichiers de format Microsoft® Access.
<b>Glob et Re</b>	Brièvement utilisées pour rechercher les noms formatés des fichiers.

L'ensemble de ces bibliothèques a été choisi pour leur adaptabilité et leurs fonctionnalités spécifiques qui répondent aux diverses exigences analytiques du projet. D'autres bibliothèques sont importées au même moment, mais elles seront présentées par la suite car elles sont utilisées dans un second temps.

Dans le contexte de l'importation des fichiers, le script a tiré parti de la fonctionnalité de lecture fournie par la bibliothèque **access\_parser** pour extraire les données des fichiers au format Access. Le format du nom de fichier lu suit une convention spécifique pour permettre de récupérer le numéro du département du RRP, non précisé autrement : il commence par 100XX (avec XX le numéro du département). Le nom du fichier pour le RRP 47 est « 10047\_mai\_2015\_Niveau\_2 ».

Pour la structuration des données, un concept clé est le "DataFrame". Il s'agit d'une structure de données tabulaire fournie par la bibliothèque **pandas**, permettant d'organiser les données de manière intuitive et flexible.

### *Etape 2 - Importation et préparation des données du RRP*

Au commencement de ce processus, se trouvent l'extraction et la préparation des données essentielles pour les calculs des indicateurs comprenant plusieurs étapes majeures.

Les informations nécessaires sont importées à partir des fichiers Access au format « .mdb » du RRP du département correspondant. Cette importation garantit la disponibilité des informations pertinentes pour les analyses à venir. Quelques spécificités sont à relever : la fonction d'import, utilisant la bibliothèque **glob** repère le format « .mdb » du fichier et pas son nom. C'est également à ce moment que le numéro du département du RRP est récupéré et stocké, en utilisant la bibliothèque **re**.

Les données extraites sont ensuite organisées dans un unique DataFrame, une structure tabulaire puissante. L'utilisation de cette structure facilite la manipulation des données et permet la mise en œuvre d'une variété d'opérations. Une étape importante a consisté à créer un identifiant unique, "**id\_strate**", en concaténant les champs "**id\_uts**" et "**no\_strate**" pour chaque table. Il permet d'identifier à tout moment chaque élément et de le relier à son UTS ou sa strate. Un tel identifiant n'existe pas dans la structure de la base de données DoneSol. Cette concaténation a servi ultérieurement de champ de jointure entre les tables.

Cet unique DataFrame créé a servi de pivot central pour les opérations analytiques. Il est issu du DataFrame importé « **strate** », mais change de nom à chaque étape du script, par souci de conservation d'une trace des changements (exemple de nom : **strate\_modif\_v0**). Les données importantes issues des autres DataFrame importés sont ajoutées à « **strate** », notamment les champs *TEXTURE AISNE*, *TEXTURE GEPPA*, *TAUX LIMON*, *TAUX ARGILE*, *TAUX SABLE ET ABONDANCE\_EG*. Pour ajouter ces derniers, un changement de modélisation de long en large sur la colonne du nom des variables est utilisé : les noms des variables deviennent les noms de colonne, puis on sélectionne seulement les variables qui nous intéressent.

Le DataFrame obtenu présente un nombre important de champs (33, la limite de Google Colab étant fixée à 20) ne permettant pas de visualiser toutes les informations en une fois dans l'environnement Google Colab. Cependant toutes les informations sont bien présentes en une seule table ce qui permet de les utiliser à n'importe quel moment du script. Un extrait du DataFrame est présent ci-dessous en tant qu'exemple (Figure 9).

	id_strate	TEXTURE AISNE	TEXTURE GEPPA	TAUX LIMON	TAUX ARGILE	TAUX SABLE	ABO
0	100321	ALO	AA	374.0	528.0	79.0	
1	100322	ALO	AA	316.0	559.0	125.0	
2	100323	A	AA	356.0	304.0	330.0	
3	102031	A	AI	350.0	400.0	228.0	
4	102032	AL	AI	537.0	340.0	123.0	

Figure 9 - Extrait du DataFrame après l'import et les ajouts des champs de STRATE\_QUAL et STRATE\_QUANT (Equipe Sols de Bordeaux Sciences Agro)

### Etape 3 - Calcul des propriétés des strates des UTS

Cette partie du script représente le cœur du processus, *i.e.* le calcul des propriétés des strates nécessaires à l'estimation du RUM. Tout d'abord des champs vides sont créés dans le DataFrame pour accueillir les résultats des calculs. Cette démarche organisée permet de les modifier au fur et à mesure du script, sans avoir à ajouter toutes les valeurs en même temps. Pour certains cas spéciaux, cela permet notamment de directement modifier la valeur de RUM de la strate et éviter de devoir la reconsidérer par la suite. Elle permet également d'éviter toute perte d'information, en accumulant des données au fur et à mesure sans jamais en supprimer.

Les caractéristiques clés des sols sont ensuite déterminées en fonction des données disponibles selon un organigramme préconçu (Lien en Annexe 2 **Erreur ! Source du renvoi introuvable.**). Cela implique : le calcul de **l'épaisseur moyenne** des différentes strates, la détermination **de la classe texturale**, l'identification des **types de strates**, ainsi que l'estimation de **l'abondance de la terre fine** et des **éléments grossiers**. Les conditions de calcul sont hiérarchisées, certaines variables conditionnent le calcul des autres, l'ordre donné ci-dessus est l'ordre réel de calcul dans le script.

Pour certains champs, les valeurs sont **quantitatives**, comme pour **l'épaisseur moyenne** des strates, **l'abondance en terres fines** des strates ou **l'abondance des éléments grossiers** des strates, elles ne demandent qu'un calcul basé sur les informations provenant de STRATE\_QUANT et STRATE. Certaines informations de STRATE\_QUAL rentrent cependant dans la décision dans certains cas précis, comme le champ type\_strate par exemple pour l'épaisseur moyenne.

Pour d'autres champs, l'information est qualitative, comme pour la **classe texturale** des strates ou le **type de strate** (strate de surface ou de profondeur) (Bruand, Duval, & Cousin, 2004). Dans la détermination de ces champs, certaines données quantitatives peuvent être utilisées, mais un choix arbitraire est fait à un moment. Il peut déjà avoir été fait à la création de la donnée et stocké dans STRATE\_QUAL, comme pour les champs de classes texturales selon le triangle de l'Aisne TEXTURE AISNE ou selon le triangle Geppa TEXTURE GEPPA. Il peut également être fait à partir de sources externes, comme le tableau de correspondance AISNE/GEPPA



d'après (Richer-De-Forges, Feller, Jamagne, & Arrouays, 2008). Enfin dans certains cas, il est juste décidé par des experts, comme plusieurs seuils déterminés.

A chaque étape, un tableau est exportable en format .csv. Cela permet de constater des modifications ou de l'utiliser pour d'autres travaux, lorsque toutes les étapes ne sont pas voulues. C'est une **fonctionnalité très importante**, l'information la plus proche de la réalité est présente dans ces tableaux, et pas dans les cartes créées par la suite, car une carte implique des choix de représentation et donc une perte d'information. L'information est récupérable pour chaque unité pédologique, chaque variable, à toute étape du script. D'autres formes de valorisation que la cartographie de la suite du script peuvent utiliser ces données.

Des vérifications par des comptages sont également effectuées à chaque étape. Cela permet de s'assurer du bon fonctionnement du script, et si des comptages semblent donner des valeurs illogiques, cela peut traduire un dysfonctionnement du script ou une erreur dans la base de données initiale. Une comparaison avec les statistiques effectuée avant les calculs peut également permettre de détecter une perte d'information.

#### *Etape 4 - Calcul des valeurs de RUM des strates*

Après avoir calculé tous les indicateurs nécessaires, on peut calculer le RUM pour chaque strate, en se basant sur l'Équation 6 - Formule générale du RUM d'un horizon  $i$  :

$$RUM_i = [HUM_{TFi} * e_i * TFi] + [(a3 * \ln(MVA_{EGi}) + b3) * MVA_{EGi} * EG_i * (\frac{e_i}{10})]$$

Tout d'abord, on calcule les RUM Terres fines et Éléments grossiers. Pour cela, des DataFrames de référence basés sur les classes et fonctions de pédotransfert évoquées plus tôt, (Bruand, Duval, & Cousin, 2004) et (Tetegan, 2011) respectivement, sont créés. À partir de ces DataFrames, les propriétés manquantes pour appliquer les formules sont déterminées pour chaque strate, il est ensuite possible de calculer les deux valeurs  $RUM_{TF}$  et  $RUM_{EG}$ .

Les résultats calculés sont alors sommés pour obtenir le **réservoir utilisable maximum en eau total** pour chaque strate. Au fil des étapes, s'il s'avère que des cas spécifiques ont rendu impossible le calcul d'une des propriétés, la valeur **-99** est attribuée au RUM de la strate, ce qui correspond à « **Données incomplètes** ». Pour le RRP 47, seules 5 strates sur les 617 présentent une valeur de -99 attribuée au RUM.

#### *Etape 5 - Passage de strate à UTS*

Avant de passer à la seconde partie du script, les résultats sont agrégés à l'UTS. Pour chaque UTS la démarche est la même : si une des strates a une valeur de **-99** pour le RUM, alors le RUM de l'UTS est également **-99**. Sinon dans le nouveau champ **RUM\_uts** créé, on ajoute la somme du RUM de chaque strate composant l'UTS. De plus, on crée pour chaque UTS le champ **Epaiss\_uts** en sommant l'épaisseur de chaque strate, ce champ sera utilisé dans la seconde partie du script.

## 2.3. Limites et perspectives

### 2.3.1. Performances du script

Le temps d'import des bibliothèques représente 20 secondes à la première exécution, mais est presque instantané par la suite. Pour ce qui est de l'import du fichier Microsoft® Access, le temps est d'environ 45 secondes pour le RRP47 par exemple (taille de 9088 Ko). Ce temps est lui répété à chaque nouveau RRP testé. Pour discuter de ces temps, il est nécessaire d'avoir les temps d'exécution des autres étapes du script. On reviendra sur toutes les performances par la suite (cf. paragraphe sur le Retour sur les performances du script).

### 2.3.2. Limites méthodologiques et perspectives d'approfondissement

Au niveau des données du RRP, il est possible que certaines erreurs soient présentes puisque les données sont rentrées manuellement. Des erreurs peuvent également être issues des mesures. Il n'est pas possible de gérer toutes les erreurs avec le script, mais certaines peuvent être détectées ou résolues. Pour le RRP 47, une strate a été remplie en double dans la base **STRATE\_QUANT**, il est possible de faire des vérifications à certaines étapes pour supprimer les doublons. Également, lorsque les variables littérales ne sont pas remplies conformément au format de la base DoneSol, il peut être possible de modifier le champ tant que les différences sont minimales, comme des majuscules ou des minuscules. Lorsque les erreurs n'ont pas été prévues, elles ne seront cependant pas gérées par le script.

Au niveau méthodologique, des choix ont évidemment été faits, que ce soit par les méthodes de calculs, comme le choix des fonctions de pédotransfert, ou par les seuils choisis, par dires d'expert ou par bibliographie. Ces choix sont justifiés mais restent des décisions arbitraires dans une certaine mesure. D'autres paramètres du RUM pourraient par exemple être utilisés, comme les remontées capillaires qui ont un impact sur le RUM de certains types de sols.

Une perspective d'amélioration serait d'utiliser d'autres méthodologies de calcul du RUM, d'autres fonctions ou classes de pédotransfert par exemple. L'objectif par la suite est notamment de comparer la méthode utilisée ici avec la fonction de pédotransfert de Roman-Dobarco récemment publiée, qui inclut des paramètres d'entrée continue : taux d'argile, en sables et en carbone organique (Roman-Dobarco, Cousin, Le Bas, & Martin, 2018).

Une autre possibilité serait aussi d'améliorer le script pour permettre de connaître le RU pour une culture donnée. Cela implique de prendre en compte l'enracinement optimal de la culture, soit l'épaisseur du sol sur laquelle les racines vont exploiter 100 % du réservoir disponible en eau.



### 3. Mise en forme cartographique des résultats

La mise en forme cartographique des résultats revêt une importance cruciale dans le script, car elle permet de présenter de manière visuellement compréhensible les données géospatiales relatives au RUM.

#### 3.1. Le script pour la représentation cartographique

##### 3.1.1. Agrégation de l'information à l'UCS pour la représentation cartographique

Comme vu plus tôt (cf. 2.1.2), l'information des RRP est stockée à l'échelle des strates et des UTS, mais l'unique support de la représentation cartographique dans la base de données DoneSol est l'UCS. Cette dernière est l'unité d'intégration spatiale.

Pour rappel, les UCS sont des objets que l'on sait localiser dans l'espace, composés d'une ou plusieurs UTS identifiées dont on connaît les proportions en % surfacique au sein de l'UCS. Pour des UCS simples (100% composées d'une seule UTS), la valeur du paramètre pédologique d'intérêt, ici le RUM, sera celle de l'UTS. Néanmoins dans le cas des UCS complexes, la valeur à l'UCS résultera de l'agrégation des valeurs renseignées pour les différentes UTS. Cette agrégation de l'information des UTS aux UCS peut être menée selon plusieurs méthodes.

##### 3.1.2. Différentes méthodes d'agrégation envisagées

Pour agréger l'information de l'UTS à l'UCS, 4 méthodes sont couramment utilisées :

- La représentation de la moyenne pondérée,
- La représentation de la valeur de l'UTS majoritaire,
- La représentation de la classe de valeurs majoritaire,
- La représentation du pourcentage de chaque classe de valeurs.

Ces méthodes sont expliquées en se basant sur la Figure 10, tirée du *Guide d'utilisation des bases de données sol pour la production de cartes thématiques* (Lemercier, et al., 2017), qui utilise la variable RU en démonstration. Ainsi, la Figure 10 montre : à gauche, une représentation schématique de polygones de sols ou UCS (UCS 1 à 3), qui sont les unités d'intégration spatiale d'une variable d'intérêt, ici le Réservoir Utilisable en eau des sols (RU) ; et à droite, une table de valeurs avec pour chaque UCS, le pourcentage (%) de chaque UTS, la valeur de RU en mm d'eau des UTS, recodée selon 5 classes de RU.

(Les classes de RU sont les suivantes : 0-40 mm, 40 - 80 mm, 80 -120 mm, 120 - 160 et > 160 mm)

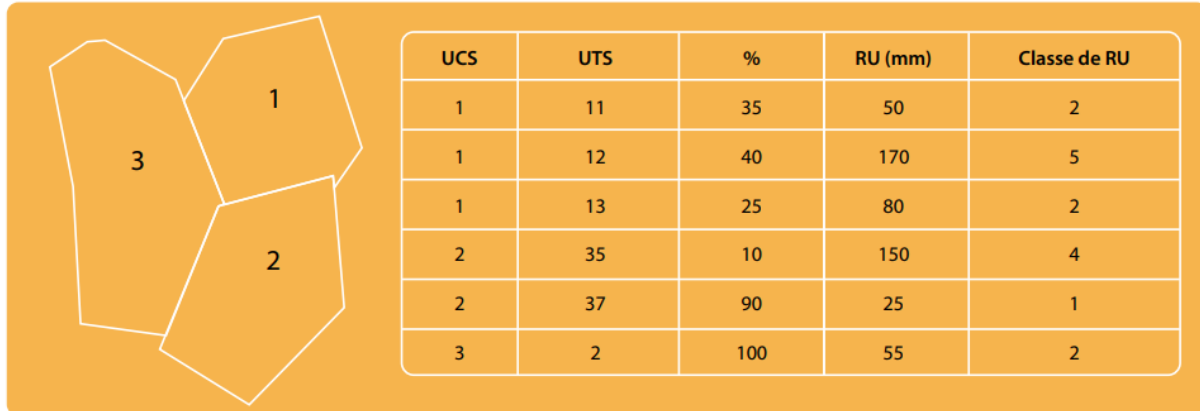


Figure 10 - Exemple de résultats du traitement de données de RU par UTS, à spatialiser par UCS (Lemercier, et al., 2017)

- **Agrégation par la moyenne pondérée** : il s'agit de faire la moyenne de la valeur de RU des UTS en pondérant par le pourcentage surfacique de chaque UTS dans l'UCS. Exemple avec la Figure 10, la moyenne pondérée de RU de l'UCS 1 sera :

$$\frac{35 * 50 + 40 * 170 + 25 * 80}{100} = 100,5 \text{ mm} \approx \mathbf{101 \text{ mm}}$$

Cette méthode renvoie une valeur réelle que l'on peut associer à la variabilité des valeurs contenues dans l'UCS par le calcul de l'écart-type par exemple. Le souci principal de cette méthode est le lissage de l'information de RU au niveau de l'UCS qui peut s'éloigner de la réalité.

- **Agrégation par la valeur de l'UTS majoritaire dans l'UCS** : la valeur retenue est celle de l'UTS majoritaire au sein de l'UCS. Exemple avec la Figure 10: le RU de l'UCS 1 est donc de **170 mm**. La valeur obtenue avec cette méthode est valable uniquement sur 40% de la surface de l'UCS 1. Cette information doit nécessairement accompagner la carte de RU des UCS pour notifier à l'utilisateur le niveau d'incertitude cartographique associé. Enfin, la légende associée à l'agrégation par la valeur de l'UTS majoritaire nécessitera de passer par une mise en classe des valeurs de RU des UTS majoritaires.
- **Agrégation par la classe majoritaire dans l'UCS** : il s'agit préalablement de mettre en classe les valeurs des UTS. Exemple de la Figure 10 : les valeurs de RU des UTS ont été recodées selon 5 classes. La classe majoritaire de RU dans l'UCS 1 est la classe 2 (RU entre 40 et 80 mm) et représente 60 % de l'UCS 1 (i.e. 35 % UTS n°11 + 25 % UTS n°13) (Figure 11).

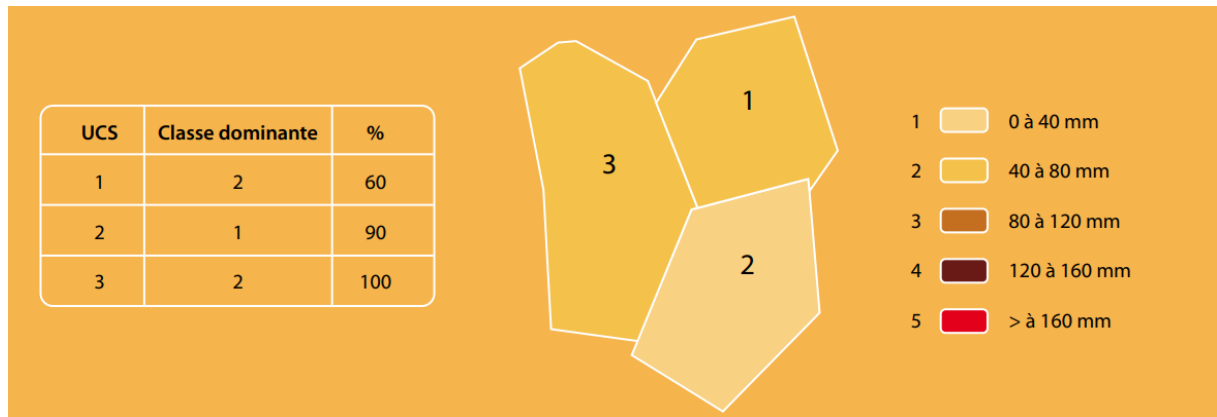


Figure 11 - Agrégation et spatialisation par UCS des valeurs des UTS selon la méthode de la classe majoritaire (Lemerrier, et al., 2017)

- **Représentation du pourcentage de chaque classe** : il s'agit de cartographier le pourcentage de représentativité de chaque classe de valeurs. Il y aura donc autant de cartes que de classes de valeurs. L'exemple renvoie donc 5 cartes avec le % de représentativité de chaque classe au sein des UCS. La valeur 0% a toute son importance ici, si une classe n'est pas présente dans une UCS, telle que la classe 3 absente pour les UCS 1, 2 ou 3.

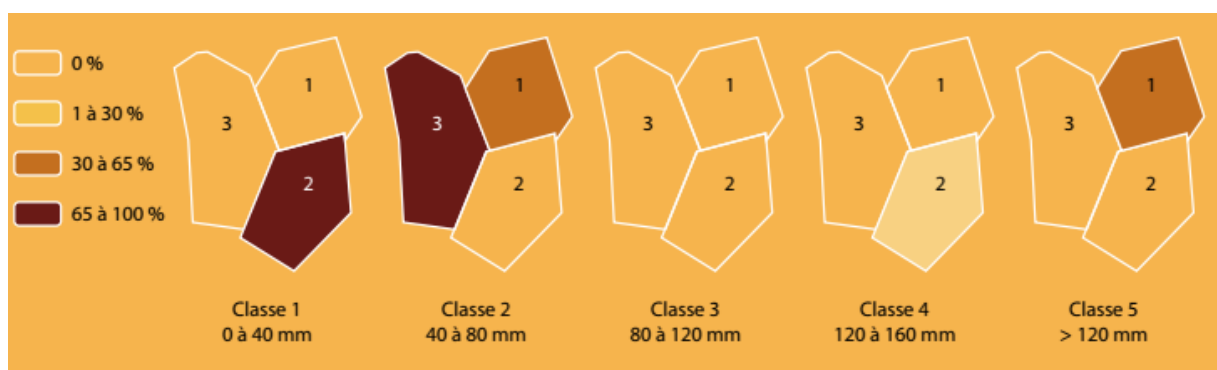


Figure 12 - Agrégation et spatialisation par UCS des valeurs de RU des UTS selon la méthode de représentation du pourcentage de chaque classe de RU (Lemerrier, et al., 2017)

#### Etape 6 - Agrégation cartographique à l'UCS

La méthode d'agrégation des valeurs de RUM retenue est celle de la **représentation de la classe majoritaire dans l'UCS**, avec 5 classes de RUM dont les seuils sont présentés en Figure 13. Avec cette méthode, il y a une perte d'information liée à la mise en classe, mais on respecte au maximum le contenu de l'UCS en gardant le niveau d'information que l'on restitue, la pureté cartographique. Idéalement il faut associer à la carte de la classe majoritaire une carte de pureté cartographique qui représente le pourcentage surfacique de la classe dominante de RU qui est cartographiée. Toutefois, des choix de représentation sont rajoutés pour se passer de cette seconde carte.



Figure 13 - Seuils des classes et légendes utilisés pour la représentation cartographique du RUM (Script, 2023)

Les choix de représentation sont liés à la variabilité de l'information représentée. Avec 5 classes de RUM, une des classes peut être dominante dans une UCS avec seulement 21% de sa composition. On fait le choix de ne cartographier uniquement les classes de RUM qui représentent **au minimum 50% de la surface de l'UCS**. Les UCS qui contiennent sur au moins 50 % de leur surface des UTS dont le RUM n'a pas pu être calculé (valeur RUM\_UTS = **-99**) auront automatiquement la classe **-99** associée, requalifiée en légende comme « **Données Incomplètes** ».

Dans le cas inverse, cela signifie que l'UCS présente une **variabilité trop importante** de classes de RUM, et on y attribue alors la valeur de **-1**.

Le Tableau 4 présente la répartition des UCS suite à l'agrégation choisie. On a tout de même 10 UCS sur 82 dont la variabilité est trop importante pour pouvoir être mis en classe.

Le nombre de classes de RUM ainsi que le choix de représenter en une seule carte le RUM et sa variabilité au sein des UCS sont inspirés des tests de représentations cartographiques menés lors du projet SERUM<sup>2</sup> *i.e.* Spatialisation et Expertises du Réservoir Utile Maximal en Grand-Est (Chambre Régionale d'Agriculture Grand-Est, 2023). Cela est dans l'optique de comparer des méthodes cartographiques entre utilisateurs de la donnée Sol et pouvoir améliorer de manière continue les outils d'extraction et de calcul de propriété des sols comme le RUM.

Tableau 4 - Répartition des UCS dans chaque classe pour le RRP 47 (Equipe Sols de Bordeaux Sciences Agro)

Classe de RUM	1	2	3	4	5	Variabilité trop importante (-1)	Données incomplètes (-99)
Nombre d'UCS	0	0	8	14	48	10	2

<sup>2</sup> Le projet SERUM Grand-Est est un projet lauréat de l'AAP IGCS 2020-2021, porté par la Chambre régionale d'agriculture Grand-Est. Il s'agissait de tester diverses manières de spatialiser le RUM des sols et son incertitude en appui d'un comité d'experts et d'utilisateurs pour évaluer ces méthodes de cartographie.

### 3.1.3. Représentation cartographique du RUM

Une fois que l'agrégation à l'UCS est réalisée, il reste alors à représenter l'information sous forme de carte.

#### *Etape 1 bis - Importation des bibliothèques nécessaires*

En plus de celles déjà présentées dans la première partie du script, quelques nouvelles bibliothèques sont utilisées pour la représentation cartographique (Tableau 5).

Tableau 5 - Bibliothèques Python importées dans la seconde partie du script

Bibliothèque	Description
<b>Geopandas</b>	Utilisée pour la lecture des fichiers en format <b>Geojson</b> , permettant ainsi de traiter des données géospatiales.
<b>Matplotlib</b>	Employée pour créer les cartes, qui sont en réalité des graphiques au sens strict du terme.
<b>Contextily</b>	Gère l'ajout de fonds de carte en associant des cartes tuilées obtenues depuis divers sites aux figures générées par <b>Matplotlib</b> .

#### *Etape 2 bis - Importation des données*

L'intégration finale de la classe de RUM est au niveau des Unités cartographiques des Sols. La couche géographique du RRP ou couche des UCS est intégrée dans le script sous le format **Geojson**. D'autres formats sont utilisables, tels que le **Shapefile** en **.shp** ou le **Geopackage** en **.gpx**. Le format Geojson a d'intéressant qu'il est disponible sur la plateforme PIGMA pour les RRP de la Nouvelle-Aquitaine, il est donc facilement récupérable et importable. C'est également une solution dont le temps d'importation reste raisonnable, on reviendra dessus par la suite.

Un fond de carte est également importé directement au moment de la réalisation de la carte avec la bibliothèque **Contextily**. La source est le fond de carte « Positron » fourni par Carto, un service de cartographie en ligne (Carto, 2023). C'est un fond de carte fréquemment utilisé pour créer des cartes de bases discrètes et sobre avec des données superposées.

#### *Etape 7 - Les choix de représentation cartographique*

Le lien est fait entre le fichier Geojson et les classes de RUM des UCS, ainsi que les autres informations voulues, grâce au champ « **NO\_UCS** ». Une fois cela fait, il ne reste qu'à créer la carte.

Pour légender le RUM, qui représente une ressource en eau, le choix du bleu est fait. Les 5 classes de RUM sont donc représentées par un dégradé de bleu. Les deux autres modalités du champ sont représentées par des couleurs qui dénotent du bleu, respectivement du magenta pour les données incomplètes, et du rouge pour les données de variabilité trop importante. La première étape du script est alors d'ajouter ces couleurs dans une liste.

La fonction «**.plot()**» de la bibliothèque **Matplotlib** est utilisée pour la représentation, en utilisant la classe de RUM et la liste des couleurs. C'est à ce moment que le fond de carte est également ajouté.

La carte est alors créée, il ne reste qu'à ajouter les éléments supplémentaires souhaités. La légende est placée à droite de la carte, avec une valeur de superficie pour chacune des classes. Un titre (Modalité dominante du RUM dans le 47), la source (Référentiel Régional Pédologique du département du Lot-et-Garonne), et une flèche indiquant le nord sont finalement ajoutés (Figure 14).

### Modalité dominante du RUM dans le 47

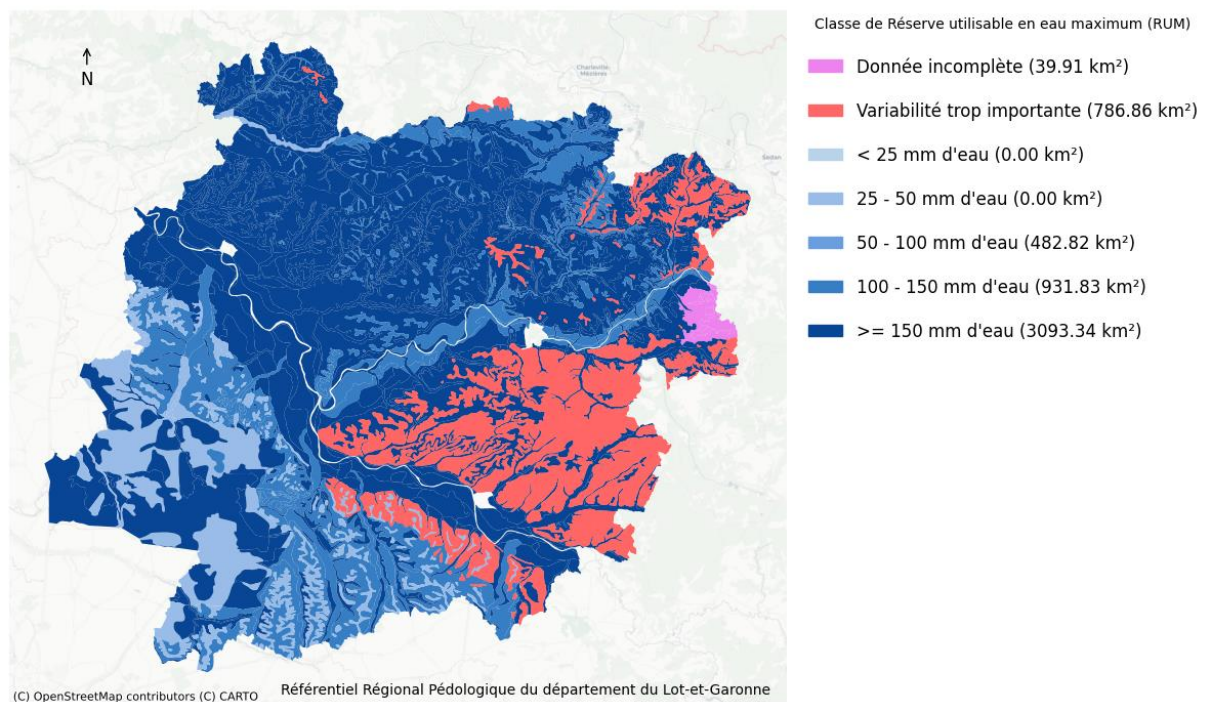


Figure 14- Rendu final de la carte obtenue avec le script (Script de valorisation cartographique - Version du 19/09/2023)

## 3.2. Limites et perspectives

### 3.2.1. Retour sur les performances du script

Maintenant que le script est terminé, il est intéressant de revenir sur la vitesse d'exécution des différentes étapes (Figure 8).

Tout d'abord, tous les essais ont été faits avec un ordinateur **Dell® OptiPlex 9020** (Dell, 2013), équipé d'un processeur Intel(R) Core (TM) i5-4570 CPU @ 3.20GHz, 3201 MHz, 4 cœurs, 4 processeurs logiques, avec 4,00 Go de RAM. L'utilisation de Google Colab passe cependant par des machines virtuelles, donc les caractéristiques de l'ordinateur ne font pas varier la vitesse d'exécution du script. La seule limite se trouve dans la connexion réseau, l'ordinateur doit être capable d'assurer une connexion à la machine virtuelle au moment où les données sont importées manuellement, où le script est modifié, et pour obtenir les résultats après exécution. Le script lorsqu'il est exécuté utilise la connexion réseau de la machine virtuelle, même pour



les imports internes (le fond de carte par exemple). Pour le RRP 47, les performances du script sont présentées dans le Tableau 6.

Tableau 6 - Temps d'exécution des différentes étapes du script pour le RRP 47

Etape dans le script	Temps approximatif	Taille du fichier pour le RRP47
<b>Importation du fichier Microsoft® Access</b>	Environ 45 secondes	9088 Ko
<b>Importation du fichier Geojson</b>	Environ 3 minutes	36 481 Ko
<b>Importation des bibliothèques nécessaires</b>	Environ 20 secondes	
<b>Importation et préparation des données du RRP</b>	Environ 7 secondes	
<b>Calcul des indicateurs des propriétés du sol</b>	Moins de 1 seconde	
<b>Calcul des valeurs de RUM des strates</b>	Moins de 1 seconde	
<b>Passage du RUM strates au RUM UTS</b>	Moins de 1 seconde	
<b>Agrégation cartographique à l'UCS</b>	Moins de 1 seconde	
<b>Représentation cartographique</b>	Environ 45 secondes	

Également, le temps d'import des bibliothèques représente 20 secondes à la première exécution, mais est presque instantané par la suite. Cela est dû au fait que les bibliothèques sont installées à la première itération du script sur la machine virtuelle de Google Colab, ce qui représente la majeure partie du temps lié à leur importation.

Le temps lié à la représentation cartographique quant à lui est dépendant du nombre de cartes souhaitées. Pour chaque carte supplémentaire générée, ce temps augmentera également, et si un grand nombre de cartes est généré, il pourrait devenir majoritaire dans le script. Vingt secondes par carte supplémentaire environ sont à prévoir.

On peut donc voir que plus de la moitié du temps du script est celui de l'importation du fichier Geojson. Il est envisageable de changer de format de fichier, un fichier Geopackage est plus rapide à être importé et lu, mais celui-ci n'a pas l'avantage d'être disponible sur l'IDG PIGMA dans cet exemple. C'est toutefois une possibilité d'amélioration de l'efficacité du script. Le temps global atteint les alentours de **5 minutes** cependant, ce qui reste un temps très raisonnable à l'échelle du nombre de RRP en France (Tableau 6).

Le temps équivalent pour un utilisateur expérimenté pour réaliser le même travail peut aller de deux à cinq jours. Pour un utilisateur non expérimenté, en comptant le temps de formation et le calcul, plusieurs semaines sont nécessaires, en plus de la nécessité d'un retour expert pédagogue.

### **3.2.2. Révision du script en cours**

Certaines améliorations restent à faire sur les représentations cartographiques de la version actuelle du script. La légende doit être hiérarchisée autrement, d'abord les classes de RUM et après la variabilité, puis les données incomplètes (Figure 13). Les valeurs de superficie sont également erronées pour le moment.

Pour la carte, elle serait également amenée à la modification. Certains bugs n'ont pas encore été résolus, ce qui fait que la carte de la Figure 14 est fautive pour le moment.

Le fond de carte est également erroné, la projection spatiale ne place pas le fond de carte correspondant au département du Lot-et-Garonne sous la carte, mais celui d'une zone autour des Ardennes.

Après ces améliorations faites, le script sera testé avec plusieurs RRP. Le script comme ses résultats seront soumis à un groupe d'utilisateurs RRP et Python, issu de l'Inrae Info&Sols et l'Institut Agro Rennes. Une fois validé par ce comité, il sera mis en ligne en OpenData, sur GitHub (GitHub, 2023) ou une plateforme similaire.

### **3.2.3. Devenir des cartes, mise en ligne sur un Mviewer**

Grâce à cet outil numérique, de nombreuses cartes peuvent être créées, même si l'exemple présenté ici se limite au RUM des sols. L'information, pour être accessible, a besoin de visibilité et l'impact est d'autant plus fort auprès des acteurs du territoire, si l'on démontre l'usage des données.

Les cartes générées par le script pourraient être partagées à travers un visualiseur à l'instar de « Sols de Bretagne » ou « Sols d'Alsace », développés sous l'application Mviewer. Les codes sources de ces Mviewer étant libre d'accès, une de mes missions en parallèle au cours de ce stage a été d'entamer le développement d'une interface simplifiée de Mviewer et d'en comprendre les rouages pour communiquer auprès d'autres gestionnaires RRP la démarche à suivre pour créer leur propre Mviewer.

Il est prévu que le Mviewer « Sols de Nouvelle-Aquitaine » dans sa première version simplifiée soit mis en production sur PIGMA courant 2024.



## Conclusion

Le projet Solid'R s'inscrit dans le programme IGCS du Gis Sol, qui se concentre sur la cartographie des sols à différentes échelles. Il vise à harmoniser les métadonnées des Référentiels Régionaux Pédologiques (RRP). Il a également pour objectif de développer des outils numériques pour la valorisation des données pédologiques, objectif d'où provient notre problématique : Comment concevoir et utiliser les outils numériques pour valoriser des données pédologiques et en quoi ces méthodes représentent-elles un avantage ?

L'état des lieux des RRP a montré qu'une grande majorité était terminée, mais que seulement une fraction d'entre eux étaient diffusés efficacement, la valorisation de ces données est un défi d'actualité.

L'objectif du stage était alors de reprendre et d'améliorer un script Python sous l'environnement de travail Google Colab pour automatiser le calcul du Réservoir en eau Utilisable Maximal des sols (RUM) à partir du RRP du Lot-et-Garonne. Le script est structuré en 7 étapes, de l'importation des bibliothèques de programmation nécessaires à la cartographie des résultats, en passant par de la correction de données et de l'agrégation à différentes échelles pédologiques.

La durée d'exécution d'un tel script est de cinq minutes, contre plusieurs jours pour réaliser le travail manuellement, la différence de temps est considérable. Toutes les étapes ont suivi une méthodologie précise, et le script permet à chacune de ces étapes de récupérer toute l'information sous forme de tableau en .csv. C'est un aspect très important car la valorisation des données du script peut être faite autrement que par la cartographie qu'il réalise déjà : les propriétés calculées peuvent servir à d'autres travaux. Une autre des grandes forces de ce script est sa facilité de partage et d'utilisation.

Cependant, nous avons également souligné les défis persistants, notamment sur le format des données et les décisions méthodologiques. Cet outil est amené à être encore amélioré, avant la mise en ligne des résultats en OpenData, et la création d'un visualiseur pour une diffusion plus large. De tels outils peuvent fortement contribuer à une meilleure compréhension des sols en France, facilitant ainsi leur exploitation et leur protection, grâce à un gain de temps et aidant de fait à répondre aux enjeux de gestion durable des sols au sein des territoires.

## Bibliographie

- Arrouays, D., Hardy, R., Schnebelen, N., Le Bas, C., Eimberck, M., Roque, J., . . . Stengel, P. (2004). Le programme Inventaire Gestion et Conservation des Sols de France. *Etude et Gestion des Sols*, 11(3), 187-197. Récupéré sur <https://hal.inrae.fr/hal-02682146v1>
- Bispo, A., Laroche, B., Fort, J., Jalabert, S., Lemercier, B., Sauter, J., & Scheurer, O. (2020, 11). *Le tour de France des maîtres d'ouvrage IGCS - Essai de diagnostic et de prospective sur la gestion des donées Sol*. Récupéré sur RMT Sols et Territoires: [https://sols-et-territoires.org/fileadmin/user\\_upload/documents/publications-marquantes/Resume\\_\\_tour\\_regions\\_VDEF.pdf](https://sols-et-territoires.org/fileadmin/user_upload/documents/publications-marquantes/Resume__tour_regions_VDEF.pdf)
- Bouthier, A., & al. (2022). *Réservoir en eau du sol utilisable par les cultures*. Arvalis.
- Bruand, A., Duval, O., & Cousin, I. (2004). Estimation des propriétés de rétention en eau des sols à partir de la base de données SOLHYDRO : Une première proposition combianant le type d'horizon, sa texture et sa densité apparente. Dans *Etude et Gestion des Sols* (pp. 323-334).
- Carto. (2023). Consulté le 09 21, 2023, sur <https://carto.com/>
- Chambre Régionale d'Agriculture Grand-Est. (2023). *Le projet SERUM*. Récupéré sur RMT Sols et Territoires: <https://sols-et-territoires.org/projets/serum>
- Dell. (2013). *Dell OptiPlex 9020 - Installation et caractéristiques*. Consulté le 09 20, 2023, sur [https://dl.dell.com/manuals/all-products/esuprt\\_desktop/esuprt\\_optiplex\\_desktop/optiplex-9020-desktop\\_setup%20guide\\_fr-fr.pdf](https://dl.dell.com/manuals/all-products/esuprt_desktop/esuprt_optiplex_desktop/optiplex-9020-desktop_setup%20guide_fr-fr.pdf)
- Doussan, C. (2017, 05 19). RUE des Sols. *La Réserve Utile : Concepts, controverses, outils*. INRA.
- Equipe Sols de Bordeaux Sciences Agro. (s.d.). Script de valorisation cartographique - Version du 19/09/2023.
- GéoSas.fr, Institut Agro Rennes & Agath.a. (2023). *Les sols de Bretagne à la carte*. Récupéré sur Cartes des Sols de Bretagne: <https://geosas.fr/solsdebretagne/>
- Gis Sol. (2023, 09 18). *Base De Données Eléments Traces Métalliques – BDETM*. Récupéré sur Gis Sol - Partageons la connaissance des sols: <https://www.gissol.fr/le-gis/programmes/base-de-donnees-elements-traces-metalliques-bdetm-65>
- Gis Sol. (2023). *Gis Sol - Présentation*. Consulté le 08 04, 2023, sur Gis Sol - Partageons la connaissance des sols: <https://www.gissol.fr/le-gis>
- GitHub. (2023). Consulté le 21 09, 2023, sur <https://github.com/>
- Google. (2023, 09 21). *FAQ Colaboratory* . Consulté le 09 21, 2023, sur Colaboratory: <https://research.google.com/colaboratory/faq.html?hl=fr>

- Grolleau, É. &. (2004). *Le système d'information national sur les sols : DONESOL et les outils associés*. Récupéré sur [https://www.researchgate.net/publication/282171076\\_Le\\_systeme\\_d'information\\_national\\_sur\\_les\\_sols\\_DONESOL\\_et\\_les\\_outils\\_associes](https://www.researchgate.net/publication/282171076_Le_systeme_d'information_national_sur_les_sols_DONESOL_et_les_outils_associes)
- Grolleau, E., Bargeot, L., Chafchafi, A., Hardy, R., Doux, J., Beaudou, A., . . . Arrouays, D. (2004). Le système d'information national sur les données pédologiques spatialisées: DONESOL et les outils associés. *Etude et Gestion des Sols*, 11(3), 255-269.
- Inrae. (2021, 12 06). *20 ans d'inventaire et de surveillance des sols : résultats et enjeux des travaux du GIS Sol*. Consulté le 09 18, 2023, sur INRAE: <https://www.inrae.fr/actualites/20-ans-dinventaire-surveillance-sols-resultats-enjeux-travaux-du-gis-sol>
- Laroche, B., Richer de Forges, A., & Jalabert, S. (2023). À la découverte des sols : panorama des données, méthodes et outils existants. *GeoDataDays 2023*. Reims. Récupéré sur <https://hal.inrae.fr/hal-04208253>
- Laroche, B., Sauter, J., & Schnebelen, N. (2013, 01). *Inventaire, gestion et conservation des sols*. Récupéré sur Gis Sol - Partageons la connaissance des sols: [https://www.gissol.fr/fiches\\_pdf/fiches\\_IGCS\\_2013.pdf](https://www.gissol.fr/fiches_pdf/fiches_IGCS_2013.pdf)
- Laroche, B., Thiry, E. S., Chenu, J., Moulin, J., & Bardy, M. (2014). Les bases de données sur les sols au service des politiques publiques : exemple des zones humides. *Etude et Gestion des Sols*, 21, 37-49. Récupéré sur <https://hal.inrae.fr/hal-02637382/document>
- Legros, J.-P. (1996). *CARTOGRAPHIE DES SOLS. De l'analyse spatiale à la gestion des territoires*. France: Presses Polytechniques et Universitaires Romandes. doi:ISBN : 2-88074-298-6
- Lemercier, B., Laroche, B., Armand, R., Chafchafi, A., Détriché, S., Ducommun, C., . . . Lehmann, S. (2017). *Guide d'utilisation des bases de données sol pour la production de cartes thématiques*. INRA.
- Messant, A., Lehmann, S., Moulin, J., Lagacherie, P., Jalabert, S., Noraz, A., . . . Sauter, J. (2021). Diffusion des référentiels régionaux pédologiques sous la forme d'une carte des sols dominants (France métropolitaine - Hors Corse) accessible sur le Géoportail. *Etude et Gestion des Sols*, 28, 57-69.
- Python. (s.d.). *Python*. Consulté le 09 21, 2023, sur Python: <https://www.python.org/>
- Richer-De-Forges, A., Feller, C., Jamagne, M., & Arrouays, D. (2008). Perdus dans le triangle des textures. *Etude et Gestion des Sols*.
- RMT Sols & Territoires. (2023, 09 18). *Axes du RMT*. Consulté le 09 18, 2023, sur RMT Sols et Territoires: <https://sols-et-territoires.org/axes-de-travail>

RMT Sols et Territoires. (2023). *Diffusion des Référentiels Régionaux Pédologiques - Le projet SOLID'R pour SOutenir La dIffusion des Référentiels régionaux pédologiques*. Consulté le 09 21, 2023, sur RMT Sols et Territoires: <https://sols-et-territoires.org/projets/soutien-a-la-diffusion-des-referentiels-regionaux-pedologiques>

Roman-Dobarco, M., Cousin, I., Le Bas, C., & Martin, M. P. (2018). Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. *Geoderma*.

*Sols de Bretagne*. (2015). Récupéré sur Sols de Bretagne: <https://sols-de-bretagne.fr/>

Tetegan, M. (2011). *Modélisation des propriétés de rétention en eau des sols caillouteux : application à l'estimation spatialisée de la réserve utile*. Orléans: Ecole doctorale Sciences et technologies.

## Annexes

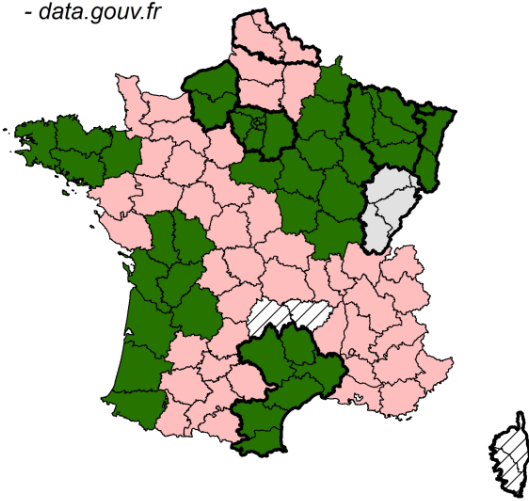
Annexe 1 - Recensement de fiches de métadonnées liées aux RRP recensées sur Internet (d'après le travail d'inventaire de Jalabert & Le Bas, 2022)

### IDG nationales

**Les métadonnées liées aux RRP sont-elles cataloguées sur au moins une IDG nationale ?**

Les IDG nationales sur lesquelles le recensement a été réalisé sont :

- Géocatalogue,
- AgroEnvGéo et
- data.gouv.fr



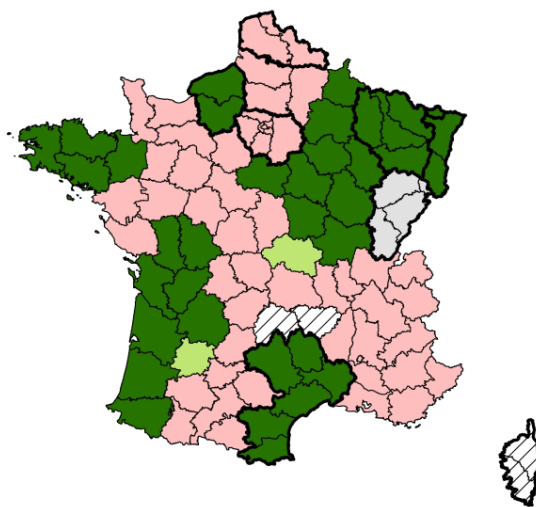
**Métadonnées RRP cataloguées sur au moins une IDG nationale ?**

- oui
- non
- non, car RRP non expertisé, non validé
- non, car RRP non finalisé

RRP réalisé à une échelle régionale (selon les limites des anciennes régions)

### IDG régionales

**Les métadonnées liées aux RRP sont-elles cataloguées sur une IDG régionale ?**



**Métadonnées RRP cataloguées sur une IDG régionale ?**

- oui
- oui, mais fiche non moissonnée sur le Géocatalogue
- non
- non, car RRP non expertisé, non validé
- non, car RRP non finalisé

RRP réalisé à une échelle régionale (selon les limites des anciennes régions)

Annexe 2 - Lien de l'organigramme de décision de calcul des propriétés des sols

<https://www.mindomo.com/mindmap/5a19579ff7ea4a04ac60cc8df49ca772>