

Analyse des besoins

et des contraintes

pour la réalisation du  
Cloud Environnemental au Bénéfice de l’Agriculture (CEBA)

Challenge 1 – CAP 20-25

SOMMAIRE

[1 Introduction 5](#_Toc511053498)

[2 Identification des besoins 5](#_Toc511053499)

[2.1 Acteurs 5](#_Toc511053500)

[2.1.1 Définition 5](#_Toc511053501)

[2.1.2 Rôles 5](#_Toc511053502)

[Consommateurs 6](#_Toc511053503)

[Producteurs 6](#_Toc511053504)

[2.2 Sources des données 8](#_Toc511053505)

[2.2.1 Réseaux déjà existants et structurés 8](#_Toc511053506)

[2.2.2 Fichiers type Excel ou CSV 8](#_Toc511053507)

[2.2.3 Flux de réseaux de capteurs 8](#_Toc511053508)

[2.2.4 Fichiers volumineux 8](#_Toc511053509)

[2.3 Usages 9](#_Toc511053510)

[2.3.1 Stockage 9](#_Toc511053511)

[2.3.2 Partage 9](#_Toc511053512)

[2.3.3 Exploitation 10](#_Toc511053513)

[2.4 Conclusion 10](#_Toc511053514)

[3 Contraintes & opportunités 11](#_Toc511053515)

[3.1 Opportunités institutionnelles 11](#_Toc511053516)

[3.1.1 Utilisation des ressources du Mésocentre Clermont-Auvergne 11](#_Toc511053517)

[3.2 Opportunités techniques 11](#_Toc511053518)

[3.2.1 Mise à disposition des données des instruments ConnecSens 11](#_Toc511053519)

[3.2.2 Interopérabilité avec le LIT Grandes Cultures en Auvergne 11](#_Toc511053520)

[3.2.3 Conformité avec les standards internationaux 11](#_Toc511053521)

[3.3 Contraintes légales 11](#_Toc511053522)

[3.3.1 Nécessité de gestion de la propriété intellectuelle des données 11](#_Toc511053523)

[3.3.2 Respect de la réglementation sur le financement des ressources 11](#_Toc511053524)

[4 Synthèse 12](#_Toc511053525)

[4.1 Synthèse Globale 13](#_Toc511053526)

[4.2 Synthèse Globale Simplifiée 14](#_Toc511053527)

[5 Livrables et jalons 15](#_Toc511053528)

[5.1 Validation de l’analyse des besoins 15](#_Toc511053529)

[5.2 Livraison du cahier des charges 15](#_Toc511053530)

[5.3 Validation du cahier des charges 15](#_Toc511053531)

[5.4 Prototype 15](#_Toc511053532)

[6 Révisions 16](#_Toc511053533)

[7 Annexes 17](#_Toc511053534)

Liste des figures

[figure 1 : Diagramme de répartitions des rôles des Acteurs du CEBA 6](#_Toc509862916)

[figure 2 : Diagramme des sources de données du CEBA 8](#_Toc509862917)

[figure 3 : Diagramme radar des principaux usages du CEBA 10](#_Toc509862918)

Lexique et définitions

CEBA : Cloud environnemental au bénéfice de l’agriculture.

DOI : Digital Object Identifier, sert à identifier une donnée de façon unique et pérenne.

IRI : International Ressource Identifier, généralise l’URI en acceptant les caractères codés en UTF-8.

IGN : Institut géographique national.

Insee : Institut national de la statistique et des études économiques.

LIT : Laboratoire d’Innovation Territorial grandes cultures en Auvergne.

Open Data : Concept de partage des données, de manière libre et gratuite.

RDF : Resource Description Framework est un modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées, de façon à permettre le traitement automatique de telles descriptions.

SGBD : Système de Gestion de Base de Données.

SI : Système d’Information, ensemble des ressources permettant à l’information de circuler au sein de l’entité considérée.

SPARQL : SPARQL Protocol and RDF Query Language (prononcé sparkle en anglais : « étincelle ») est un langage de requête et un protocole qui permet de rechercher, d'ajouter, de modifier ou de supprimer des données RDF disponibles à travers Internet.

Triplestore : Un triplestore est une base de données spécialement conçue pour le stockage et la récupération de données RDF.

Triplet RDF : Un triplet RDF est l'unité de données la plus petite contenue dans un graphe RDF au sein d'une base de données de type triplestore (ex. :(sujet, prédicat, objet)).

URI : l’Uniform Resource Identifier doit permettre d'identifier une ressource de manière permanente, même si la ressource est déplacée ou supprimée (exemple : URL, ISBN, …).

URL : Uniform Resource Locator, est un type d’URI, permettant de localiser une ressource sur un réseau ex : cipiere.fr

URN : Uniform Resource Name, est un type d’URI, comme le numéro ISBN qui permet de catégoriser une publication.

# Introduction

Ce document présente une analyse des besoins et des contraintes exprimés par les usagers du Cloud Environnemental au Bénéfice de l’Agriculture (CEBA), observatoire numérique de l’environnement (eau, sol, air, biodiversité) et des agroécosystèmes à l’échelle du territoire auvergnat. L’objectif du CEBA est de rassembler tous les indicateurs environnementaux pertinents pour la surveillance des agroécosystèmes et de construire des modèles pour prédire leur évolution et leur impact sur les territoires. Lieu de rencontre entre monde académique, monde agricole et acteurs institutionnels, le CEBA permettra d’évaluer le rôle des facteurs environnementaux sur le rendement et la qualité des produits agricoles et l’impact des systèmes agricoles sur l’évolution et la qualité de l’environnement.

L’analyse des besoins et contraintes du CEBA a été construire sur la base des informations obtenues grâce à des interviews conduites par Sébastien Cipière au mois de février et mars 2018 auprès de 19 personnes impliquées dans le CEBA. Pour rappel,   
les partenaires du CEBA sont : LaMP, GEOLAB, LMGE, ICCF, PIAF, LMV, UREP, GDEC, LPC, LIMOS, TSCF, Centre Michel de l'Hospital, Unité Expérimentale Herbipôle, Fédération des Recherches en Environnement, Cluster d’excellence E2IA, Weather Measures, Céréales Vallée.

La synthèse des interviews est disponible dans le document de travail intitulé « compte-rendu des interviews ». ce document est consultable par les membres du comité de pilotage du CEBA.

# Identification des besoins

Le cloud environnemental au bénéfice de l’agriculture (CEBA) a pour but de fournir un observatoire numérique de l’environnement et des agroécosystèmes (eau, sol, air, biodiversité) à l’échelle du territoire auvergnat. Nous définissions ici les Acteurs du CEBA comme étant des personnes physiques ou morales, ainsi que des groupes de travail.

## Acteurs

### Définition

La notion d’Acteurs sera concrétisée au sein du CEBA par une identité numérique protégée par un couple login et mot de passe unique permettant de l’identifier au sein du système d’information (SI). L’identification de l’Acteur permettra une gestion des droits d’accès à la donnée ainsi qu’une traçabilité des actions effectuées. Il existe une catégorie d’usagers qui ne nécessite pas d’authentification pour accéder à certaines informations contenues dans le cloud. Cette exception est liée aux données Open Data qui sont rendues accessibles au plus grand nombre librement.

### Rôles

Les Acteurs du CEBA se verront affectés des droits d’accès, d’ajout, de modifications et de gestions de la donnée, en fonction de leurs accréditations, et du rôle qu’ils occupent au sein du CEBA. Ces rôles pourront s’additionner sur un compte en fonction des responsabilités de la personne.

#### Consommateurs

Les consommateurs sont des Acteurs qui consultent des données. Les acteurs ne nécessitant pas d’authentification et souhaitant accéder aux données labélisées Open Data, ainsi que certains des acteurs authentifiés, dont le but est uniquement de consulter les informations présentes sur le cloud environnemental, sont classés dans la catégorie des consommateurs. On pourra donc retrouver dans la catégorie consommateur : des agriculteurs, des collectivités territoriales, des sociétés, des chercheurs, etc.

#### Producteurs

Les producteurs sont des Acteurs qui seront nécessairement identifiés pour pouvoir déposer leurs données à l’intérieur du cloud. Lors du dépôt des données en vue de la mise à disposition des informations auprès des autres Acteurs, il leur sera demandé :

* + De saisir les métadonnées relatives à l’information déposées
    - Type de la donnée
    - Le protocole suivi pour obtenir la donnée
    - Standard suivi / niveau de qualité de la donnée
    - Géolocalisation
    - Description / DOI / licence / etc.
  + Le périmètre dans lequel ils souhaitent partager leurs données (ex. privé, groupe, ou public)

Dans le cas du producteur automate, ces informations pourront être définies de façon standardisée.

Aucune contrainte autre que la fourniture des métadonnées précédentes ne sera imposée pour l’intégration d’une donnée (notamment le format et la structure).

1. Rôles des acteurs interviewés pour préparer l’analyse des besoins et des contraintes

La figure 1 présente la répartition des rôles des acteurs interviewés pour préparer l’analyse des besoins et des contraintes. Elle met clairement en évidence le fait que de nombreux acteurs sont à la fois producteurs et consommateurs des données du CEBA.

## Sources des données

Nous avons catalogué les différentes sources de données qui peuvent être partagées dans le CEBA. Les données peuvent être disponibles de diverses façons. Chaque méthode dispose d’avantages et d’inconvénients qui lui sont propres. La figure 2 ci-dessous présente les 4 sources de données identifiées au cours des interviews et que nous allons détailler : les réseaux déjà existants, les fichiers structurés, les flux de données issues de réseaux de capteurs et les fichiers volumineux.

### Réseaux déjà existants et structurés

Certains Acteurs remontent actuellement leurs données vers des réseaux comme des infrastructures de recherche européennes comme ICOS ou ACTRIS ou des établissements publics comme les agences de l’eau, etc. Il sera intéressant d’identifier les réseaux qui peuvent mettre à disposition du CEBA les données déjà fournies par ces acteurs pour éviter des doublons de procédure. Nous pouvons ici mentionner les observatoires ou les structures comme le BRGM dont les méthodes sont très rodées et qui n’attendent pas de données en retour du CEBA.

### Fichiers type Excel ou CSV

Pour des résultats d’analyses ou des mesures réalisées par un acteur, les données sont disponibles dans des fichiers, sur un ordinateur de travail. La majorité de ces fichiers sont des tableurs de différents types (Excel, CSV, etc.). Pour l’instant lorsque ces fichiers nécessitent un partage, il s’effectue par clef USB ou email. Les propriétaires de ces fichiers sont très motivés pour prendre part à l’initiative CEBA pour pouvoir plus facilement échanger leurs données. Ils sont de plus fortement intéressés à ce que le CEBA puisse améliorer la traçabilité de leurs jeux de données par l’application d’un DOI.

### Flux de réseaux de capteurs

Le déploiement de réseaux de capteurs pour la surveillance d’agroécosystèmes, notamment dans le contexte du projet de site ConnecSens génère le besoin de stocker, structurer, pérenniser, et rendre disponible les données collectées par les capteurs de différentes natures (piézomètre, humidité du sol, dendromètre, station météorologique, etc.). D’autres acteurs n’utilisent pas les mêmes technologies sans fil que celles utilisées par ConnecSens. Des Acteurs comme le LAMP nécessitent des données en temps réel (T+30 secondes) pour certains de leurs pluviomètres (notamment durant les événements météorologiques intenses). D’autres peuvent récupérer à J+1 le relevé au quart d’heure, qui n’est envoyé qu’une fois par jour.

### Fichiers volumineux

Une approche de certains Acteurs nécessite le partage de fichiers volumineux de type voix, vidéos, imagerie (scan, etc.). Pour répondre à leurs besoins le CEBA devrait mettre à disposition des services de types Gestion Electronique de Documents GED ou drive (répertoires partagés).

1. Diagramme des sources de données du CEBA

## Usages

Au travers des différents interviews réalisés en février et mars 2018 certains usages ont été cités de façon récurrente (cf. annexes). Ils sont classés ici en trois catégories : besoins de stockage, de partage et d’exploitation des données.

### Stockage

Si certains Acteurs mettront leurs données à disposition au travers de services qui ne nécessitent pas le stockage des données au sein du CEBA, d’autres Acteurs souhaitent héberger leurs données sur les serveurs du CEBA. Un cas extrême de cette approche est celui des herbiers universitaires qui nécessiterait une volumétrie d’une centaine de Téra octets. A titre de comparaison la totalité des autres usages à moyen terme ne devrait pas dépasser le Téra octets (1 000 Go). Le besoin de pérennisation du CEBA a été exprimé afin de garantir un accès aux données à long terme.

Les usages de stockage identifiés sont les suivants :

- Stocker des flux : il s’agit notamment de stocker les données collectées par des réseaux de capteurs ou des sites de mesure

- Stocker de la cartographie

- Stocker des fichiers très volumineux

- Stocker des fichiers semi-structurés (Excel, textes)

### Partage

De nombreux producteurs souhaitent partager leurs données avec les autres usagers. Les différents cas exprimés de partage sont les suivants :

* Partage de données normalisées
* Partage de données labélisées : besoin de définir et labéliser la qualité de la donnée
* Partage de données formatées : changer le format
* Partage contrôlé des données : garder la donnée pour moi ou mes amis
* Partage ouvert des données : mettre mes données à la disposition de tous
* Accès à des données internes au cloud et externes
* Mise à disposition de données traçables en termes de paternité et d’utilisation de la donnée
* Mise à disposition de données historiques

Pour refléter les modes de fonctionnement naturels, les différents acteurs pourront former des groupes de travail et des sous-groupes avec des responsabilités différentes. On peut ainsi envisager qu’un groupe de saisie ne puisse pas rendre la donnée accessible sans qu’elle soit validée par sa hiérarchie.

### Exploitation

Nous avons vu que la donnée sera mise à disposition dans le CEBA sous forme de flux ou de fichiers. Une demande récurrente exprimée par les personnes interviewées est la possibilité de situer et visualiser les informations à travers une vue cartographique. Cette vue doit éventuellement être sur-couchée avec des données géographiques. Pour faciliter le partage des informations les données pourront être restructurées automatiquement pour les rendre interopérable avec d’autres services. Les différents cas exprimés d’exploitation sont les suivants :

* Découvrir des données, des producteurs ou des consommateurs de données
* Télécharger et visualiser les données de tout type (données géoréférencées, spatialisées et cartographiques)
* Traiter les données : traitements standards et à la carte, avec une possibilité d’automatisation.
* Produire des données prévisionnelles à l’aide de modèles

Un guichet unique équipé d’un moteur de recherche multicritères (thématique, géographique, temporel, fournisseur…) permettra de faciliter l’exploitation des données.

## Conclusion

La figure 3 présente un diagramme radar résumant les principaux usages tels qu’ils ont été exprimés par les personnes interviewées.

1. Diagramme radar des principaux usages du CEBA exprimés par les acteurs interviewés (réponses multiples)

# Contraintes & opportunités

## Opportunités institutionnelles

### Utilisation des ressources du Mésocentre Clermont-Auvergne

De par la nature du projet qui est l’un des livrables du challenge 1 de CAP 20-25 les ressources informatiques qui permettront la mise en place du CEBA seront localisées au Mésocentre Clermont-Auvergne.

## Opportunités techniques

### Mise à disposition des données des instruments ConnecSens

Le projet ConnecSens est identifié comme une source privilégiée de données issues de réseaux de capteurs dans l’environnement. Ces données en Open Data ainsi que les ressources informatiques acquises au titre du projet font partie des fondations solides du CEBA.

### Interopérabilité avec le LIT Grandes Cultures en Auvergne

Le LIT est un espace d’accueil et d’émergence de projets visant à concevoir des solutions innovantes inspirées des principes de l’agroécologie dans une démarche de co-création avec les agriculteurs, en lien avec les autres acteurs du territoire Limagne-Val d’Allier. La mise en place d’une infrastructure numérique agro-environnementale performante et attractive est stratégique pour le LIT pour favoriser l’émergence de projets porteurs de solutions innovantes sur ce territoire. Il sera donc nécessaire de définir les conditions juridiques et financières d’interaction entre le LIT et le CEBA, ainsi que les modalités d’interopérabilité entre les bases de données du CEBA et du LIT. Les projets du LIT pourront en effet s’appuyer notamment sur des données technico-économiques collectées/stockées/gérées directement par le LIT et d’intérêt pour les acteurs du CEBA.

### Conformité avec les standards internationaux

Pour faciliter les échanges entre les différents Acteurs il est souhaitable d’être en conformité avec les différents standards internationaux que ce soit au niveau des outils ou des infrastructures. Cette bonne pratique permettra une meilleure visibilité du projet en permettant d’être interopérable avec des infrastructures nationales et internationales. Par contre, il ne sera pas demandé aux fournisseurs de données que celles-ci soient mises à disposition à la condition de respecter des normes internationales comme celles de la directive INSPIRE.

## Contraintes légales

### Nécessité de gestion de la propriété intellectuelle des données

Différentes conventions nous lieront avec nos Acteurs et il est donc primordial de pouvoir leur garantir que nous ne trahirons pas le périmètre de distribution des données qu’ils nous confient.

### Respect de la réglementation sur le financement des ressources

Certains financements publics sont accompagnés de limites dans l’usage à des fins commerciales. C’est le cas notamment du réseau RENATER et des ressources acquises avec des fonds structurels européens (FEDER).

# Synthèse

Ce travail se base sur les interviews d’une partie des Acteurs (19) du CEBA. Nous avons agrégé dans les tableaux présentés en 4.1 et 4.2 un résumé de l’ensemble de leurs réponses. Le tableau 4.1 Synthèse Globale, présente les éléments sans doublons et regroupés par thèmes et concepts. Le tableau 4.2 Synthèse Globale Simplifiée, est une vue simplifiée du tableau 4.1 qui permet de gagner en lisibilité en factorisant les différents éléments de détails derrière des mots et concepts clefs.

## Synthèse Globale

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rôle | Données | Politique sur la donnée | Usages | Attentes |
| Producteur/  Consommateur  Producteur  Consommateur | Fond de cartes  Orthophotoplan  Lidar, 400 jeux de données  Carte au 1/250 000  Cadastre  Données des sous-sols  Qualité des sols  Hydrogéologie  Qualité de l’eau  Analyse eau/sédiment  Risques  Capteurs ConnecSens  Flux de données  Station météo / Météo  Satellite  Données ICOS  Bio géo chimique  Physico chimie  Bilan énergétique  Flore  Végétation  Paysage  Dendromètres ConnecSens  Phénologie  Prélèvements et bibliographie  Historique (10 ans)  Historique depuis 1788  Historique en Excel  Fichiers volumineux  250To + 100 To | Open Data  Partage en groupe  Licence  Avec temps de rétention  Réserve sur la distribution  Formulaire de traçabilité  Privée  Convention/pécuniaire  A définir | Collecte  Saisie  Stockage  Structuration  Traçabilité  Pérennisation  Partage  Réutilisation  Modélisation  Exploitation | Nouveaux partenaires  Données de nouveaux partenaires  Outils de stockage  Apposition de DOI  Traçabilité de la donnée  Outils de partage  Facilité de partage  Rendu cartographique  Utilisation des autres données du cloud environnemental  Interconnectibilité interne  Données météo  Données de qualité  Amélioration de la qualité des données et métadonnées  Outils de reformatage  Automatisation  Rendu visuel des carottes |

## Synthèse Globale Simplifiée

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rôle | Données | Politique sur la donnée | Usages | Attentes |
| Producteur/  Consommateur  Producteur  Consommateur | Cartographie  Flux de données  Fichiers  Historique  Fichiers volumineux | Open Data  Partage en groupe  Privée  Convention/pécuniaire | Stockage  Partage  Exploitation | Données de nouveaux partenaires  Traçabilité de la donnée  Rendu cartographique  Données météo  Données de qualité  Outils de reformatage  Automatisation  Rendu visuel des carottes |

# Livrables et jalons

Liste des échéances à l’horizon 2020.

## Validation de l’analyse des besoins

Fin mars 2018.

## Livraison du cahier des charges

Fin juin 2018.

## Validation du cahier des charges

Mi-juillet 2018.

## Prototype

Juin 2019

# Révisions

|  |  |  |
| --- | --- | --- |
| **Date création** | **Auteur** | **Référence** |
| 27/03/2018 | S.Cipière | 20180327SC |

|  |  |  |  |
| --- | --- | --- | --- |
| **Date révision** | **Auteur** | **Révision** | **Commentaires** |
| 27/3/2018 | V. Breton, S. Cipière, B. Gouhier | Relecture et édition du document dans son ensemble |  |
| 28/3/2018 | D. Sarramia | Relecture et édition du document |  |
| 28/3/2018 | G. Mailhot | Relecture et édition du document |  |
| 28/3/2018 | B. Gouhier | Ajout des contributions du LIT |  |
| 28/3/2018 | V. Breton | Relecture et édition du document dans son ensemble |  |

# Annexes