

Cahier des charges

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

Challenge 1 – CAP 20-25

Version 20180627-1800-SC

SOMMAIRE

[1 Introduction 1](#_Toc517947473)

[2 Système d’information 2](#_Toc517947474)

[3 Acteurs 3](#_Toc517947475)

[3.1 Définition 3](#_Toc517947476)

[3.2 Rôles 3](#_Toc517947477)

[3.2.1 Consommateurs 3](#_Toc517947478)

[3.2.2 Producteurs 3](#_Toc517947479)

[3.2.3 Administrateurs 3](#_Toc517947480)

[3.2.4 Super Administrateurs 4](#_Toc517947481)

[3.2.5 Rôles complémentaires 4](#_Toc517947482)

[4 Données 5](#_Toc517947483)

[4.1 Définitions 5](#_Toc517947484)

[4.2 Sources de données 6](#_Toc517947485)

[4.2.1 Réseaux déjà existants et structurés 6](#_Toc517947486)

[4.2.2 Fichiers type Excel ou texte (type CSV) 6](#_Toc517947487)

[4.2.3 Flux de réseaux de capteurs 6](#_Toc517947488)

[4.2.4 Fichiers volumineux 6](#_Toc517947489)

[4.3 Types de données 7](#_Toc517947490)

[4.3.1 Données d’observation 8](#_Toc517947491)

[4.3.1.1 ConnecSens et réseaux de capteurs 8](#_Toc517947492)

[4.3.1.2 Saisie humaine 8](#_Toc517947493)

[4.3.2 Données expérimentales 9](#_Toc517947494)

[4.3.3 Données de simulation numérique 9](#_Toc517947495)

[4.3.4 Données dérivées ou compilées 9](#_Toc517947496)

[4.3.5 Données de référence 9](#_Toc517947497)

[4.4 Format des fichiers 9](#_Toc517947498)

[4.4.1 Flux de réseaux de capteurs 9](#_Toc517947499)

[4.4.1.1 Solutions basées sur un réseau Internet Protocol IP 9](#_Toc517947500)

[4.4.1.2 Solutions basées sur le LoRaWAN 9](#_Toc517947501)

[4.4.1.3 Solutions basées sur la radiotéléphonie (GSM-SMS) 9](#_Toc517947502)

[4.4.2 Tableaux de données 10](#_Toc517947503)

[4.4.3 Rapports 10](#_Toc517947504)

[4.4.4 Audiovisuels 10](#_Toc517947505)

[4.4.5 Autres types de fichiers 10](#_Toc517947506)

[4.5 Flux (webservices) 11](#_Toc517947507)

[4.6 Cartographique 11](#_Toc517947508)

[4.6.1 Flux (Webservices) 11](#_Toc517947509)

[4.6.2 Fichiers cartographiques 11](#_Toc517947510)

[4.7 Historique 11](#_Toc517947511)

[4.7.1 Fichiers historiques 11](#_Toc517947512)

[4.7.2 Non numérisé 11](#_Toc517947513)

[4.8 Droits sur les données 12](#_Toc517947514)

[4.8.1 Open Data 12](#_Toc517947515)

[4.8.1.1 Open Data complète 12](#_Toc517947516)

[4.8.1.2 Open Data avec limitations 12](#_Toc517947517)

[4.8.2 Partage en groupe 12](#_Toc517947518)

[5 Métadonnées 13](#_Toc517947519)

[5.1 Définition 13](#_Toc517947520)

[5.2 Exemple 13](#_Toc517947521)

[5.3 Bonnes pratiques 2](#_Toc517947522)

[5.3.1 Approche FAIR 2](#_Toc517947523)

[6 Périmètre temporel 2](#_Toc517947524)

[7 Périmètre géographique 3](#_Toc517947525)

[8 Périmètre culturel 3](#_Toc517947526)

[9 Fonctionnalités 4](#_Toc517947527)

[9.1 Interfaces 4](#_Toc517947528)

[9.1.1 Interfaces graphiques 4](#_Toc517947529)

[9.1.2 Lignes de commandes 4](#_Toc517947530)

[9.2 Authentification des Acteurs 4](#_Toc517947531)

[9.2.1 Fédération d’identité 4](#_Toc517947532)

[9.2.2 Couple : Login et mot de passe 5](#_Toc517947533)

[9.2.3 Accès libre 5](#_Toc517947534)

[9.3 Gestion des Acteurs/utilisateurs 6](#_Toc517947535)

[9.3.1 Définitions 6](#_Toc517947536)

[9.3.2 Création d’utilisateurs 6](#_Toc517947537)

[9.3.3 Suppression d’utilisateurs 7](#_Toc517947538)

[9.3.4 Droits 7](#_Toc517947539)

[9.3.5 Groupes 7](#_Toc517947540)

[9.4 Gestions des données 7](#_Toc517947541)

[9.4.1 Création 7](#_Toc517947542)

[9.4.2 Modification 7](#_Toc517947543)

[9.4.3 Suppression 7](#_Toc517947544)

[9.4.4 Droits 7](#_Toc517947545)

[9.4.5 Traçabilité 7](#_Toc517947546)

[9.4.5.1 Paternité 7](#_Toc517947547)

[9.4.5.2 Réutilisation 8](#_Toc517947548)

[9.4.6 Recherche 8](#_Toc517947549)

[9.4.6.1 Moteur de recherche multicritères 8](#_Toc517947550)

[9.4.6.2 Interface graphique 8](#_Toc517947551)

[9.4.7 Normalisation 8](#_Toc517947552)

[10 Solutions envisagées 9](#_Toc517947553)

[10.1 9](#_Toc517947554)

[11 Hébergement 9](#_Toc517947555)

[11.1 Mésocentre Clermont Auvergne 9](#_Toc517947556)

[12 Contraintes & opportunités 9](#_Toc517947557)

[12.1 Opportunités institutionnelles 9](#_Toc517947558)

[12.1.1 Utilisation des ressources du Mésocentre Clermont-Auvergne 9](#_Toc517947559)

[12.2 Opportunités techniques 9](#_Toc517947560)

[12.2.1 Mise à disposition des données des instruments ConnecSens 9](#_Toc517947561)

[12.2.2 Interopérabilité avec le LIT Grandes Cultures en Auvergne 9](#_Toc517947562)

[12.2.3 Conformité avec les standards internationaux 9](#_Toc517947563)

[12.3 Contraintes légales 10](#_Toc517947564)

[12.3.1 Nécessité de gestion de la propriété intellectuelle des données 10](#_Toc517947565)

[12.3.2 Respect de la réglementation sur le financement des ressources 10](#_Toc517947566)

[13 Cas d’utilisation & partenariats 10](#_Toc517947567)

[13.1 Observatoire de l’allier 10](#_Toc517947568)

[13.2 ConnecSens et réseau de capteurs sans fils 10](#_Toc517947569)

[13.3 Laboratoire d’Innovation Territorial (LIT) 10](#_Toc517947570)

[13.4 beeHope 10](#_Toc517947571)

[14 Gouvernance 11](#_Toc517947572)

[14.1 Comité éthique et scientifique 11](#_Toc517947573)

[14.1.1 Composition 11](#_Toc517947574)

[14.1.2 Rôle 11](#_Toc517947575)

[14.1.2.1 Autorisation de nouvelles données ou de nouveaux projets 11](#_Toc517947576)

[14.1.2.2 Suppression de données 11](#_Toc517947577)

[15 Livrables et jalons 12](#_Toc517947578)

[16 Synthèses 12](#_Toc517947579)

[16.1 Synthèse Globale 13](#_Toc517947580)

[16.2 Synthèse Globale Simplifiée 14](#_Toc517947581)

[17 Révisions 15](#_Toc517947582)

[18 Annexes 16](#_Toc517947583)

[18.1 Analyse des besoins du CEBA 16](#_Toc517947584)

[18.2 Compte rendu des interviews du CEBA 16](#_Toc517947585)

[18.3 Licences applicables aux données 16](#_Toc517947586)

[18.3.1 Licence ouverte Etalab 16](#_Toc517947587)

[18.3.1.1 Licence ODC Open Database License (ODbL) 17](#_Toc517947588)

[18.3.1.2 Creative Commons (CC) 17](#_Toc517947589)

[19 Webographie 21](#_Toc517947590)

Liste des figures

[figure 1 : Classification des différents types de données 5](#_Toc517947591)

[figure 2 : Types de données d’après Gaillard 2014 source : URFIST Rennes 7](#_Toc517947592)

[figure 3 : Schéma du réseau de capteurs ConnecSens 8](#_Toc517947593)

[figure 4 : Exemple d’une page de login utilisant la FÉR 5](#_Toc517947594)

[figure 5 : Exemple de page permettant la connexion de membres extérieurs à la FÉR 5](#_Toc517947595)

[figure 6 : Logos de la Licence Ouverte 2.0 16](#_Toc517947596)

[figure 7 : Exemple de logos Creative Commons 19](#_Toc517947597)

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 vise à définir les périmètres du Cloud Environnemental au Bénéfice de l’Agriculture en Auvergne CEBA. Ce document s’appuie sur « l’analyse des besoins et des contraintes du CEBA » et « le compte rendu des interviews du CEBA ». Nous détaillerons les besoins exprimés ainsi que les objectifs attendus.

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 finalité est la création d’un service d’observation numérique de l’environnement et des agroécosystèmes (eau, sol, air, biodiversité) à l’échelle du territoire auvergnat pour répondre à différentes problématiques qu’elles soient scientifiques, culturelles ou socio-économiques, au bénéfice de l’agriculture.

Le "Cloud Environnemental" incarne l’ambition de créer un « grand » observatoire de l’environnement en Auvergne, en mettant à la disposition de la communauté scientifique un environnement numérique interconnecté valorisant les données environnementales existantes ou à acquérir. Apportant des fonctionnalités en matière de stockage, gestion, sécurisation des données environnementales mesurées sur les différents dispositifs instrumentés en région Auvergne, le « Cloud environnemental » offrira des facilités en termes d’organisation et d’extraction des informations pour :

1. progresser dans notre compréhension des compartiments environnementaux face au changement global (évolutions, interactions),
2. en comprendre l’impact sur les agro-écosystèmes et réciproquement, et
3. conceptualiser les interrelations entre les compartiments pour comprendre, modéliser et prédire le comportement des agroécosystèmes.

# Système d’information

L’objectif du CEBA est de mettre à disposition des acteurs un système d’information (SI) qui répondent aux besoins listés dans le « CR des interviews du CEBA » et dans « l’analyse des besoins du CEBA ». Les usages considérés sont : Collecte, Saisie, Stockage, Structuration, Traçabilité, Pérennisation, Partage, Réutilisation, Modélisation, Exploitation. Ces termes sont explicités dans ce document pour éviter toute ambiguïté liée à l’usage du Français. Le SI sera constitué d’un ou plusieurs serveurs qui seront dimensionnés pour supporter convenablement les charges imposées par les usages attendus. Les technologies retenues devront garantir une exploitation pérenne et être capable de supporter un passage à l’échelle en cas de modification ultérieure d’un ou plusieurs périmètres.

Nous bénéficierons du soutien des équipes du Mésocentre Clermont Auvergne (MCA). Pour garantir la meilleure efficacité entre nos services nous respecterons les recommandations faites par les ingénieurs du MCA.

Les machines physiques sont gérées par le MCA, elles sont intégrées au sein du datacenter ce qui garantit la fourniture des fluides (électricité et froid), essentiels au bon fonctionnement de ces matériels. Ces serveurs sont supervisés avec des produits VMware qui permettent de virtualiser les machines physiques en machines virtuelles. Concrètement le MCA fournira au CEBA un service de type IaaS (*Infrastructure as a Service*) sur lesquels nous pourrons déployer les Systèmes d’exploitations (OS *Operating System*) et les différents composants logiciels nécessaire à la mise en œuvre du CEBA. L’expérience et l’expertise du MCA permettra en cas de besoin d’exploiter d’autres types de technologies tel que les containers, ou autres, pour répondre aux problématiques logicielles.

Les machines seront préférentiellement installées avec des systèmes d’exploitation CentOS 7 qui permettent une fréquence de mise à jour moins élevée, d’autres version de Linux ou Windows seront envisageable en fonction des besoins.

Les connexions des super administrateurs s’effectuerons par SSH (*Secure Shell*) avec ou sans interface graphique.

Nous profiterons autant que possible de l’expérience et des solutions déjà mise en œuvre par le MCA si nos périmètres opérationnels respectifs le permettent.

Les machines virtuelles sont reliées aux réseaux CRATERE et RENATER, ce qui garantit une excellente connexion à internet. Elles sont cloisonnées à l’intérieur d’un VLAN (78-CEBA) et éventuellement d’une VRF pour favoriser la sécurité et la confidentialité des données, sur les différents sites de l’UCA.

Les intérêts d’être hébergé au MCA sont multiples, nous avons évoqué la qualité de la gestion des machines et les connexions réseau, nous pourrons aussi bénéficier de leur expérience en termes de sauvegarde des données sur un temps long. L’expertise et l’expérience des personnels du Mésocentre Clermont Auvergne et de la DSI UCA seront mises à disposition du projet CEBA.

# 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.

### Consommateurs

Les consommateurs sont des Acteurs qui consultent des données. On peut diviser les consommateurs en deux groupes, les acteurs ne nécessitant pas d’authentification et des acteurs authentifiés. Sans authentification seules les données Open Data seront accessibles. 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 les métadonnées les décrivant. Aucune contrainte autre que la fourniture des métadonnées précédentes ne sera imposée pour l’intégration d’une donnée (notamment le format et la structure).

Ces 2 rôles pourront s’additionner sur un compte en fonction des responsabilités de la personne. Nous obtenons ainsi trois rôles « Consommateurs », « Producteurs » et « Producteur/Consommateurs » que nous avions déjà défini dans l’analyse des besoins. À ces rôles s’ajoute des rôles d’administration.

### Administrateurs

Les « administrateurs » peuvent créer et gérer des groupes d’Acteurs. Ils peuvent aussi gérer les droits de lecture et d’écriture des Acteurs placés sous leur tutelle. Ainsi l’informaticien référent d’un laboratoire peut créer un groupe qui représente un des axes des recherches de son laboratoire. Il peut déléguer son droit d’administration à un DR qui en aura fait la demande. Ces deux administrateurs peuvent continuer de peupler le groupe qui est la représentation de l’axe avec les identités numériques de leurs collègues. Chacun de ces collègues pourra avoir des droits d’administration plus ou moins étendu, voir aucun. Nous détaillerons les différents droits dans le paragraphe 4.4.

### Super Administrateurs

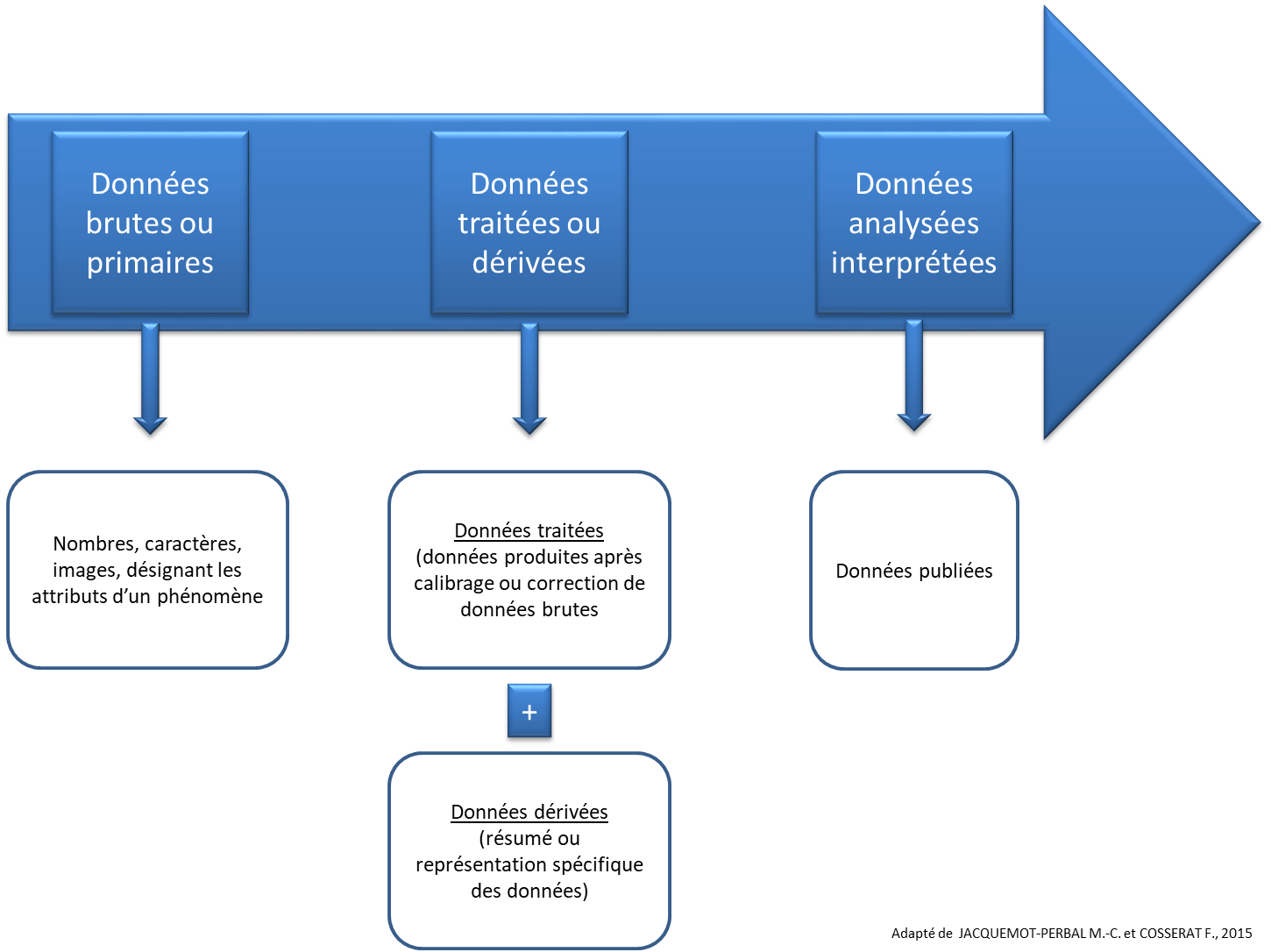
Les informaticiens responsables du projet disposeront de droits de « super administrateurs » qui leur octroie un contrôle complet sur les utilisateurs et les données de façon à pouvoir rapidement répondre aux besoins des acteurs. Ainsi avec l’aval du comité scientifique ils peuvent créer de nouveaux groupes d’acteurs et de données. Ces utilisateurs seront techniquement capables d’accéder à toutes les données du CEBA, ils seront soumis à des clauses de confidentialités et de non divulgations.

### Rôles complémentaires

Des rôles additionnels pourront être envisagés en fonction des besoins. Il pourra s’agir de sous rôles qui traduiront de besoins fonctionnels particuliers. L’objectif étant de fournir suffisamment

# Données

## Définitions



1. Classification des différents types de données

Pour nous permettre d’utiliser un vocabulaire commun lorsque nous nous référons aux différents types de données, nous utiliserons comme base la représentation de la figure 1 : adaptée de (JACQUEMOT-PERBAL and COSSERAT 2015).

Données : « Informations qualitatives ou quantitatives (…) qui sont factuelles. Ces données peuvent être brutes ou primaires (directement issues d’une mesure), ou dérivées de données primaires mais ne sont pas encore le produit d’analyses ou interprétations autres que de calcul » (The Royal Society, Science as an open enterprise : summary report, juin 2012, p. 9).

Jeu de données : « Agrégation, sous une forme lisible, de données brutes ou dérivées présentant une certaine unité, rassemblées pour former un ensemble cohérent ». Cela peut donc consister en un enregistrement de données sous la forme d’un ou plusieurs fichiers électroniques, téléchargeables, citables (notamment par le biais d’un DOI) et intelligibles, ce jeu étant accompagné des métadonnées descriptives suffisantes (Rémi Gaillard, De l’open data à l’open research data : quelle(s) politique(s) pour les données de recherche ? janvier 2014, p. 19).

## Sources de 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.

### 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 lister les observatoires ou les structures comme le BRGM, leurs méthodes sont bien rodées, et généralement ils n’attendent pas de données en retour du CEBA, ce sont souvent des producteurs uniquement.

### Fichiers type Excel ou texte (type CSV)

Pour des résultats d’analyses ou des mesures réalisées par un opérateur, 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. De plus les outils qui pourront être proposés par le CEBA telle que l’application de DOI qui permet la traçabilité de jeux de données les intéresse fortement.

### Flux de réseaux de capteurs

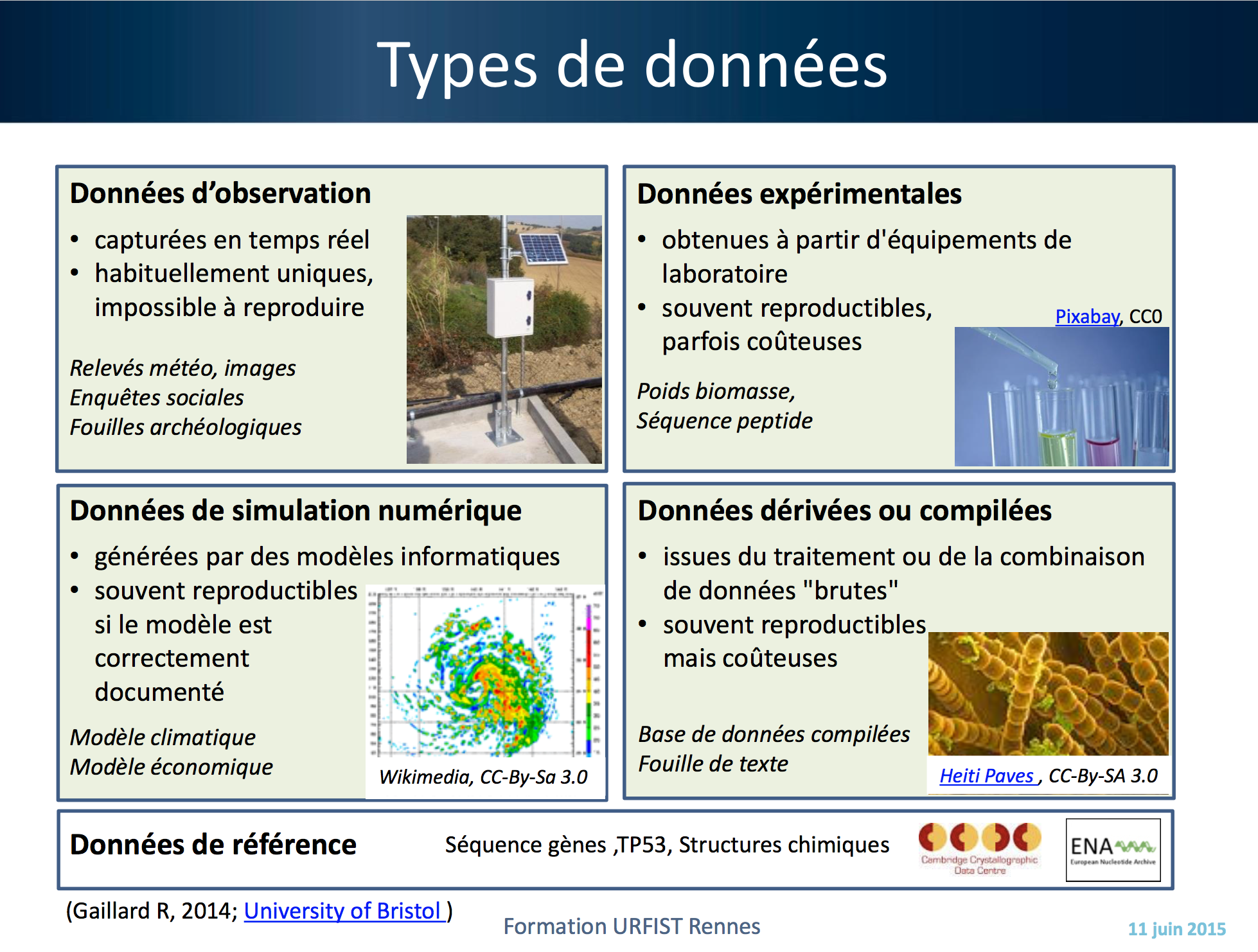
Le projet ConnecSens est un bon exemple de réseau de capteurs sans fils. Le CEBA pourra partager des données qui proviennent d’Acteurs n’utilisant pas les mêmes technologies sans fil que celles utilisées par ConnecSens. Il s’agit ici de stocker, structurer, pérenniser, et rendre disponible les données collectées par les capteurs de différentes natures (piézomètre, humidité du sol, dendromètre, station météorologique, etc.). 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).

## Types de données

Nous listons ici les différents types de données qui pourront être échangées à l’aide du CEBA. Nous nous baserons sur les travaux de Gaillard, R (2014) [[1]](#footnote-1).



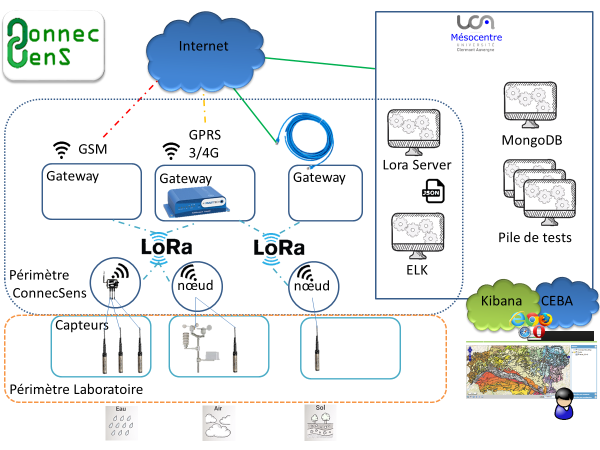
1. Types de données d’après Gaillard 2014 source : URFIST Rennes

Cinq types de données doivent être différenciés : observation, expérimentales, simulation numérique, dérivées ou compilées, références. Nous détaillerons dans les paragraphes suivants quelles sont les spécificités de ces cinq types de données et pourquoi il est intéressant de les traiter distinctement.

### Données d’observation

Les données d’observation sont capturées en temps réel, elles sont impossibles à reproduire. Elles peuvent être mesurées par des capteurs ou être issues d’une saisies humaines relative à une interview, une étude, des fouilles archéologiques, etc.

#### ConnecSens et réseaux de capteurs



1. Schéma du réseau de capteurs ConnecSens

Pour permettre l’accès aux données provenant de réseaux de capteurs, qu’elles soient transmises par GSM, ou sous forme de données, le SI du CEBA s’adaptera aux différents formats qui pourront être transmis. Dans le cas du producteur automate les métadonnées pourront être définies de façon standardisée.

Parmi les partenaires potentiels du CEBA nous pouvons ici envisager à termes des flottes de bouées mesurant les qualités physico-chimiques de l’eau, des capteurs de polluant atmosphérique, des stations météorologiques, du monitoring de ruches etc.   
La fréquence des mesures peut être définie au préalable ; toutes les 15 minutes ou horaires ; ou varier en fonction d’un élément déclencheur. Cette fonctionnalité permet d’ajuster le pas de temps lors d’événements rares.

#### Saisie humaine

L’automatisation des mesures n’est pas toujours possible, certains relevés nécessitent l’expertise humaine. Les données seront alors transmises dans un second temps au CEBA. Suite aux auditions que nous avons effectuées les données sont fréquemment saisies dans un tableur sur un ordinateur portable. Les résultats d’analyses provenant d’échantillons prélevés sur le terrain seront considérés comme saisie humaine.

### Données expérimentales

Les données expérimentales sont obtenues à partir d’équipements de laboratoire, elles sont souvent reproductibles et parfois coûteuses.

### Données de simulation numérique

Les données de simulation numériques sont générées par des modèles informatiques. Ces données sont souvent reproductibles si le modèle le permet et qu’il est suffisamment documenté. Certains de nos partenaires souhaitent mettre à disposition ce type de données, notamment à travers des résultats de simulations de modèles climatiques. C’est d’ailleurs l’une des singularités du CEBA qui pourra mettre à disposition des données « futures » ainsi les partenaires pourront disposer d’une multitude de climat potentiel pour un périmètre géographique et une année donnée.

### Données dérivées ou compilées

Les données dérivées ou compilés sont issues du traitement ou de la combinaison de données « brute », souvent reproductibles mais coûteuses.

### Données de référence

Les données de référence peuvent soit être des données existantes fournies par des acteurs de références (exemple des données génomique, Météo France, ou du réseau RNM…) soit être des données stockées qui doivent devenir après analyse des données de référence.

## Format des fichiers

### Flux de réseaux de capteurs

Pour permettre l’accès aux données des réseaux de capteurs sans fils ; telles que celles de ConnecSens ; qu’elles soient transmises par GSM, ou sous forme de données ; le SI du CEBA s’adaptera aux différents formats qui pourront être transmis tant qu’ils sont exploitables par les outils et techniques sélectionnés. Le flux peut être plus ou moins continue, dans les faits nous sommes face à des fréquences d’envoi variant du quart d’heure à la journée. C’est le médium qui distinguera les flux entre eux. En effet le format de la donnée sera tributaire des solutions techniques permettant sa transmission.

#### Solutions basées sur un réseau Internet Protocol IP

C’est le cas le plus simple techniquement car nous restons dans le paradigme le plus courant. Les protocoles pourront varier d’une implémentation à l’autre, la totalité des formats de fichiers pourra transiter par ces solutions. Il sera préférable de favoriser les protocoles sécurisés lorsque cette alternative sera possible.

#### Solutions basées sur le LoRaWAN

.

#### Solutions basées sur la radiotéléphonie (GSM-SMS)

A partir du GPRS et au-dessus (2/3/4/5 G) les solutions de transmission de données par radiotéléphonie se comporte comme un réseau IP. Nous considérons donc ici les données transmises par radiotéléphonie (type SMS) qui ont la particularité de nécessiter une technologie de téléphonie cellulaire et un abonnement au niveau du serveur de réception. Le SMS (*Short Message Service*) a besoin d’être reçu et transmis au serveur avant de pouvoir être traité comme une donnée brute standard. Ce genre de technologie est souvent mis en œuvre pour que le serveur envoie des SMS à des utilisateurs, il s’agit ici d’une situation opposée, où s’est le serveur qui est le destinataire du message. Les technologies pourront être propriétaires comme c’est le cas pour la bouée qui sera mise en œuvre à Aydat. Nous ferons en sorte de pouvoir répondre aux différents besoins soulevés par nos partenaires. Notre objectif est de rendre ces phases techniques et technologiques transparentes pour nos utilisateurs finaux, qu’ils soient producteurs ou consommateurs.

### Tableaux de données

Les fichiers de types Excel, Calc ou \*.CSV représentent des tableaux de données. Pour pouvoir être exploité le fichier devra être décrit, lignes et colonnes utiles, typage des caractères (ex : présence d’accents, etc.) ; toutes ces informations s’ajoutent aux métadonnées du fichier. Conscient du poids de la démarche demandée au producteur différentes solutions de déposes pourront être proposées pour essayer de minimiser au maximum ce frein à la dépose. Des modèles pourront être crées pour importer des tableaux de séries similaires. Ces tableaux seront sauvegardés dans des bases de données BDD. Il sera intéressant de définir les champs ajoutés à chaque entrée du tableau, les besoins pourront varier d’un producteur à l’autre.

### Rapports

En fonction du type de fichier éditable comme un \*.txt ou figé comme un \*.pdf, le fichier pourra être rendu accessible en tant que tel à l’instar d’une image, ou stocké sous une forme d’une page Web. Il sera aussi possible de rendre ce type de fichier accessible au travers d’une interface type « *drive* » de partage de fichiers. Les avantages et inconvénients de chaque approche seront mis en exergues durant les phases de prototypages.

### Audiovisuels

Certains types de fichiers souvent volumineux, nécessitent des lecteurs spécifiques, ou seront rendu accessible au travers d’une interface type « *drive* » de partage de fichiers, comme les autres fichiers qui ne nécessitent pas de traitement.

###### Images

En fonction de la taille initiale de l’image il pourra être intéressant de la présenter sous forme d’une miniature. Pour permettre de naviguer à travers un stock important d’images, avant d’afficher en pleine résolution les images d’intérêt.

###### Vidéos

Comme pour les images les vidéos pourront être présentées sous forme d’une image unique, avant d’être parcourue. Il faudra définir les types de fichiers vidéo qui seront supportés par le CEBA.

###### Audios

Les types de fichiers audios supportés devront être définis. Le lecteur audio pourra ressembler au lecteur vidéo.

### Autres types de fichiers

Les types de fichiers non cités précédemment seront traités au fur et à mesure de l’apparition du besoin. Ils pourront être traités d’une façon similaire aux fichiers précédents, ou ils nécessiteront une approche spécifique.

## Flux (webservices)

Certains réseaux structurés mettent à disposition leurs données à travers des services WEB. On peut notamment mentionner hub’eau [[2]](#endnote-1) qui a pour vocation de rendre accessible d’une manière simplifiée toutes les données libres relatives à l’eau. Les politiques Européennes et Françaises s’orientent fortement en faveur de l’*OpenData* ainsi il sera envisageable d’accéder aux données pertinentes de la plateforme ouverte des données publiques françaises[[3]](#endnote-2).

## Cartographique

### Flux (Webservices)

Certains de nos partenaires CRAIG, IGN, BRGM, proposent de fournir leurs services à travers des webservices. Les avantages de ce genre d’approche sont nombreux (pas de stockage de notre côté, …). L’objectif est de pouvoir proposer tous ces flux à travers la même interface et de les afficher à la demande (comme des calques dans éditeur d’images).

### Fichiers cartographiques

Contrairement au flux de nos partenaires géographes, d’autres acteurs disposent de leurs propres fichiers cartographiques. Qu’ils souhaitent les partager ou juste les consulter le CEBA fournira un support permettant de visionner ces cartes, comme autant de calques supplémentaires.

## Historique

Les données historiques bien que souvent moins précises sont précieuses car elles sont un témoignage du passé et ne pourront plus être mesurées pour la plupart.

### Fichiers historiques

S’ils sont du même type que les fichiers mentionnés précédemment leur intégration sera relativement transparente. Il faudra autoriser que la datation de la mesure soit suffisamment ancienne se pose alors la question du paragraphe périmètre temporelle ??

### Non numérisé

Pour les documents non numérisés, la demande n’est pas assez forte pour que le CEBA fournisse une solution.

## Droits sur les données

### Open Data

*Open Data* signifie Données ouvertes, ce sont des données dont l’accès, l’utilisation et la réutilisation sont publics et libres de droits.

L’Open Data se caractérise par 3 critères essentiels :

• Disponibilité et accès : les données doivent être accessibles, moyennant un coût de reproduction raisonnable, et pouvoir être téléchargées sur Internet. La forme des données doit être pratique et modifiable.

• Réutilisation et redistribution : les données doivent être fournies sous des conditions permettant la réutilisation et la redistribution, incluant le mélange avec d’autres ensembles de données.

• Participation universelle : tout le monde doit être en mesure d’utiliser, de réutiliser et de redistribuer les données. Il ne doit y avoir aucune discrimination à l’égard des utilisateurs (restrictions à certains secteurs par exemple) ou concernant les fins d’utilisation.

Il est possible de revendiquer la paternité des données publiées dans des archives ouvertes par le biais   
d’un DOI (*Digital Object Identifier*), qui lie de façon pérenne un document à son auteur. Le DOI est un mécanisme unique d’identification de ressources numériques comme un film, un rapport, des articles scientifiques, des bases de données,   
des logiciels, etc. Il permet de protéger des données ou des jeux de données.

#### Open Data complète

#### Open Data avec limitations

### Partage en groupe

# Métadonnées

## Définition

Les métadonnées sont les données qui décrivent la donnée.

Dans le cas du producteur automate les métadonnées pourront être définies de façon standardisée.

Pour pouvoir exploiter pleinement les données il faudra 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 le producteur souhaite partager ses données (ex. privé, groupe, ou public) sera un élément essentiel à communiquer, pour remplir cet objectif nous avons mis en place un DMP ??[réf]

## Exemple

Prenons l’exemple d’un échantillon d’eau prélevé à Aydat aujourd’hui (2018-04-20). Cet échantillon sera apporté au laboratoire d’analyse sur le campus des Cézeaux et analysé le lendemain. Suite à l’analyse nous obtiendrons une teneur un sel de 4 mg pour 100 grammes.

La donnée ici est teneur en sel du lac d’Aydat le 20 avril 2018 - 4mg pour 100 grammes.

La donnée pourrait aussi être teneur en sel du lac d’Aydat en avril 2018 - 4mg pour 100 grammes.

Considérons maintenant les métadonnées autour de cette valeur.

En fonction de l’existence ou non d’un protocole de prélèvement et/ou d’un protocole de transport et/ou d’un protocole d’analyse. Nous pourrions avoir comme métadonnées protocoles suivi à la lettre sans encombre. Ou protocole de prélèvement OK, protocole de transport avec bouchon sur la N89 durant 1h possibilité de température plus élevée que d’habitude durant le transport et protocole d’analyse OK.

Une version très détaillée pourrait être

Préleveur : Opérateur 101

Date de prélèvement : 14 :15

Prélèvement à 45°39'58.9"N+2°59'01.0"

Environnement : plage vide

Qualité visuelle de l’eau : limpide sans écume

T°C de l’eau à 30 cm du bord 16°C

T°C de l’air 24°C (source OPGC)

Protocole sceau de 10 litres

puis éprouvette stériles 50ml – sterigold n° de lot 20180101

Placée dans pour éprouvette puis glacière

Départ de Aydat 14 :30

Arrivée au Laboratoire 15 :00

Placée au réfrigérateur à 24°C à 15 :10

Personne en charge de l’analyse : Laborantin N°42

Sortie réfrigérateur à 23°C à 09 :10 le 21/04/2018

Méthode d’analyse évaporation par chauffe au gaz

Pesée avec balance de précision X.Y.Z tarée le 02/04/2018

Mesure masse avant chauffe

Mesure masse après chauffe

Produit en croix pour rapporter la mesure 50ml à 100g

Saisie des valeurs dans un tableur

## Bonnes pratiques

### Approche FAIR

Nous suivrons le principe FAIR[[4]](#endnote-3) (***F****indable,* ***A****ccessible,* ***I****nteroperable, and* ***R****e-usable*) pour les données et les métadonnées. En effet il est important que les données et les métadonnées soient faciles à trouver, que les utilisateurs puissent y accéder, qu’elles soient interopérables et réutilisables.

Pour être facilement trouvable :

F1. Les (méta)données dispose d’un identifiant unique type DOI.

F2. Les données sont accompagnées et décrites par des métadonnées de qualité.

F3. Les (méta)données sont enregistrées ou indexées dans un moteur recherche.

F4. Les métadonnées contiennent l’identifiant de la donnée.

Pour être facilement accessible :

A1 Les (méta)données sont récupérable à l’aide de protocoles standards.

A1.1 Le protocole est libre, gratuite et universellement implémentable.

A1.2 Le protocole permet une authentification lorsque c’est nécessaire.

A2 Les métadonnées sont accessibles même si les données ne sont pas accessibles.

Pour être interopérable :

I1. Les (méta)données utilisent un langage formel, accessible, largement partagé pour décrire les connaissances.

I2. Les (méta)données utilisent un vocabulaire qui suit le principe FAIR.

I3. Les (méta)données incluent des liens vers d’autres (méta)données de références.

Pour être réutilisable :

R1. Les (méta)données sont composées d’attributs précis et pertinents.

R1.1. Les (méta)données publiées avec des licences claires et accessibles.

R1.2. Les (méta)données sont associées à leur provenance.

R1.3. Les (méta)données répondent aux normes standard de leur domaine.

# Périmètre temporel

Les données pourront être enregistrées avec leurs métadonnées. Les métadonnées pourront comporter une date de création de la donnée, une date de numérisation, des dates liées aux traitements et à la validation de la qualité des données, une date ou un intervalle temporel durant lequel la donnée est pertinente. Le format de la date peut être aussi précis que la   
microseconde 10-6 seconde (YYYY-MM-DD HH:MM:SS.000000) . Pour des raisons de conformité avec les standards internationaux les dates seront stockées sous le format YYYY-MM-DD et pourront être fournies sous le format JJ/MM/AAAA.   
Certaines fonctions renvoient le nombre de millisecondes (10-3 seconde) entre le 1er janvier 1970 à minuit, nous mettrons à disposition les outils pour rendre ces informations facilement lisibles et compréhensibles. Le standard temps universel coordonné (UTC) sera préféré à celui de *Greenwich Mean Time* (GMT). De par la paternité Française du projet le fuseau horaire par défaut sera UTC +01 :00 (ce fuseau couvre environs 10% de la population mondiale). Les dates pourront être des dates futures. Nous pourrons ainsi adresser les besoins de partage de résultats de modélisation, de projection climatique. Les images résultantes de la numérisation d’archives comme les herbiers pourront aussi être traitées. Les données dont la datation est moins précise disposeront d’un intervalle de pertinence.

# Périmètre géographique

Le périmètre initial et minimal est l’Auvergne (le territoire auvergnat). Ce périmètre s’il est évident qu’il doit être couvert ne doit pas être limitant, si notre SI est capable de répondre aux problématiques auvergnates il doit être capable d’adresser celles de la région AURA. De même si l’on considère la thématique rivière Allier il serait contreproductif de se priver des informations provenant de sa source ou en aval des limites de notre territoire. Notre SI devra donc pouvoir passer à l’échelle en fonction des évolutions futures (Auvergne -> AURA -> France métropolitaine -> France -> Europe -> Monde). Les standards de projection (RGF93 / Lambert-93 (EPSG :2154)) que nous utiliserons sont les références de nos partenaires cartographes jusqu’à la granularité France métropolitaine. Les extensions futures s’appuieront sur leurs expertises.

# Périmètre culturel

Comme pour le périmètre géographique nous devrons à minima partager les informations aux seins du territoire auvergnat. Cependant nos partenaires envisagent déjà des échanges avec leurs collègues Européens et à l’heure de la mondialisation il semble essentiel de pouvoir couvrir le spectre complet des caractères envisageable. En effet l’un de nos partenaires a déjà évoqué la possibilité d’exporter notre solution au japon. Ainsi pour être en accord avec ISO/CEI 10646 :2012 et pouvoir apporter une solution globale les données seront encodées préférentiellement en Unicode 10.0.0, UTF 32, UTF 16 et UTF 8 ; en cas de doute, les versions (Big Endian (BE) – Byte Order Mark (BOM ou indicateur d’ordre des octets IOO) seront retenues. Les standards ISO-8859-1 ou Windows 1252, seront évités pour limiter les problèmes liés à l’encodage.

# Fonctionnalités

## Interfaces

Pour accéder aux différents services proposés par le CEBA les utilisateurs doivent disposer d’interfaces, en fonction de leurs attentes et de leurs besoins nous nous proposons de mettre à leur disposition deux types d’interfaces : les interfaces graphiques et les interfaces en lignes de commandes.

### Interfaces graphiques

Les interfaces permettent un accès visuel à l’information, peuvent se présenter sous forme de pages WEB. La première page rencontrée par l’utilisateur dispose d’un bouton qui permet de se connecter, de s’identifier (voir 9.2). Des menus permettent de naviguer entre les différentes pages. Le contenu des menus sera à définir, on peut envisager des accès par thématique (eau, air, sols, biodiversité), un espace de recherche multi critère, une approche visuelle (cartographique) avec un système de calques. Les représentations varieront en fonction du niveau d’accréditation de l’utilisateur.

Les différents services, nécessiteront des vues qui leurs sont propres (gestion des utilisateurs, des projets, des données).

### Lignes de commandes

En fonction des traitements effectués par les utilisateurs il sera appréciable de pouvoir télécharger la donnée directement à travers une interface en lignes de commandes.

## Authentification des Acteurs

Dans notre SI certaines informations sont accessibles librement, c’est le cas des données en OpenData, et d’autres sont soumises à des restrictions des partages au sein de groupe. Pour permettre cette gestion il sera nécessaire d’identifier les utilisateurs du système. Deux moyens, complémentaires, permettent de valider l’identité numérique d’un usager, la fédération d’identité et l’utilisation d’un couple login et mot de passe.

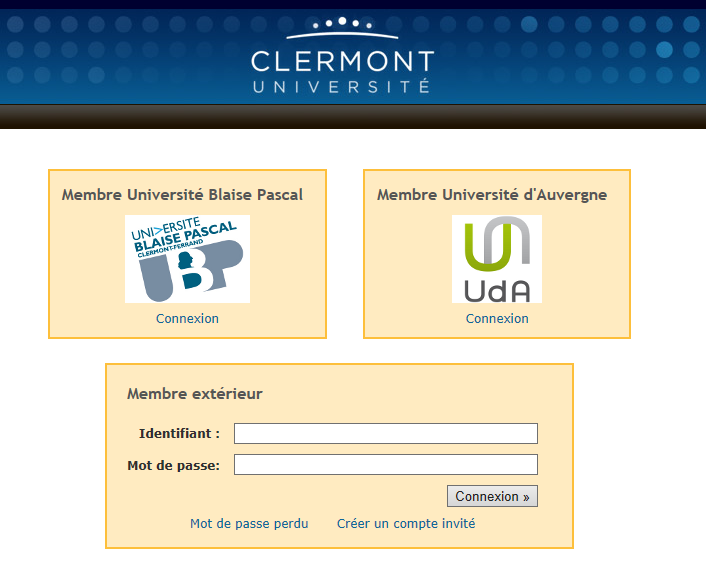
### Fédération d’identité

Le CEBA bénéficie du support de l’UCA et du Mésocentre Clermont Auvergne (MCA). Les tutelles des futurs utilisateurs du CEBA sont nombreuses à appartenir à la Fédération Éducation Recherche[[5]](#endnote-4) (FÉR). La FÉR s’appuie sur un système de *Single Sign-On* (SSO) qui naviguer sur plusieurs sites différents en ne s'authentifiant qu'une seule fois. Pour les personnes ne disposant pas de compte dans la FÉR nous utiliserons un couple login et mot de passe.



1. Exemple d’une page de login utilisant la FÉR

### Couple : Login et mot de passe



1. Exemple de page permettant la connexion de membres extérieurs à la FÉR

### Accès libre

## Gestion des Acteurs/utilisateurs

Les acteurs peuvent être des personnes, des équipes, des projets, des capteurs, des réseaux. Si l’acteurs n’est pas une personne physique il doit être rattaché à une personne physique, pour garantir la traçabilité des données ainsi qu’un moyen de contact en cas d’anomalie sur les données. À l’intérieur du SI nous nommerons les Acteurs, utilisateurs.

### Définitions

Nous définissons ici les champs possibles et minimaux nécessaires à la création d’un utilisateur, pour se faire nous avons déposé une déclaration à la CNIL.

Type d’utilisateur : (personne physique, équipe, projet, capteur, réseau)

Nom dans le système : (chaine de caractères 50)

Civilité : (M. , Mme, Robot, Réseau, …)

Nom : (chaine de caractères 50)

Prénom : (chaine de caractères 50)

Email – ou email du responsable : (chaine de caractères 50)

Téléphone : (chaine de caractères 06 xx xx xx xx ou +336 xx xx xx xx)

Laboratoire : (chaine de caractères 50)

Equipe : (chaine de caractères 50)

Responsable au sens N+1 ou rattachement d’un capteur à un responsable : (chaine de caractères 50)

Site web de référence : (chaine de caractères 50)

Projet : (chaine de caractères 50)

### Création d’utilisateurs

Pour créer un nouvel utilisateur, une requête doit être effectuée.

Si l’utilisateur appartient à la fédération éducation-recherche[[6]](#endnote-5) (FÉR) ses informations seront récupérées à travers le webservice.

Si l’utilisateur n’appartient pas à la FÉR il pourra créer un compte à partir de la page de login, il lui sera alors demandé d’entrée les informations minimales requises et décrites dans le paragraphe précédent 9.3.1. Le compte sera automatiquement créé cependant il ne disposera à ce stade que de l’accès aux informations disponibles en Open Data.

Les automates comme les capteurs de réseaux de capteurs, sont des producteurs particuliers. Ils ne disposent pas de droits d’accès aux données, ils peuvent simplement fournir des données. Ils seront créés par un utilisateur disposant de droits suivant comme un administrateur ou un super-administrateur. Ce type d’utilisateurs sera par défaut rattaché à la personne qui créée son identité numérique. L’automate pourra aussi être rattaché à un projet. Dans ce cas, la responsabilité est transitive. Le responsable du projet, sera par défaut le responsable de l’automate, sauf si la responsabilité est clairement déléguée à une autre personne physique disposant d’un compte au sein du CEBA.

### Suppression d’utilisateurs

Seul les super-administrateurs seront habilités à supprimer le compte d’un utilisateur.

### Droits

Tableaux des droits en fonction du type d’utilisateur

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Création  D’utilisateurs | Octroie des droits  Super-administrateur | Octroie des droits  Administrateur | Création  De projets |  |
| Super-administrateur | X | X | X | X |  |
| Administrateur | X |  | X | X |  |
| Utilisateur enregistré |  |  |  |  |  |
| Automate |  |  |  |  |  |
| Utilisateur non enregistré |  |  |  |  |  |

### Groupes

## Gestions des données

### Création

### Modification

### Suppression

### Droits

### Traçabilité

#### Paternité

#### Réutilisation

### Recherche

#### Moteur de recherche multicritères

#### Interface graphique

### Normalisation

# Solutions envisagées

## 

# Hébergement

## Mésocentre Clermont Auvergne

# 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).

# Cas d’utilisation & partenariats

## Observatoire de l’allier

## ConnecSens et réseau de capteurs sans fils

## Laboratoire d’Innovation Territorial (LIT)

## beeHope

# Gouvernance

## Comité éthique et scientifique

### Composition

### Rôle

#### Autorisation de nouvelles données ou de nouveaux projets

#### Suppression de données

# Livrables et jalons

Liste des échéances à l’horizon 2020.

**Livraison de l’analyse des besoins :**  Mars 2018.

**Validation de l’analyse des besoins :**  Fin mars 2018.

**Livraison du premier cahier des charges :**  Fin juin 2018.

**Validation du premier cahier des charges :**  Mi-juillet 2018.

**Prototype d’évaluation technique :** Juin 2019.

**Révision du cahier des charges :**  Fin juin 2019.

**Démonstrateur :**  Juin 2020.

# Synthèses

Ce travail se base sur les interviews d’une partie des Acteurs (19) du CEBA. Nous avons agrégé dans les tableaux présentés en 16.1 et 16.2 un résumé de l’ensemble de leurs réponses. Le tableau 16.1 Synthèse Globale, présente les éléments sans doublons et regroupés par thèmes et concepts. Le tableau 16.2 Synthèse Globale Simplifiée, est une vue simplifiée du tableau 16.1 qui permet de gagner en lisibilité en factorisant les différents éléments de détails derrière des mots et concepts clefs. Les rapports travaux préliminaires sont disponibles sur le site de documentation doc.ceba.uca.fr. Ce cahier des charges a vocation d’être le support de référence pour les réflexions et développements futurs.

## 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 |

# Révisions

|  |  |  |
| --- | --- | --- |
| **Date création** | **Auteur** | **Référence** |
| 27/06/2018 | S.Cipière | 20180627SC |

|  |  |  |  |
| --- | --- | --- | --- |
| **Date révision** | **Auteur** | **Révision** | **Commentaires** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Annexes

## Analyse des besoins du CEBA

## Compte rendu des interviews du CEBA

## Licences applicables aux données

### Licence ouverte Etalab

Dans le cadre de la politique du Gouvernement Français en faveur de l’ouverture des données publiques (« Open Data »), Etalab[[7]](#endnote-6) a conçu la « Licence Ouverte / Open Licence ». Cette licence, élaborée en concertation avec l’ensemble des acteurs concernés, facilite et encourage la réutilisation des données publiques mises à disposition gratuitement.

La publication du décret n° 2017-638 prévu par l’article L 323-2 du CRPA fait de la Licence Ouverte 2.0 (LO 2.0) la licence de référence pour les administrations pour la publication de données publiques, aux côtés de l’ODbL (cf. 19.3.1.1), et permet ainsi son utilisation par l’ensemble des administrations.

La « Licence Ouverte / Open Licence » présente les caractéristiques suivantes :

1. Une grande liberté de réutilisation des informations :
   1. Une licence ouverte, libre et gratuite, qui apporte la sécurité juridique nécessaire aux producteurs et aux utilisateurs des données publiques.
   2. Une licence qui promeut la réutilisation la plus large en autorisant la reproduction, la redistribution, l’adaptation et l’exploitation commerciale des données.
   3. Une licence qui s’inscrit dans un contexte international en étant compatible avec les standards des licences Open Data développées à l’étranger et notamment celles du gouvernement britannique (*Open Government Licence*) ainsi que les autres standards internationaux (ODC-BY, CC-BY 2.0).
2. Une exigence forte de transparence de la donnée et de qualité des sources en rendant obligatoire la mention de la paternité.
3. Une opportunité de mutualisation pour les autres données publiques en mettant en place un standard réutilisable par les collectivités territoriales qui souhaiteraient se lancer dans l’ouverture des données publiques.

Le logo de la « Licence Ouverte / Open Licence » est également librement réutilisable.



1. Logos de la Licence Ouverte 2.0

#### Licence ODC Open Database License (ODbL[[8]](#endnote-7)[[9]](#endnote-8))

La licence ODbL (Open Database License), supportée par L’*Open Data Commons* ODC, est un contrat de licence ayant pour objet d’autoriser les utilisateurs à partager, modifier et utiliser librement la présente Base de données initiale tout en maintenant ces mêmes libertés pour les autres. De nombreuses bases de données étant protégées par des droits d'auteur, les présentes règles ont pour objet de céder ces droits. Certains États, principalement au sein de l’Union européenne, prévoient des droits spécifiques régissant les bases de données, de ce fait ces droits sont également concernés par la licence ODbL.

#### Creative Commons[[10]](#endnote-9) (CC)

Creative Commons (CC) propose des contrats-type ou licences pour la mise à disposition d’œuvres en ligne. Inspirés par les licences libres, les mouvements *open source* et *open access*, ces licences facilitent l’utilisation d’œuvres (textes, photos, musique, sites web, etc.).

Ces licences s’adressent aux auteurs souhaitant :

* Partager et faciliter l’utilisation de leur création par d’autres
* Autoriser gratuitement la reproduction et la diffusion (sous certaines conditions)
* Accorder plus de droits aux utilisateurs en complétant le droit d’auteur qui s’applique par défaut
* Faire évoluer une œuvre et enrichir le patrimoine commun (les biens communs ou Commons)
* Économiser les coûts de transaction
* Légaliser le peer to peer de leurs œuvres.

Les licences Creative Commons sont fondées sur le droit d’auteur. Alors que le régime du droit d’auteur classique vous incite à garder l’exclusivité sur la totalité de vos droits (« tous droits réservés »), ces licences vous encouragent à n’en conserver qu’une partie (« certains droits réservés »). Creative Commons travaille avec des experts en droit d’auteur dans le monde entier pour que ces licences soient valides quelle que soit la juridiction. Ces licences permettent au public d’utiliser vos œuvres, sous certaines conditions, selon vos préférences. Les licences sont modulables et existent sous 3 formes :

1. Un résumé explicatif destiné aux utilisateurs non-juristes, il décrit de manière simple les actes que le public a le droit d’effectuer sur l’œuvre.
2. Un contrat destiné aux juristes.
3. Une version en code informatique, permettant d’établir un lien vers le résumé et d’associer des métadonnées à l’œuvre.

Conditions communes à toutes les licences Creative Commons :

* Offrir une autorisation non exclusive de reproduire, distribuer et communiquer l’œuvre au public à titre gratuit, y compris dans des œuvres dites collectives.
* Faire apparaître clairement au public les conditions de la licence de mise à disposition de cette création, à chaque utilisation ou diffusion.
* Chacune des conditions optionnelles peut être levée après l’autorisation du titulaire des droits.
* Les exceptions au droit d’auteur ne sont en aucun cas affectées.
* Il est interdit d’utiliser des mesures techniques contradictoires avec les termes des licences.
* Le partage de fichiers (*peer to peer*) n’est pas considéré comme une utilisation commerciale.

Comment faire pour placer vos œuvres sous l’une des licences Creative Commons ?

La procédure se fait en ligne, il n’y a aucun document à signer. Si vous êtes auteur, ou avec l’accord des titulaires de droits, vous pouvez simplement choisir votre licence parmi les 6 combinaisons d’options existantes en répondant à quelques questions sur l’interface mise à disposition par [www.creativecommons.fr](http://www.creativecommons.fr/) (d’où provient le paragraphe 19.3.1.2).

Lorsque vous sélectionnez une licence Creative Commons, vous obtiendrez un morceau de code html/rdf qui peut facilement être inséré sur une page web. Ce code reproduira sur le site le logo Creative Commons avec un lien vers la version résumée de la licence sélectionnée. Vous pouvez insérer à côté de ce logo une phrase pour expliquer que les œuvres placées sur votre site sont sous l’une des licences Creative Commons.

Toute copie ou communication de l’œuvre au public doit être accompagnée de la licence selon laquelle elle a été mise à la disposition du public, ou d’un lien vers le texte de cette licence. Certains formats peuvent être marqués directement (<http://creativecommons.org/technology/usingmarkup>).

Creative Commons propose gratuitement six licences qui permettent aux titulaires de droits d’auteur de mettre leurs œuvres à disposition du public à des conditions prédéfinies. Les licences Creative Commons viennent en complément du droit applicable, elles ne se substituent pas au droit d’auteur.

Simples à utiliser et intégrées dans les standards du web, ces autorisations non exclusives permettent aux titulaires de droits d’autoriser le public à effectuer certaines utilisations, tout en ayant la possibilité de réserver les exploitations commerciales, les œuvres dérivées et les conditions de redistribution.

##### Les options des licences CC

Les auteurs ou titulaires des droits d’auteur peuvent choisir un ensemble de conditions qu’ils souhaitent appliquer à leurs œuvres :

ATTRIBUTION : Toutes les licences Creative Commons obligent ceux qui utilisent vos œuvres à vous créditer de la manière dont vous le demandez, sans pour autant suggérer que vous approuvez leur utilisation ou leur donner votre aval ou votre soutien.

PAS D’UTILISATION COMMERCIALE : Vous autorisez les autres à reproduire, à diffuser et (à moins que vous ne choisissiez ‘Pas de Modification’) à modifier votre œuvre, pour toute utilisation autre que commerciale, à moins qu’ils obtiennent votre autorisation au préalable.

PARTAGE DANS LES MEMES CONDITIONS : Vous autorisez les autres à reproduire, diffuser et modifier votre œuvre, à condition qu’ils publient toute adaptation de votre œuvre sous les mêmes conditions que votre œuvre. Toute personne qui souhaiterait publier une adaptation sous d’autres conditions doit obtenir votre autorisation préalable.

PAS DE MODIFICATION : Vous autorisez la reproduction et la diffusion uniquement de l’original de votre œuvre. Si quelqu’un veut la modifier, il doit obtenir votre autorisation préalable.

##### Les licences CC

Ces quatre options peuvent être arrangées pour créer six licences différentes, les six licences Creative Commons :

1. Attribution (BY) : Le titulaire des droits autorise toute exploitation de l’œuvre, y compris à des fins commerciales, ainsi que la création d’œuvres dérivées, dont la distribution est également autorisée sans restriction, à condition de l’attribuer à son l’auteur en citant son nom. Cette licence est recommandée pour la diffusion et l’utilisation maximale des œuvres.

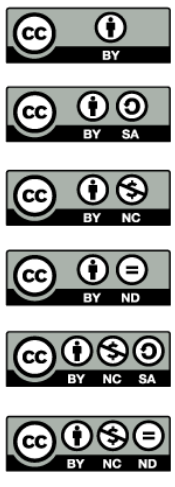
2. Attribution + Pas de Modification (BY ND) : Le titulaire des droits autorise toute utilisation de l’œuvre originale   
(y compris à des fins commerciales), mais n’autorise pas la création d’œuvres dérivées.

3. Attribution + Pas d’Utilisation Commerciale + Pas de Modification (BY NC ND) : Le titulaire des droits autorise l’utilisation de l’œuvre originale à des fins non commerciales, mais n’autorise pas la création d’œuvres dérivés.

4. Attribution + Pas d’Utilisation Commerciale (BY NC) : le titulaire des droits autorise l’exploitation de l’œuvre, ainsi que la création d’œuvres dérivées, à condition qu’il ne s’agisse pas d’une utilisation commerciale (les utilisations commerciales restant soumises à son autorisation).

5. Attribution + Pas d’Utilisation Commerciale + Partage dans les mêmes conditions (BY NC SA) : Le titulaire des droits autorise l’exploitation de l’œuvre originale à des fins non commerciales, ainsi que la création d’œuvres dérivées, à condition qu’elles soient distribuées sous une licence identique à celle qui régit l’œuvre originale.

6. Attribution + Partage dans les mêmes conditions (BY SA) : Le titulaire des droits autorise toute utilisation de l’œuvre soriginale (y compris à des fins commerciales) ainsi que la création d’œuvres dérivées, à condition qu’elles soient distribuées sous une licence identique à celle qui régit l’œuvre originale. Cette licence est souvent comparée aux licences « copyleft » des logiciels libres. C’est la licence utilisée par Wikipédia.



1. Exemple de logos Creative Commons[[11]](#endnote-10)

# Webographie

1. Gaillard, R. (2014). De l’Open data à l’Open research data : quelle(s) politique(s) pour les données de recherche ? (Mémoire de fin d’études de conservateur de bibliothèque, Enssib, Lyon, France). http://www.enssib.fr/bibliotheque-numerique/documents/64131-de-l-open- data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche.pdf [↑](#footnote-ref-1)
2. [http://hubeau.eaufrance.fr](http://hubeau.eaufrance.fr/) – date d’accès avril 2018. [↑](#endnote-ref-1)
3. <https://www.data.gouv.fr/fr/> – date d’accès avril 2018. [↑](#endnote-ref-2)
4. <https://www.force11.org/group/fairgroup/fairprinciples> - date d’accès avril 2018 [↑](#endnote-ref-3)
5. <https://services.renater.fr/federation/technique/attributs> - date d’accès avril 2018. [↑](#endnote-ref-4)
6. <https://services.renater.fr/federation/index> - date d’accès avril 2018 [↑](#endnote-ref-5)
7. <https://www.etalab.gouv.fr/licence-ouverte-open-licence> - date d’accès avril 2018. [↑](#endnote-ref-6)
8. <https://opendatacommons.org/licenses/odbl/summary/> - originale – date c’accès avril 2018. [↑](#endnote-ref-7)
9. <https://vvlibri.org/fr/licence/odbl-10/legalcode/unofficial> - version française - date d’accès avril 2018. [↑](#endnote-ref-8)
10. <http://creativecommons.fr/licences> - date d’accès avril 2018 [↑](#endnote-ref-9)
11. Ce logo renvoie vers la licence située à <https://creativecommons.org/licenses/by/4.0/> - date d’accès avril 2018 [↑](#endnote-ref-10)