



AgGate

Portail de données
pour l'innovation en agriculture

MISSION
Jean-Marc BOURNIGAL

Octobre 2016

SOMMAIRE



Remerciements	5
Membres du comité de pilotage	6
Liste des personnes rencontrées	7
Résumé	9
Lexique	15
Introduction	16
1_ Contexte et enjeux : le développement de l'agriculture numérique	17
1.1_ La transition numérique, levier d'une agriculture multi-performante répondant aux enjeux de l'agro-écologie	18
1.2_ De plus en plus de données générées	19
1.3_ Le portail de données agricoles	19
1.4_ Revue des dispositifs existants en lien avec l'ouverture de données agricoles	22
2_ Gouvernance et structure juridique	29
3_ Les jeux de données à intégrer	31
3.1_ La démarche	32
3.2_ Origine des données	32
3.3_ Caractérisation des données et métadonnées	34
3.4_ Accessibilité des données	34
3.5_ Les aspects juridiques de la donnée	34
4_ Au cœur du portail, le guichet unique	39
4.1_ Utilisateurs et usages du guichet	40
4.2_ Fonctionnalités communes à tous les utilisateurs	43
4.3_ Fonctionnalités pour le consommateur de données	44
4.4_ Fonctionnalités pour le fournisseur de données	46
4.5_ Fonctionnalités d'administration du portail et gestion des droits	48

5_	Les autres fonctionnalités indispensables au portail	51
5.1_	« Cloud agriculteurs » : espace d'hébergement des données privées des agriculteurs	52
5.2_	Le « magasin numérique »	53
5.3_	Espace vitrine des outils et services à l'agriculture et d'évaluation participative	53
5.4_	Espace d'échange entre acteurs de la profession agricole	54
5.5_	Conclusion	55
6_	Points techniques critiques	57
6.1_	Recherche des données, intégration, métadonnées et sémantique	58
6.2_	Interconnexion	62
6.3_	Anonymisation	63
6.4_	Sécurisation	66
6.5_	Ergonomie, design et charte graphique	66
7_	Animation du projet	67
7.1_	L'équipe d'animation	68
7.2_	Animer l'écosystème de l'innovation	68
7.3_	Convaincre les producteurs et fournisseurs de données	69
7.4_	Liens avec la recherche	70
8_	Mise en œuvre	73
8.1_	Calendrier	74
8.2_	Evaluation économique préliminaire	75
9_	Conclusion	79
9.1_	Les enjeux du portail	80
9.2_	Facteurs de succès et risques principaux	80
9.3_	Recommandations	83
10_	ANNEXES	85
10.1_	Annexe 1 - Lettre de mission pour JM Bournigal	86
10.2_	Annexe 2 – Questionnaire pour la caractérisation des bases de données agricoles	89
10.3_	Annexe 3 - Extrait de l'inventaire des bases de données à caractère agricole	95
10.4_	Annexe 4 - Liste des métadonnées pressenties	120
10.5_	Annexe 5 - Listes des standards, ontologies, schémas et référentiels recommandés	122
10.6_	Annexe 6 - La structure juridique d'un portail de données	123

Remerciements

La rédaction de ce rapport est le fruit d'un travail collectif conduit par une équipe mise en place au sein d'Irstea :

- Véronique Bellon-Maurel, Directrice du Département Ecotechnologies
- Pascal Odot, Directeur des affaires juridiques
- Laurence Freyt-Caffin, Adjointe au directeur des affaires juridiques
- Céline Delmas, Juriste
- Emmanuelle Jannes-Ober, Directrice adjointe déléguée à l'IST
- et Charles Moszkowicz de la société ENEO à laquelle il a été fait appel pour une assistance à maîtrise d'ouvrage dans la réalisation de la mission.

Merci à chacun d'entre eux pour leur implication et contribution, ainsi qu'à Pierrick Givone Directeur Général Délégué à la Recherche et l'Innovation d'Irstea pour ses conseils avisés, à la Direction de la Prospective de la Veille, de la Valorisation de l'Information Scientifique et Technique pour son apport documentaire, ainsi qu'aux membres du comité de pilotage de la mission.

Je tiens à remercier les parties prenantes - éditeurs de logiciels, agro-équipementiers, agriculteurs, chambres d'agriculture, instituts techniques, organismes de recherche, ministères - réunies en atelier le 22 juin dernier.

Plus généralement mes remerciements vont à l'ensemble des personnes et organismes ayant contribué à cette mission lors d'un entretien ou en nous aidant à affiner notre vision. Je tiens à adresser toute ma gratitude à plusieurs personnes avec lesquelles les interactions ont été répétées et fécondes (Catherine Migault et Isabelle Serrot, du Crédit Agricole, Muriel Monteiro et Natalia Danon-Boileau, de la société Bearing Point, Pierre-Olivier Gibert de la société Digital Ethics, Gaylord Marchesseau de la société OpenDataSoft), mais également celles qui nous ont aidées à dresser l'inventaire des données utilisables (Laurent PAVARD, Président du conseil des systèmes d'information au Ministère de l'agriculture, de l'agroalimentaire et de la forêt, Agnès Salaün, chargée de mission chez Arvalis, GIS GC HP2E).

Enfin, merci aux spécialistes de la donnée qui nous ont éclairés sur les subtilités de ces nouveaux territoires et sans lesquels nous aurions pu nous perdre, en particulier Romain Tales, mission ETALAB, Patrick Valduriez et Marie-Laure Mugnier de l'Inria, Catherine Roussey, François Pinet et Jean-Pierre Chanet d'Irstea à Clermont-Ferrand.

Membres du comité de pilotage

Noms	Prénoms	Fonctions	Organismes
BADUEL	Valérie	Inspectrice générale de la santé publique vétérinaire	DGER (MAAF)
BAILLE	Annick	Conseillère pour l'enseignement agricole, la recherche et les relations sociales	Cabinet agriculture
BELLON-MAUREL	Véronique	Directrice du département Écotecnologies	Irstea
BOURNIGAL	Jean-Marc	Président	Irstea
DUBARRY	Cécile	Chef du service des nouvelles technologies de l'information et de la communication	DGE (MAAF)
KAO	Cyril	Sous-directeur de la recherche, de l'innovation et des coopérations internationales	DGER (MAAF)
MAHE	Sylvain	Chargé de mission	Direction générale de la recherche et de l'innovation (MENESR)
METEYER	Hervé	chef du bureau Logiciel	DGE (MAAF)
MONTUSCLAT	Sébastien	responsable sectoriel numérique	BPI France
NORDEN	Gary	Chargé de mission	DGE (MAAF)
ODOU	Grégoire	Conseiller auprès de la Secrétaire d'Etat au numérique	Ministère de l'Economie, de l'Industrie et du Numérique
TALES	Romain	Responsable du recensement des données publiques	Etalab
TURENNE	Julien	Chef du service compétitivité et performance environnementale	DGPE (MAAF)
VERGES	Elisabeth	Chef du Secteur environnement, agronomie, écologie, sciences du système Terre et de l'univers	DGRI (MENESR)
VOYATZAKIS	Ariane	Responsable du secteur Agroalimentaire	BPI France

Liste des personnes rencontrées

Noms	Prénoms	Fonctions	Organismes
BABINET	Gilles	Responsable de l'économie numérique	Commission européenne
BARBE	Eric	Directeur adjoint UMR TETIS	Irstea
BARDOU	Aida	Commerciale	Orange
BARRAQUE	Gabrielle		FNSafer
BEULIN	Xavier	Président	FNSEA
BOSCHET	Agnès		WIKANE
BRAULT	Michel	Directeur général	CCSMSA
BUCHE	Patrice	UMR IATE	INRA
CHAPALAIN	Catherine	Directrice générale	ANIA
CORBEAU	Bertrand	Directeur général adjoint	Crédit Agricole S.A.
DANON-BOILEAU	Natalia		Bearing Point
DAVID	Xavier	Directeur	ALLICE
DELAGE-PAGANINI	Fabrizio	Account manager	SIGFOX
DESPREY	François	Directeur général	GS1 France
DUBARRY	Cécile	Chef de service de l'économie numérique	MEIM
DUBOURG	Régis	Directeur général	APCA
DURAND	Patrice	Directeur	FNEDT
FAURAX	Clément	Directeur général adjoint	FNSEA
FERRAND	Rémi	Délégué aux affaires publiques	SYNTEC
FOG HANSEN	Nicolai	Digital Business Manager	Seges
GIBERT	Paul-Olivier	Président	CEO Digital Ethics
GIRARD	Jean-Philippe	Président	ANIA
GOZAL	Muriel	Directrice générale	FNSafer
HAEZEBROUCK	Théo-Paul	Chargé de mission - Projets numériques, Animateur plateforme API-AGRO	ACTA
HERMAN	Mathias	Technico-commercial	Orange
HUNAUULT	Jean-Louis	Président	SIMV
HUYGHE	Christian	directeur scientifique adjoint Agriculture	INRA
LAURENT	François	membre du Directoire Opérationnel du GIS GC HP2E	Arvalis
LE PALLEC	Sylvie	Responsable des relations institutionnelles	GS1 France
LEGRAIN	Claire	Chef du Service Entreprises et Marchés	FranceAgriMer
LEVEAU	Valérie	animatrice du groupe « Connaissance des pratiques » du GIS GC HP2E	Arvalis

LHERMITE	Serge	Délégué aux IAA	MAAF
LION	Catherine	Directrice générale	FNSEA
LONGO	Aldo	Directeur - Direction générale de l'Agriculture et du développement rural –	Commission européenne
MARCEL	Stéphane	Directeur général	SMAG
MARCHESSEAU	Gaylord	Account manager	OpenDataSoft
MARTEL	Laurent	DG Pôle Agriculture	INVIVO
MIGAULT	Catherine	Responsable des affaires agricoles	FNCA
MOLLER	Jens	Directeur	DKE
MONTEIL	Jean-Luc	Directeur général	FNCUMA
MONTEIRO	Muriel		Bearing Point
MUGNIER	Marie-Laure	Directrice GraphIK	LIRMM-INRIA
MUNIER-JOLAIN	Nicolas	Animateur Agrosyst	INRA Dijon
PELC	Alain	Directeur des études des répertoires et des statistiques	CCSMSA
PERAULT	Régine	Commerciale	Orange
PILLAUD	Hervé	Agriculteur-éleveur	--
RIVIERE	Loic	CEO	TechInFrance
ROUDAUT	Françoise	Directrice des Systèmes d'Information	Irstea
SALAUN	Agnès	Chargée de mission « Connaissance des pratiques » du GIS GC HP2E	Arvalis
SINE	Mehdi	Responsable Agriculture Numérique	Arvalis
SUCEVIC	Perica	Conseiller juridique	Secrétariat général pour la modernisation de l'action publique (SGMAP)
TALES	Romain	Responsable data.gouv.fr	ETALAB
THATCHER	Mary Kay	Responsable AgDataCoalition	Farm Bureau (US Agriculture Department)
THOMOPOULOS	Rallou	Chercheur	INRA, UMR IATE
TIFFIN	Richard	Chief Scientific Adviser Executive Management Team	Agrimetrics
VALDURIEZ	Patrick	Directeur de recherche	Inria
VINE	Pascal	Délégué général	Coop de France
WAINSTAIN	Jérémié	Président	TheGreendata

À cette liste s'ajoutent toutes les personnes interviewées pour établir l'inventaire des bases de données (cf Annexe 3).

Résumé

Suite aux recommandations de la Mission Agriculture Innovation 2025, les ministres de l'agriculture, de la recherche et de l'économie ont confié à Jean Marc Bournigal¹ une mission visant d'une part à préparer la mise en place opérationnelle d'un portail de données agricoles, d'autre part à proposer un plan d'action pour favoriser l'émergence de nouveaux services à partir des données du portail.

Comme dans les autres secteurs économiques, il y a en agriculture un enjeu fort à collecter et à traiter des données en vue d'en tirer des informations inédites et des modèles innovants sur la base desquels développer des services nouveaux. Cette opportunité n'a pas échappé non seulement aux géants de l'agrofourmiture qui cherchent à intégrer l'ensemble de la chaîne, de la collecte de données à la vente de produits accompagnés de services, mais aussi aux GAFAS qui commencent à s'intéresser à l'agriculture, nouveau domaine d'action après le commerce, la banque et la santé. La concentration de telles ressources dans les mains de quelques acteurs génère des inquiétudes quant à l'indépendance des agriculteurs, voire la souveraineté alimentaire. C'est pourquoi, il est urgent de proposer une alternative, dans laquelle la profession agricole reste maître des données qu'elle produit, et qui facilite leur valorisation, via un schéma d'innovation ouverte à de multiples acteurs. Cette stratégie d'innovation ouverte permet d'opposer à la puissance de grands opérateurs la multiplicité et l'agilité de petites sociétés dans un écosystème dans lequel concurrence et synergies seront favorisées.

C'est pourquoi il a été proposé de créer un portail de données à vocation agricoles, avec l'objectif de faciliter l'accès à des données d'intérêt, de différentes natures et interconnectables, issues du secteur public et du secteur privé. L'accès à ces données accélèrera les processus de développement de connaissances, de modèles, et in fine de services à l'agriculture.

Lors de cette mission visant à instruire la mise en place opérationnelle de ce portail, ont été plus particulièrement étudiés: la gouvernance, la structure juridique, les jeux de données à mobiliser, les fonctionnalités attendues, les points techniques critiques (sécurisation, accès, interconnexion, anonymisation...), les besoins de recherche en accompagnement, le modèle économique, le plan de financement et le calendrier prévisionnel... en liaison avec la profession agricole. Après l'analyse des facteurs de risque et des facteurs de réussite, des propositions concrètes sont faites. Cette synthèse présente les points essentiels de ce travail.

¹ Lettre de mission du 15 Avril 2016 (cf Annexe 1)

Les fonctionnalités : le « guichet unique » au cœur du portail

Si le cœur du portail est le guichet unique pour accéder aux données d'intérêt, nous avons identifié d'autres fonctionnalités susceptibles d'intéresser la profession. Les cinq fonctionnalités identifiées à ce jour sont les suivantes :

- 1- Un « Guichet Unique » pour l'accès à des jeux de données d'intérêt agricole, publiques et privées, qui restent stockées sur les plateformes de leurs propriétaires ; ce guichet assurera l'inventaire, le référencement, la recherche et le croisement des données, qui peuvent être ouvertes sous conditions, par exemple avec des licences spécifiques ou contre rétribution ;
- 2- Un « Magasin Numérique » pour un retour rapide vers les fournisseurs de données ; cet espace offrira des outils simples de traitement (statistiques, géostatistiques) ou de présentation des données (cartographie...) ; il pourra donner accès à des référentiels par région, par secteur, par type d'exploitation, services gratuits destinés aux agriculteurs ;
- 3- Un "Cloud Agriculteurs" pour héberger les données professionnelles des agriculteurs ;
- 4- Un « Espace d'échanges » entre acteurs de la profession agricole, pour partager expériences, savoir-faire, idées, etc. et pour créer une dynamique d'innovation ;
- 5- Une « Vitrine d'OAD » assortie d'une évaluation participative, espace dans lequel des Outils d'Aide à la Décision (OAD) et les services existants sur le marché seront présentés et évalués ou commentés par les utilisateurs (un concept dénommé "TripAdvisor des OAD »).

Ces différentes fonctionnalités visent à susciter l'intérêt des agriculteurs et des fournisseurs vis-à-vis du portail, et contribuent à son modèle économique. Elles ont été identifiées suite aux entretiens avec des représentants de la profession agricole, à un atelier de 50 utilisateurs potentiels du portail (le 22 Juin 2016), et à un benchmark sur des initiatives approchantes. Parmi la demi-douzaine de dispositifs approchants, le benchmark n'a révélé aucune initiative de cet ordre dans le monde. Les structures ont des points communs avec le portail : le Seges (Danemark), société gouvernée par les agriculteurs, qui collecte leurs données et les valorise sous forme de modèles et de services, le projet API-AGRO (France), lancé par l'ACTA et les ITA, qui ouvre des référentiels agricoles à ses membres via des API, la mission Etalab, qui a le même objectif de réutilisation des données mais qui diffère quant aux données offertes (exclusivement publiques et non spécifiquement agricoles), la AgDataCoalition (USA), dont l'objectif unique est aujourd'hui de collecter les données des agriculteurs pour les sécuriser, le système d'information AGROSYST, géré par l'INRA, qui collecte les données des fermes du réseau Dephy et les réutilise dans des projets de recherche. Si chaque dispositif a un ou deux points de convergence avec le portail, aucune ne conjugue simultanément « collecte de données à vocation agricole », « ouverture », « interconnexion » et « animation du portail ». La gouvernance n'est confiée aux agriculteurs que dans deux cas (le Seges et la AgDataCoalition).

Structure et gouvernance : la profession agricole aux commandes d'un outil mutualisé inédit

L'enjeu principal du portail (mobiliser les données des exploitations agricoles pour en assurer la valorisation au service de l'agriculture) milite pour une gouvernance privilégiant la confiance et la mobilisation du monde agricole. Ainsi, il est essentiel que ce projet soit porté par un consortium de quelques acteurs économiques du monde agricole (coopératives, filières, banques, assurances, structures d'investissement...). Pour assurer son opérationnalité, une structure de type SAS est préconisée ; si la mutualisation est effective (à savoir si aucune grande entreprise ne détient plus de 25% du capital), celle-ci pourra être reconnue « pôle d'innovation » et bénéficier de la part de l'État d'une avance remboursable allant jusqu'à 50% des coûts de mise en place. Pour sa gouvernance stratégique, un comité d'orientation stratégique garant des conditions d'utilisation, de la déontologie et de l'éthique, mais aussi de l'animation de l'écosystème d'innovation, en lien avec les professions concernées devra rassembler a minima, les représentants des syndicats agricoles, l'APCA, COOP de France, la CNMCCA, des représentants du secteur de l'agrofourniture (AXEMA, UFS, UIPP, UNIFA ; l'ANELA), des représentants de la recherche (INRA, Irstea, ACTA), des représentants des ministères en charge de de l'agriculture et du numérique, et des personnalités qualifiées du secteur numérique. L'équilibre de ce comité comme les principes de la création du portail pourraient utilement être validés lors d'un comité d'orientation stratégique du ministère en charge de l'agriculture.

Les données à mobiliser : bâtir une gouvernance des données et une protection juridique appropriée pour exploiter des gisements immenses

Le portail a pour objet de mettre à disposition du public des données et services numériques. Créer un portail d'une telle nature suppose, d'une part, de respecter un certain formalisme juridique, incluant la mise en ligne de mentions légales sur le site et, d'autre part, de protéger juridiquement les données accessibles sur le site, notamment par la mise en place d'une gouvernance des données, via un cadre contractuel définissant les modalités de mise à disposition et de réutilisation des données publiques ou privées qui alimenteront le portail, pour chaque utilisateur privé ou public. Une attention particulière doit être portée à la gouvernance des données. Deux dispositifs sont prévus: les Conditions Générales (CG), qui s'adressent aux fournisseurs de données et édictent les règles liées au dépôt, et les Conditions Générales d'Utilisation (CGU), qui s'adressent aux utilisateurs de données. Un fournisseur de données souhaitant rendre ses données accessibles devra préalablement déterminer les droits qu'il souhaite accorder sur ses données, en choisissant le type de licence approprié. Les données agricoles peuvent, pour partie, être géo-référencées (parcellaire), et donc être raccordables à une personne physique, ce qui nécessite son accord avant diffusion. Les données seront, sauf accord du titulaire, anonymisées avant leur diffusion.

Un inventaire des données à mobiliser a été dressé. Il a permis d'identifier un premier socle d'une soixantaine de bases « qualifiées » (c'est-à-dire bien décrites du point de vue des métadonnées), publiques ou privées, le double avec des bases non décrites. Cet inventaire devra être régulièrement mis à jour.

Les points-clés techniques : concilier performance du moteur de recherche, anonymisation et sécurité

Un cahier des charges a été dressé pour lister les fonctions nécessaires aux 3 acteurs du portail : les administrateurs, les fournisseurs et les utilisateurs de données. Au-delà de ce cahier des charges qui décrit des fonctions assez classiques de portails de données, il convient d'insister sur les « point durs » liés à la nature particulière des données et à l'ambition du portail.

Le cœur du portail est le guichet unique et la recherche de données. L'approche préconisée, pour conjuguer rapidité de mise en place et efficacité du portail, est la fédération des bases de données distribuées, c'est-à-dire de données qui resteront stockées chez les différents fournisseurs. Cette approche, dite « médiateur », consiste à créer un dispositif de « traduction » de chaque base de données (l'adaptateur) qui soit facilement interrogeable par un utilisateur, via une interface unique (le médiateur). Vu de l'utilisateur, le système est « transparent », comme s'il n'y avait qu'une seule base de données. Cela exige de connaître les métadonnées et les schémas de données des différentes bases, indispensables à une recherche efficace, et donc d'avoir une collaboration active de la part des fournisseurs de données (éditeurs de logiciels, agro-équipementiers, services de l'Etat...). Dans une deuxième version, la performance du portail pourra être accrue en mettant en place des ontologies, rendant plus précise l'identification des données.

Un deuxième point essentiel du portail est l'anonymisation des données, lorsque celle-ci est réclamée. De nombreuses informations d'intérêt sont liées au parcellaire (géo-référencées), ce qui peut mener à identifier le producteur de la donnée. Plusieurs techniques d'anonymisation devront être déployées en fonction de l'usage qui sera fait des données : accroissement de la granularité, confidentialité différentielle, brouillage (par décalage aléatoire) des coordonnées géographiques. L'anonymisation est donc possible et doit être mise en œuvre à la demande du producteur, mais elle risque de diminuer fortement l'intérêt des données. Ainsi, il s'agira en parallèle de sensibiliser la profession agricole sur l'enjeu de ne garder confidentielles que les données les plus stratégiques.

En lien avec l'anonymisation, l'interconnexion des données est le procédé qui va permettre au portail de retourner à l'utilisateur une combinaison de données de natures différentes et issues de sources différentes (par exemple données météo, données pédologiques, données de rendement), sur la base de leurs identifiants spatiaux (géoréférencement) et temporels. D'une part, le portail doit être capable d'identifier ces données dans différentes bases via l'approche médiateur, ensuite, il doit être capable de les fusionner. L'interconnexion pourrait être un service monnayable du portail.

Enfin, la sécurité est un point crucial. L'approche « bases de données distribuées » est un atout car elle évite de gérer la sécurité des bases de données, qui relève de chacun des fournisseurs. En revanche, la mise en place d'une « cloud agriculteurs » nécessitera de considérer cette question. L'accès au système devra être sécurisé et ne pas présenter de faille de sécurité connue, avec des mises à niveau régulières du dispositif de sécurité. Les échanges monétaires devront être tout particulièrement sécurisés.

L'animation, essentielle pour créer la confiance et « acquérir du trafic »

Ce portail de données est une initiative unique au monde pour favoriser l'innovation ouverte au service de l'agriculture. Il est en accord avec la vision fédératrice des données, qui considère que les données acquièrent une valeur nouvelle (au-delà de celle pour laquelle elles ont été collectées) lorsqu'elles sont regroupées et interconnectées, car susceptibles de réutilisation et de création de connaissance via des traitements appropriés. Cette vision est en rupture avec une vision protectionniste des données, qui considère la donnée comme un bien matériel, qui perd de la valeur lorsqu'on la réutilise. Il est donc indispensable d'accompagner la construction technique du portail par une animation visant à sensibiliser les producteurs et fournisseurs de données à ce nouveau paradigme et à attirer les spécialistes du traitement de la donnée (data sciences) vers les enjeux de l'agriculture.

Un premier niveau d'animation consiste à une communication efficace et une collaboration avec les fournisseurs et les producteurs de données : co-construction des fonctionnalités du portail dans un approche de design thinking, information régulière sur les nouveautés du portail (par exemple dans le magasin numérique), statistiques de production pour les agriculteurs, voire activités plus ludiques (concours, crowd-sourcing...). Un canal privilégié pour construire la confiance est celui des organisations agricoles, que le portail devra largement associer à l'animation.

Côté « utilisateurs de données », un programme sera élaboré pour générer rapidement des applications emblématiques à partir des données du portail, faisant ainsi pour l'agriculture la démonstration de l'intérêt de la démarche. Il est préconisé de mettre en œuvre un plan d'action public-privé pour accompagner le développement du portail, sur la base d'actions dans la dynamique « French Tech » thématisée autour de l'agriculture : appels à projets, concours de startups, hackathons, compétition numérique, challenge numérique, etc. Des fonds spécifiques devront être levés pour accompagner ces événements.

Enfin, il s'agit d'inscrire le portail dans une perspective d'amélioration continue, soutenue par une recherche dédiée, à financer par la structure porteuse du portail. Deux champs de questions ont été identifiés : la question des ontologies - essentielles pour améliorer la performance du moteur de recherche - et celle de la chaîne de valeur de la donnée (agricole). Outre ces recherches destinées à préparer les versions futures, le portail sera en relation directe avec des grands programmes ou projets de recherche sur des thèmes connexes : le programme CASDAR du MAAF, l'institut Convergences en agriculture numérique #DigitAg, des Laboratoires d'Innovation Territoriale. Ces projets seront soit sources de données nouvelles, soit utilisateurs de données pour créer des OAD et services promus sur le portail.

Construction du portail : calendrier et modèle économique

Le temps nécessaire à la mise en place du portail est estimé à 3 ans avec une première phase de construction (12-15 mois) et une phase de lancement (21-24 mois). La « phase de construction » est destinée à la mise en place de l'équipe, à la construction informatique des modules pour les fonctionnalités centrales (a minima le « guichet unique ») et aux actions d'animation (challenges, hackathons) dans l'objectif d'avoir des illustrations d'application à placer sur le portail dès son ouverture. La « phase de lancement » a vocation à développer les modules complémentaires et à

assoir l'acquisition de trafic. Dans cette première phase, l'accès aux données publiques très nombreuses dans le domaine agricole sera déterminant.

Les ressources nécessaires à la mise en place du portail sont des ressources humaines pour l'implémentation (on estime qu'une équipe de 5 personnes est suffisante), des apports sous forme de prestations (infrastructure, assistance juridique, business plan, communication ...), les actions de recherche pour préparer les versions ultérieures. La mise en place du portail nécessite un investissement de 4,8 à 5,4 M€ dont 1,5M€ pour l'animation. Le modèle économique devrait s'équilibrer via la contribution des actionnaires, la mobilisation d'aides publiques et des recettes diverses (publicités, stockage de données dans le « cloud agriculteurs », sponsoring privé).

En conclusion, facteurs de réussite et facteurs de risques

Les enjeux-clés pour la réussite du portail sont 1) l'excellence, à savoir la capacité à trouver des données pertinentes, 2) la maîtrise des risques techniques (anonymat, sécurité) et d'image, 3) la maîtrise des coûts et 4) du modèle économique. Un facteur central est celui de l'appropriation et de la confiance : à l'encontre d'une vision protectionniste de la donnée, la marque de ce portail doit être celle de l'ouverture des données pour créer de la valeur avec un retour le plus juste possible vers les acteurs de la chaîne de valeur. C'est pourquoi cette question, qui est aussi un facteur de transparence, doit faire l'objet d'un approfondissement méthodologique particulier. L'implication des organisations agricoles et des fournisseurs de données est essentielle.

Lexique

- **API** : "Application Programming Interface". Une API est une interface de programmation qui permet de se "brancher" sur une application pour échanger des données. Une API est ouverte et proposée par le propriétaire du programme.
- **Donnée publique** : les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission
- **Licence copyleft** : Le copyleft est une méthode générale pour rendre libre un programme (ou toute autre œuvre) et obliger toutes les versions modifiées ou étendues de ce programme à être libres également.
- **Licence ouverte** : La licence ouverte / open licence est une licence libre française créée par la mission ETALAB afin d'encadrer l'ouverture des données de l'État français. Cette licence, présentée le 18 octobre 2011, a été voulue comme une licence compatible avec les licences Open Government Licence (OGL) du Royaume-Uni, Open Data Commons Attribution (ODC-BY) de l'Open Knowledge Foundation et Creative Commons Attribution 2.0 (CC-BY 2.0) de Creative Commons. La Licence Ouverte n'est pas une licence copyleft puisqu'elle n'impose pas d'utiliser une licence identique sur les travaux dérivés des données couvertes.
- **Métadonnées** : Une métadonnée est une donnée servant à définir ou décrire une autre donnée quel que soit son support (papier ou électronique). Un exemple type est d'associer à une donnée la date à laquelle elle a été produite ou enregistrée, ou à une photo les coordonnées GPS du lieu où elle a été prise.
- **Propriétaire de la donnée** : le créateur de la donnée (au sens de l'article L. 111-1 du code de la propriété intellectuelle).
- **Titulaire de la donnée** : le détenteur des droits sur la donnée.

Introduction

Une mission, intitulée « Portail de données agricoles », a été confiée à JM. Bournigal par les Ministres en charge de l'économie, de l'agriculture et de la recherche et de l'enseignement supérieur (cf. Lettre de mission en Annexe 1). Elle fait suite aux recommandations de la Mission Agriculture Innovation 2025 (volet agriculture numérique).

Cette mission doit étudier les points suivants :

1. La structure juridique du portail et le modèle économique associé
2. La gouvernance du portail
3. Les jeux de données nécessaires pour le développement de services innovants
4. La sécurité des données et l'anonymisation
5. Les conditions de développement d'un écosystème innovant autour du portail
6. Un calendrier prévisionnel de mise en œuvre

Pour répondre aux points 3, 4 et en partie 5, un cahier des charges fonctionnel approfondi est proposé dans ce document. Il décrit les fonctionnalités attendues pour la construction du portail d'accès aux données et services d'intérêt agricole, conçu comme l'outil central d'un dispositif complet d'innovation. Le document porte sur la création de ce portail, sa maintenance évolutive ainsi que l'ajout éventuel, *a posteriori*, de nouvelles briques de services.

Contexte et enjeux :



le développement
de l'agriculture numérique



1. Contexte et enjeux : le développement de l'agriculture numérique

1.1 La transition numérique, levier d'une agriculture multi-performante répondant aux enjeux de l'agro-écologie

Comme toute l'économie, l'agriculture entre dans l'ère du numérique et du « big data ». Les enjeux de la transition numérique en agriculture sont nombreux avec un champ des possibles très vastes :

- réduction des coûts et des impacts environnementaux
- réduction de la charge de travail et de la charge mentale, en particulier en élevage, culture pérennes et maraîchage
- modification en profondeur de la relation entre l'agriculteur et la filière aval
- nouvelles approches de traçabilité et de lutte contre la contrefaçon (ex : vin)
- modification en profondeur des pratiques de service technique ou financier à l'agriculteur : conseil technique et/ou aide à la décision stratégique, tactique et opérationnelle, assurance, financement de l'investissement, etc.

Le numérique va donc modifier en profondeur les pratiques de service à l'agriculteur. L'agriculture numérique doit tout à la fois permettre d'accroître la productivité agricole, dans un contexte de forte pression foncière et environnementale, d'évoluer vers l'agro-écologie, et de contribuer au développement économique par l'ouverture de nouveaux marchés. Autre exemple, dans le domaine de l'assurance, la gestion de très grands « lacs de données »² -données d'origines très variées stockées sans structuration- ouvre la voie à l'élaboration de services nouveaux et plus ciblés. Concernant le financement de l'investissement en agriculture, de nouvelles modalités, faisant notamment appel au financement participatif, deviennent également possibles via le numérique.

Cette transition numérique de l'agriculture s'inscrit dans l'évolution numérique de l'ensemble de la société française, et au-delà, dont l'une des manifestations, avec la parution de la LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (dite loi Lemaire) et les débats qui s'ensuivent (encore en cours à la date de rédaction du présent rapport), semble être l'avènement d'un véritable service public de la donnée. A ce titre, il faut souligner la forte attente des opérateurs vis-à-vis de l'accès aux données publiques.

² Entrepôt de données massives non structurées, dont tous les attributs sont conservés, en vue d'une utilisation encore non définie au moment du stockage.

1.2 De plus en plus de données générées

Les données vont être générées de manière massive par un nombre croissant de capteurs et objets connectés, qui se multiplient (80 milliards d'objets connectés sont attendus en 2020 dans le monde, tous secteurs confondus, cf. étude IDATE Digiworld). Fixes (installés dans les parcelles ou les bâtiments) ou mobiles (attachés aux agroéquipements, aux personnes ou aux animaux d'élevage), ils fournissent des données qui complètent les informations satellite, météo, ... comme celles des fournisseurs des exploitations. La conjonction du développement de ces technologies d'acquisition massive de données et des capacités de stockage et de traitement informatique embarqué ou déporté (puissances de calcul importantes accessibles par des réseaux de communication sans fil très haut débit) rend possible l'émergence de nouvelles connaissances, de nouveaux services et d'outils d'aide à la décision (OAD), simples d'utilisation et rapides, améliorant la précision et la pertinence des interventions. L'intérêt est avant tout de valoriser ces données à différentes échelles.

- La mutualisation à grande échelle (nationale ou internationale) peut permettre une valorisation via des travaux de prospective ou d'analyse comparative, la conception d'OAD généralistes visant les marchés globaux ou en appui aux prises de décisions politiques, économiques et stratégiques.
- À l'échelle de l'exploitation et du territoire rural, la valorisation des données visera plutôt l'émergence de services et conseil de conduite des itinéraires techniques, d'optimisation énergétique de l'exploitation et des externalités, ou au choix des agro-produits adaptés aux contraintes et enjeux technico-économiques locaux, sur la base de modèles biotechniques ; la mise en place de systèmes de diagnostic et d'alerte pour une plus grande réactivité face aux pressions parasitaires et environnementales ; une mutualisation et un partage des expériences.
- Ces données peuvent également être valorisées à une échelle plus fine, dans le système de culture, ses parcelles ou lots d'animaux, dans le cadre d'une agriculture de haute précision (gestion des hétérogénéités au sein des parcelles et des lots d'animaux).
- Enfin, l'ensemble des informations collectées peut contribuer à la valorisation des produits en améliorant la traçabilité des produits et des modes de production, mais aussi à l'amélioration des modèles économiques des exploitations comme des filières de valorisation.

1.3 Le portail de données agricoles

1.3.1 Élément essentiel de l'agriculture numérique

Pour accélérer la création de services innovants en agriculture et garantir la meilleure redistribution de la valeur vers les acteurs nationaux, notamment agricoles, le dispositif le plus adapté semble être de fédérer les données via un portail d'accès sous gouvernance partagée entre les acteurs de la profession agricole. Ce portail est intégré dans un projet plus vaste d'innovation numérique dans le domaine agricole, proposé dans le rapport « Agriculture Innovation 2025 » et qui comprend trois volets :

- La mise en place d'un point d'entrée unique pour l'accès aux données à vocation agricole : le portail.
- La création d'un centre interdisciplinaire de recherches sur l'agriculture numérique pour développer des objets et services plus innovants (tels que des outils d'aide à la décision, à différentes échelles, de nouvelles méthodes de travail et d'échange) pouvant se baser sur ce portail. L'institut #DIGITAG, lauréat de l'AAP « Instituts Convergences » (Février 2016), sera créé en 2017 à Montpellier pour renforcer la recherche et l'enseignement dans le domaine de l'agriculture numérique.
- Le développement des capteurs pour l'agriculture et l'élevage (capteurs fiables, économiques, polyvalents, biocapteurs) et les réseaux de communication nécessaires. L'objectif est de mieux caractériser l'état du procédé agricole pour le comprendre et le modéliser dans l'objectif de l'optimiser (composition des plantes, du sol, présence de maladies, infestation par des ravageurs...).

1.3.2 Objectifs du portail

Le Big Data agricole est une masse de données générées par diverses sources (capteurs, machines connectées, données de traçabilité, ...), qui sont aujourd'hui stockées dans de nombreuses bases de données réparties chez différents acteurs du monde agricole, de l'agriculteur, la coopérative, les chambres d'agriculture, les éditeurs de logiciels jusqu'au constructeur de machines, en passant par les services publics. Aucun de ces acteurs n'a la capacité, seul, à valoriser le Big Data. Avec la mise en place du portail, l'enjeu est de créer un écosystème dans lequel chacun des acteurs pourra avoir un rôle pour mieux valoriser la donnée, en la réutilisant ; ces derniers seront en outre incités à concevoir des outils de production de la donnée pensés dès le départ avec cet objectif de réutilisations multiples.

L'objectif du portail est de permettre l'accès, l'utilisation et l'exploitation d'un ensemble important de données produites par et pour tous types d'acteurs directement ou indirectement liés au monde agricole. Il permettra de recenser et mettre à disposition ces données, publiques ou privées, qui sont stockées sur les plateformes de leurs propriétaires. Cela concerne aussi bien des données publiques, par exemple la plateforme nationale www.data.gouv.fr (développée par ETALAB), que des données privées, qui seront rendues accessibles sous conditions, dans la mesure où leur propriétaire accepte leur partage.

Le portail devra aussi mettre en avant et donner accès à des OAD et services dédiés, et permettre la mise en relation entre les acteurs de la filière agricole.



À terme, ce portail pourrait contribuer à la création d'un dispositif européen par interconnexion d'une partie des données nationales des pays de l'UE. Cette option devra être prise en compte dès la création du portail.

1.3.3 Les attendus : 5 briques fonctionnelles à implémenter

Le portail sera développé sur 3 ans. Cinq briques fonctionnelles ont été identifiées, elles seront mises en place en plusieurs lots, en fonction de la demande des utilisateurs, avec l'objectif d'une acquisition rapide de trafic. Le portail agricole sera une plateforme web accessible sur les navigateurs couramment utilisés (Internet Explorer, Firefox, Chrome, Safari, et versions « mobiles »). La figure 1 donne un rapide aperçu des attendus, qui seront détaillés dans la suite du document. Le portail devra démarrer avec la brique fonctionnelle « Guichet unique », qui est sa fonction n°1, accompagnée d'une ou deux autres fonctionnalités destinées à acquérir rapidement du trafic (c'est-à-dire attirer un maximum d'utilisateurs sur le portail).

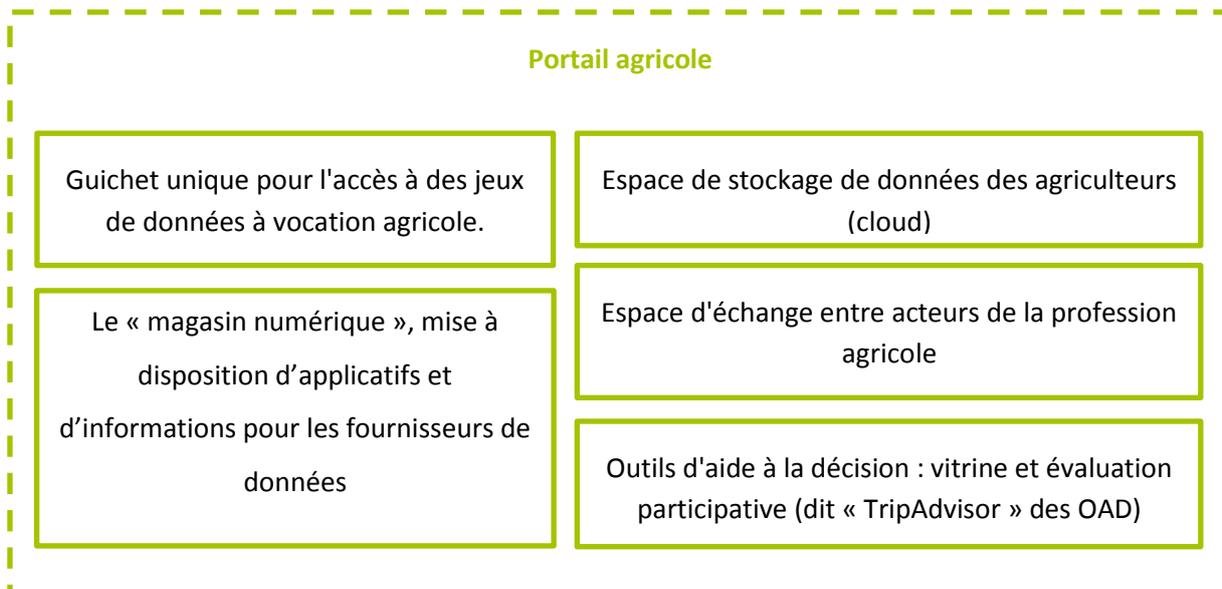


Figure 1 - Schéma des principales fonctionnalités attendues du portail agricole

- **Le guichet unique**

Il s'agit en priorité de proposer un guichet unique, un espace permettant de référencer, de rechercher et de télécharger des données d'intérêt agricole diverses. Différents acteurs auront ainsi la possibilité d'exploiter et de valoriser les données disponibles : entreprises de data science, de conseil, institutions publiques, recherche, associations, voire agriculteurs et citoyens etc. L'accès aux données sera ouvert sous conditions (financières et de licences) en fonction de l'origine des données. Le guichet unique sera en relation avec d'autres plateformes françaises (et à termes européennes) de données ouvertes et/ou à vocation agricole voire agro-alimentaire.

- **Le « magasin numérique »**

Le « magasin numérique » est un espace qui permettra des retours rapides vers les fournisseurs de données, sous forme de petits outils, d'agrégations de résultats mis en forme avec une visée informative et attractive, de traitements très primaires mais immédiatement disponibles des données du portail, d'informations complémentaires utiles. Ce pourrait être par exemple la visualisation de données géographiques sur une carte, le calcul de statistiques (moyenne, écart-type, min-max,) à différentes échelles (de la région à la parcelle), etc.

- **Cloud agriculteurs**

La mise à disposition d'un espace de stockage des données pour les agriculteurs (dit « cloud agriculteurs ») est une demande qui émane des syndicats agricoles : les données sont sécurisées chez un tiers gouverné par la profession agricole, elles peuvent être ouvertes aux personnes autorisées.

- **Espace d'échanges**

À cela s'ajouterait également un espace d'échanges entre les acteurs de la profession agricole et des métiers associés pour partager expériences, expertise, savoir-faire et idées. Cet espace pourrait prendre la forme d'un forum / wiki / « réseau social » pouvant participer activement à la dynamique d'une démarche d'innovation – priorisation.

- **Vitrine d'OAD et évaluation participative**

Une réflexion doit également être engagée quant au développement d'un espace de présentation d'OAD, de services et d'évaluation participative (dit "TripAdvisor des OAD") : c'est un espace où les OAD et services existants sur le marché sont présentés et sur lesquels les utilisateurs peuvent faire des commentaires voire donner une note.

1.4 Revue des dispositifs existants en lien avec l'ouverture de données agricoles

Les démarches d'innovation ouverte à partir de données se font de plus en plus présentes dans de multiples domaines en France comme à l'étranger. Le benchmarking des initiatives déjà lancées a plusieurs avantages. Il permet de réutiliser le travail qui a été mené au sein de certains de ces projets, d'éviter les écueils identifiés, de bénéficier des retours d'expérience (y compris sur la question de la création de l'écosystème de l'innovation), voire pour certains dispositifs, d'identifier des pontages possibles. L'ensemble des dispositifs présentés ci-dessous concernent l'agriculture.

1.4.1 France : AGROSYST

Le Système d'Information AGROSYST accueille et met à disposition (des collectifs impliqués) les données issues du réseau des fermes DEPHY. Développé dans le cadre du programme national Ecophyto, il vise prioritairement à faciliter l'identification des systèmes de culture économes en pesticides les plus performants, et à favoriser leur transmission à d'autres conseillers et d'autres agriculteurs. L'objectif est de capitaliser dans un système d'information des données sur les systèmes de culture afin de faciliter leur valorisation et leur analyse transversale. Le système d'information assure les fonctions suivantes :

- L'acquisition et l'hébergement de données de natures différentes (succession des cultures, interventions pratiquées sur les cultures, règles de décision, mesures, observations, etc.),
- Le calcul de variables de synthèse, de bilans, de différents indicateurs, de rapports, etc.
- L'édition de schémas décisionnels,
- L'interopérabilité avec d'autres systèmes via une interface dédiée.

1.4.2 France : API Agro

API-AGRO (<http://www.api-agro.fr/>) est un projet lauréat de l'appel à projets CASDAR 2013 «recherche finalisée», réunissant la majorité des Instituts Techniques Agricoles (ITA) et quelques partenaires de la Recherche & Développement (INRA, l'APCA, Agro-Transfert Ressources et Territoires) au niveau national. Ce projet vise à co-construire une plateforme informatique optimisant les services rendus en termes de résultats d'évaluation des systèmes grâce à un partage de références agronomiques communes mises à disposition par les partenaires. Opérationnelle depuis juin 2016, la plateforme API-Agro est une interface de programmation (API pour Application Programming Interface) qui met à disposition des services web standards souhaitant accéder à des référentiels ou des services agronomiques (comme clients) ou les diffuser (comme fournisseurs). La finalité du projet est de construire une plateforme viable économiquement favorisant l'interopérabilité des références agronomiques des partenaires du projet. Bien que d'une ambition volontairement limitée à ce stade, API-AGRO est doté de fonctionnalités proches de celles recherchées pour le portail. L'expérience liée à sa mise en place mérite d'être valorisée dans le projet et son articulation avec le portail numérique au moins pour l'accès aux données devra être trouvée.

1.4.3 Danemark : SEGES

Le SEGES est né de la fusion du Knowledge Centre for Agriculture et du Danish Pig Research Centre, au 1er janvier 2015. C'est un centre de recherche et d'essais, fort de 650 personnes, gouverné par la profession agricole. Il effectue environ 1 000 essais par an et plusieurs projets de recherche, en collaboration avec le DLBR (the Danish Agricultural Advisory Service). Il couvre tous les aspects de l'agriculture et de la gestion agricole : de production agricole (environnement, élevage, agriculture biologique, etc) au financement, la législation fiscale, l'architecture informatique, la comptabilité, les ressources humaines, la formation.

Concernant les données, le SEGES collecte depuis une vingtaine d'années des données à vocation agricole. Il dispose aujourd'hui d'une base de données professionnelles (LandbrugsInfo) qui comprend plus de 120 000 pages détaillant les résultats d'essais, d'analyses et des informations sur la réglementation. La base de données, organisée en 700 catégories, contient plus de 280 000 références croisées et reçoit 50 000 visites par mois. Les informations sont soit disponibles directement, sans inscription, soit via un abonnement annuel de 270 euros à 670 euros. La base est utilisée à 90% comme une base de connaissance. Le SEGES utilise ces données pour mettre au point de nouveaux outils d'aide à la décision. Les données peuvent également être vendues à des tiers dès lors que leur propriétaire donne son accord, mais la base n'a pas vocation à une ouverture large facilitant l'innovation ouverte, mais reste à destination préférentielle du SEGES.

À côté de Landbrugsinfo, environ 20.000 agriculteurs utilisent le portail « landmand.dk » aussi géré par SEGES. Le portail et LandbrugsInfo sont étroitement interconnectés, et les données issues des fermiers sont utilisées pour personnaliser le contenu de LandbrugsInfo : par exemple un fermier qui cultive de l'orge recevra des alertes et des informations sur l'orge, émises à partir de LandbrugsInfo, quand il se connecte à landmand.dk.

1.4.4 Allemagne : Data Hub DKE

La société DKE ('Digitale Kommunikation und Entwicklung' [Communication et Développement Numérique] - <http://www.dke-data.com/fr/>), construite sur un consortium d'entreprises (majoritairement des constructeurs de machines agricoles) basées en Allemagne, développe actuellement une plateforme d'échange de données (Data Hub) indépendante. Ses clients bénéficieront d'une prestation de service : le transport de données sûr et indépendant de toute machine associée. Le lancement officiel de cette plateforme se fera en 2017. Selon DKE la plateforme permettra de réduire substantiellement les problèmes de conversion lors de la communication de données. L'utilisation de cette nouvelle plateforme sera gratuite pour les agriculteurs et les entreprises de travaux agricoles. Le transport des données se financera par les contributions des fabricants de machines agricoles, des développeurs d'applications et d'autres fournisseurs de services. Ainsi, cette plateforme n'est pas une plateforme de stockage ou d'ouverture de données, mais de transfert entre machines. Il n'y a donc que peu de liens avec le concept de portail de données à vocation agricole.

1.4.5 Etats-Unis : The Agricultural Data Coalition (ADC)

En 2013-2014, le Farm Bureau (l'un des principaux syndicats agricoles aux USA) a été alerté par ses adhérents du fait que les données de leurs exploitations étaient « captées » par les fournisseurs (intrants, agro-équipementiers), via des contrats peu explicites sur la politique de gestion des données. Le Farm bureau a réagi avec 3 actions successives :

- Une charte d'une douzaine de principes de gestion des données, établie en concertation avec les agro-fournisseurs³;
- Le label « AgData transparent », mis au point en concertation avec d'autres syndicats agricoles et des agrofournisseurs, pour renseigner les producteurs sur les contrats qui les lient aux fournisseurs de logiciels ou d'équipements. Ce label se base sur un outil, le « AgData transparency evaluator », qui analyse le contrat, sur une base de 10 questions concernant le type de données collectées, leur propriété, l'utilisation et partage des données, la gestion de la confidentialité, les niveaux de sécurité, etc. (voir encadré ci-dessous) ;

³ <http://www.fb.org/tmp/uploads/PrivacyAndSecurityPrinciplesForFarmData.pdf>

- 1-Quelles catégories de données le produit ou service recueille-t-il chez moi, l'agriculteur?
- 2-Est-ce que la convention avec le Fournisseur d'agroéquipement (FAE) traite de la propriété de mes données après que mes données ont été transférées?
- 3-Si le FAE contracte avec d'autres entreprises pour fournir des données pour des services connexes, est-ce que le FAE exige que ces entreprises adhèrent à la politique de confidentialité que l'ATP a mis en place avec moi?
- 4-Est-ce que le FAE doit obtenir mon consentement avant de fournir l'accès à mes données à d'autres entreprises?
- 5-Après avoir téléchargé mes données chez le FAE, me sera-t-il possible de récupérer mon jeu de données complet d'origine dans un format original ou équivalent?
- 6-Est-ce que le FAE m'informerait si ces accords changent?
- 7-Est-ce que le FAE me préviendra en cas de violation de la sécurité des données, se traduisant par la divulgation de mes données à un tiers?
- 8-Mon jeu de données d'origine peut-il être supprimé chez le FAE, à ma demande, lorsque mon contrat avec le FAE se termine?
- 9-Est-ce que la convention avec le FAE établit la durée pendant laquelle mes données d'origine seront conservées ?
- 10-Est-ce que la convention avec le FAE traite du devenir de mes données si le FAE est vendu à une autre société?

- La AgData coalition (ADC)

L'ADC (<http://agdatacoalition.org/>) a été créée pour aider les agriculteurs à mieux contrôler et gérer leurs données - de manière indépendante des sociétés privées qui assurent l'acquisition et la mise en forme - pour stocker ces données dans des espaces sécurisés et les ouvrir à la demande de l'agriculteur à des tiers autorisés. L'objectif de l'ADC est de construire une base de données agricoles neutre et indépendante dans laquelle les agriculteurs pourront stocker leurs données en toute sécurité et de manière pérenne : informations collectées chaque jour dans leurs parcelles par leurs machines et autres outils d'acquisition (imagerie aérienne par exemple). Ces données peuvent alors être nettoyées, synchronisées et transmises d'une manière efficace et uniforme à des tiers, qu'ils soient chercheurs, assureurs, responsables gouvernementaux, gestionnaires agricoles, fournisseurs d'intrants ou tout autre acteur autorisé par l'agriculteur. C'est l'agriculteur qui choisira la personne avec qui il souhaite partager et la durée du partage. Le coût d'accès à ce service devrait être faible : 250\$ par an. La plateforme choisie est "Onsite" de AgIntegrated⁴.

1.4.6 Royaume-Uni : Agrimetrics

Agrimetrics est une société à but non lucratif, financée par le gouvernement du Royaume-Uni dans le cadre de sa Stratégie Agritech. Son ambition est d'être un centre d'excellence sur le Big Data en agriculture et en agro-alimentaire : data sciences, modélisation... Il s'appuie sur quatre organisations fondatrices (Rothamsted Research, l'Université de Reading, l'Institut National de Botanique Agricole / NIAB et le Collège Rural d'Ecosse / SRUC) - et plus de 50 entreprises et organisations environnementales et académiques.

L'objectif d'Agrimetrics est de bâtir une plate-forme de données et un écosystème associé pour soutenir l'intégration, la modélisation et l'analyse des données et fournir l'accès à des données

⁴ <http://www.agintegrated.com/onsite-platform/>

provenant d'un large éventail de domaines pertinents pour le secteur. L'accent est mis sur la filière, de l'agriculture à l'alimentaire, plus que sur la production primaire. Le projet a débuté début 2016 avec une équipe d'une vingtaine de personnes (construction de la plateforme, collecte des données, traitement de la donnée). La plateforme sera partiellement opérationnelle en 2017 et devra être autosuffisante au bout de 3 ans. Un enjeu-clé est que les entreprises du secteur agricole et alimentaire comprennent l'impact que les données peuvent avoir sur leur organisation, fournissent des données et puissent accéder aux produits d'Agrimetrics. A terme, Agrimetrics proposerait aussi un coffre-fort de données.

Agrimetrics s'attachera à :

- Créer un « data-hub » (plateforme de données) pour les secteurs agricole et alimentaire : Créer un grand répertoire de données agricoles et alimentaires interconnectées (« linked data ») et en faciliter l'accès pour permettre de construire des approches holistiques des défis de la production agricole et alimentaire
- Construire des outils intégrant des données et une expertise associée pour une meilleure prise de décision fondée sur des données
- Créer un réseau pour mettre en relation les différentes entreprises du domaine
- Construire la confiance et partager la valeur : en développant une culture de partage des données et des décisions fondées sur les données, dans l'optique d'une intensification durable de la collecte.

1.4.7 Conclusion

Les différents dispositifs présentés ci-dessus concernent le secteur de l'agriculture. Des initiatives similaires existent dans d'autres secteurs d'activité, mais il nous a paru intéressant de relater en particulier l'expérience française de la mission ETALAB, car plusieurs similitudes ont été identifiées.

France : Data.gouv (mission ETALAB)

Etalab (<https://www.etalab.gouv.fr/>) coordonne l'action des services de l'Etat et des établissements publics pour faciliter la réutilisation la plus large possible des données publiques. Etalab administre le portail interministériel data.gouv.fr destiné à rassembler et à mettre à disposition librement l'ensemble des informations publiques de l'Etat, de ses établissements publics et, si elles le souhaitent, des collectivités territoriales et des personnes de droit public ou de droit privé chargées d'une mission de service public. Etalab collabore étroitement avec les services chargés de la modernisation de l'action publique, notamment ceux responsables de l'innovation au service des usagers et de la transformation numérique de l'Etat. Le portail national data.gouv.fr donne accès à plus de 13.000 informations publiques gratuites et réutilisables. Etalab dispose donc de données susceptibles d'intérêt pour le portail de données agricoles. L'animation de l'écosystème d'innovation passe aujourd'hui essentiellement par des fonctionnalités sociales proposées sur la plateforme (discussions, republication de données, transferts de ressources d'un producteur à un autre...).

L'analyse comparative de ces différentes initiatives a permis de mettre en évidence des expériences utiles sur lesquelles s'appuyer pour la mise en place du portail, mais elle a aussi démontré qu'aucune n'a la même ambition que le futur portail AgGate, en particulier dans son objectif de favoriser l'innovation ouverte. Voici un tableau qui caractérise ces projets selon trois critères :

	Objectif	Ouverture	Animation
Etalab	Réutilisation des données	Public	Non
Agrosyst	Réutilisation des données	Ouvert au Collectif	Non
API-Agro	Réutilisation des données	Ouvert au Collectif (puis plus large ?)	Non
SEGES	Stockage et Réutilisation des données	Limité au producteur, SEGES et tiers autorisés	Non
Agrimetrics	Réutilisation des données (et stockage, à terme)	Ouvert	Oui
AgDataCoalition	Stockage des données	Limité au producteur et tiers autorisés	Non
AgGate	Réutilisation des données (et stockage si option « cloud »)	Ouvert	Oui

Voici les similitudes et les différences de ces initiatives:

- Etalab :
 - Similitude : recensement de données dans le but de les rendre publiques pour les réutiliser
 - Différence : uniquement données publiques, non ciblé sur l'agriculture, vocation limitée dans l'animation de l'innovation ouverte.
- Le Seges
 - Similitude : collecte de données à vocation agricoles
 - Différence : pas d'ouverture, la valorisation est faite en interne par Seges
- Agrosys
 - Similitude : collecte de données à vocation agricole dans l'objectif de les réutiliser (transfert de savoir-faire, élaboration d'OAD)
 - Différence : pas d'ouverture en dehors de la recherche.
- API-AGRO
 - Similitude : collecte et mise à disposition de référentiels agricoles générés par des acteurs de la R et D agricole, dans le but de les réutiliser, ouverture dans un premier temps au collectif du projet, puis plus largement
 - Différence : contour plus limité, pas d'animation lourde, pas d'interconnexion de données
- Agrimetrics
 - Similitude : collecte de mise à disposition de données agricoles et alimentaires dans le but de les réutiliser ; interconnexion des données ;
 - Différence : animation de l'écosystème limité

Au final, aucune des initiatives ne combine « ouverture des données agricoles » / « interconnexion de données » / « animation de l'écosystème de l'innovation », triptyque de base pour créer un écosystème de l'innovation. L'initiative la plus proche (y compris dans les aspects techniques) est la plate-forme Agrimetrics, la seule différence étant un aspect « animation » (organisation de challenge, hackathons etc) moins développé. Un intérêt a été émis par le contact de la plateforme Agrimetrics d'étudier les pontages possibles avec la démarche AgGate.

Gouvernance et structure juridique



2. Gouvernance et structure juridique

L'enjeu principal étant de mobiliser les données des exploitations agricoles pour en assurer la valorisation à leur service, la gouvernance retenue doit privilégier la confiance mais aussi la mobilisation des acteurs du monde agricole. Dans cette notion de confiance et de mobilisation des acteurs économiques, l'accès ouvert et large aux données publiques reste un des enjeux.

Ce sujet plusieurs fois abordé au sein du Conseil de l'Agriculture Française mobilise les acteurs et a fait clairement l'objet de soutien dans le cadre des suites du rapport « innovation agriculture 2025 » par les acteurs économiques et syndicaux du monde agricole et des ministres en charge de l'agriculture et du numérique.

La structure juridique de la gouvernance la plus adaptée par sa souplesse de mise en œuvre apparaît la SAS. En charge du montage et de la gestion opérationnelle du portail, elle mobilisera quelques acteurs économiques opérationnels du monde agricole (coopératives, filières, banques, assurances, entreprises, structures d'investissement...). Correspondant à une mutualisation, elle mérite d'être reconnue comme « pôle d'innovation » au titre des critères du CGI, pour bénéficier d'un appui public allant jusqu'à 50% des coûts pour sa mise en place. Pour cela, il est souhaitable de créer un actionnariat diversifié, en cherchant à ce qu'aucune grande entreprise ne détienne plus de 25% du capital.

Pour sa gouvernance stratégique, il convient de mettre en place un comité d'orientation stratégique qui sera garant des conditions d'utilisation, de la déontologie et de l'éthique du portail numérique, mais aussi de l'animation de l'écosystème d'innovation. Il s'assurera d'un lien avec les professions concernées pour une écoute et une anticipation des besoins exprimés par les professionnels et les organismes publics. Il devra rassembler au moins, les représentants des syndicats agricoles, l'APCA, COOP de France, la CNMCCA, des représentants du secteur de l'agrofourriture (AXEMA, UFS, UIPP, UNIFA ; l'ANELA), des représentants de la recherche (INRA, Irstea, ACTA) et des représentants des ministères en charge de l'agriculture et du numérique, et des personnalités qualifiées du secteur numérique. L'équilibre de ce comité comme les principes de la création du portail pourraient utilement être validés lors d'un comité stratégique d'orientation (CSO) du ministère en charge de l'agriculture pour en conforter la légitimité.

Cette composition méritera d'être affinée voire d'évoluer au cours du temps en fonction de l'extension du portail, par exemple à l'ANIA s'il s'étend à l'industrie agro-alimentaire. Beaucoup d'autres acteurs ont manifesté un intérêt dans le portail, la FNSAFER, la MSA, les centres de gestion, la FNCUMA, la FNEDT, l'ordre des vétérinaires...

Les jeux de données à intégrer



3. Les jeux de données à intégrer

Un travail d'inventaire a été lancé pour identifier les bases de données à vocation agricole existantes en France. Ce travail a plusieurs objectifs :

- Identifier les fournisseurs disposant de données,
- Impliquer ces fournisseurs dans le projet et comprendre leurs attentes éventuelles,
- Identifier les thématiques agricoles pour lesquelles des données existent,
- Caractériser plus finement les données et leur structuration afin d'identifier les éventuelles contraintes auxquelles le guichet unique devra répondre.

3.1 La démarche

Les bases de cette démarche ont été établies en collaboration avec Etalab qui réalise ce genre d'inventaire pour les données d'origine publique. Dans cette mission, nous ne nous sommes pas limités aux données publiques et avons également recherché les fournisseurs potentiels de données privées (par exemple, éditeurs de logiciels, agro-équipementiers...). La démarche est la suivante, une fois une source potentielle identifiée :

1. Contact avec le potentiel fournisseur de données afin d'identifier le référent ressource.
2. Envoi d'un questionnaire (cf Annexe 2) afin que le référent prépare les réponses.
3. Entretien semi-directif avec le référent en prenant le questionnaire comme guide.
4. Rédaction du compte-rendu de l'entretien.
5. Renseignement du tableur qui résume l'ensemble des entretiens réalisés.

3.2 Origine des données

Dans le temps imparti, la mission s'est attachée à initier un inventaire, en référençant les fournisseurs de données et en élaborant une méthodologie pour identifier et décrire les bases de données. Ce travail est donc loin d'être exhaustif et méritera d'être approfondi.

Plusieurs fournisseurs de données ont été référencés (cf Annexe 3), mais le portail devra pouvoir s'enrichir de nouvelles sources au fil de l'eau, en fonction de l'évolution des besoins et de l'offre. Les fournisseurs sont des administrations publiques, des agro-équipementiers, des éditeurs de logiciel de traçabilité, des éditeurs d'outils d'aide à la décision, des centres de recherche, etc. Les agriculteurs pourraient également déposer directement leurs données sur le site grâce à la mise à disposition d'un cloud agricole, qui est une des déclinaisons du portail. Cette brique fonctionnelle sera abordée ultérieurement dans ce document.

Certaines données publiques sont en partie accessibles depuis les portails nationaux, notamment "data.gouv.fr" en France. Néanmoins nombreuses sont encore les données publiques qui ne sont pas disponibles sur des portails. L'étude de la mise en place du portail AgGate a permis d'initier le travail d'inventaire et de mise à disposition de ces données. Voici quelques exemples de données publiques: parcellaire, occupation du sol, données du recensement agricole, données Eurostat, données économiques sur les marchés, météorologie Météo France, topographie IGN.

D'autres données seront d'origine privée (entreprises, coopératives, syndicats...). Elles seront rendues accessibles par une participation volontaire de leurs propriétaires à la démarche. Ceux-ci

décideront du type d'accès (données visibles et/ou téléchargeables), des modalités d'accès, des licences associées à la réutilisation de leurs données, et des coûts d'accès aux jeux de données et/ou aux services développés dans le respect du cadre légal (cf Annexe 6).

Liste des données disponibles

Dans le cadre de l'inventaire, une vingtaine d'entretiens en direct ont été réalisés (présentiel ou par téléphone), majoritairement avec des acteurs privés. Ces entretiens ont permis d'identifier et de « caractériser » une trentaine de bases de données, c'est-à-dire d'en appréhender le contenu (type de données, moyen de collecte, étendue spatiale, etc). Le questionnaire a été transmis au sein des ministères. À ce jour les retours ont permis de caractériser également une trentaine de bases de données. L'inventaire devra régulièrement être remis à jour et amendé avec les nouvelles bases de données identifiées⁵.

Plusieurs acteurs ont par ailleurs initié le recensement de bases de données agricoles, dans différents secteurs. A défaut d'être exhaustifs, citons 2 exemples :

- Le Groupement d'Intérêt Scientifique GC HP2E, dans le cadre des missions du groupe thématique « connaissance des pratiques pour l'évaluation des performances » co-animé par Arvalis et l'INRA, s'est engagé dans un travail de recensement des bases de données existantes utilisées pour l'évaluation des performances en grandes cultures. Cette démarche a pour but d'étudier les opportunités, les besoins et usages, ainsi que la faisabilité d'un système d'information sur les pratiques des producteurs de grandes cultures permettant l'évaluation de leurs performances. Près de 130 bases de données « réseaux, enquêtes et observatoires » ont été identifiées et environ 80 ont été caractérisées. Il est à noter que le questionnaire utilisé est quasiment identique à celui utilisé pour le portail agricole.
- Le Groupement d'Intérêt Scientifique sur l'Elevage a également réalisé un inventaire des bases de données se rapportant à cette thématique en 2012. Une vingtaine de bases a été identifiée et caractérisée.

Le travail d'identification / caractérisation réalisé dans le cadre de ces 2 inventaires pourra être valorisé dans le portail agricole.

A ce stade, nous avons d'ores et déjà caractérisé près de 110 bases de données (dont certaines très importantes dans leur étendue – la quantité « d'objets » reportés – ou leur profondeur). La liste des bases identifiées ou caractérisées est présentée en Annexe 3. La caractérisation n'est pas toujours complète : il reste souvent quelques caractéristiques à préciser. Par exemple, l'étendue temporelle des données n'est pas toujours bien connue du fournisseur, de même que la projection géographique utilisée ou la volumétrie de la base de données.

D'autre part, certaines ressources sont déjà accessibles sur des portails Web et pourront être référencées au sein du portail agricole. C'est notamment le cas de certaines ressources présentes sur

⁵ Il serait intéressant de prévoir une mise à disposition en open data de cet inventaire, côté données publiques, dans l'idée de contribuer à la réalisation d'une cartographie collaborative des bases de données de l'Etat.

le portail API-Agro (<http://www.api-agro.fr/>), sur www.data.gouv.fr ou sur www.open-data.europa.eu. Le site AgriOntology (<https://sites.google.com/site/agriontology/home>) recense également de nombreuses ressources en agriculture, environnement et géographie accessibles en France et à l'étranger.

3.3 Caractérisation des données et métadonnées

Les jeux de données disponibles via le guichet unique sont de nature hétérogène de par :

- Leur qualité, qui peut varier selon la précision de la mesure, le capteur utilisé ou l'observateur.
- Leur échelle spatiale (pour les données géolocalisées) : de l'intra-parcellaire (type et nature des sols, hydrométrie ou biomasse, cartographies d'apports d'intrants de densités de peuplement, de rendements et de qualité des récoltes, etc.) aux données globales (météo, cartographie géologique, carte topologique, etc).
- Leur temporalité : date de prise de l'information, fréquence des mesures.
- Leur nature : données cartographiques, photos, informations génétiques, référencement de produits phytosanitaires et fertilisants, etc.

Ces différentes caractéristiques ne sont qu'une partie des métadonnées que le guichet unique devra prendre en compte et utiliser pour mettre les données à disposition de manière efficace. Les métadonnées sont indispensables à l'exploitation des données. La description des métadonnées envisagées sera abordée dans la partie "6. Points techniques critiques".

Certaines données ont un caractère personnel ou relèvent du secret des affaires. Leur diffusion ne peut être faite qu'avec l'autorisation de la personne concernée (consentement éclairé) ou après anonymisation. Il est donc également important de maîtriser les notions d'anonymisation, aussi bien d'un point de vue juridique (qu'est-ce que la notion d'anonymat) que d'un point de vue technique. Ces points cruciaux seront également abordés dans la partie "6. Points techniques critiques" et dans l'annexe 6 « étude juridique ».

3.4 Accessibilité des données

Les données qui seront disponibles sur le guichet seront soit conservées chez le fournisseur, soit stockées sur le cloud « agriculteurs ». Le guichet unique devra proposer des solutions aux fournisseurs pour leur permettre de rendre leurs données accessibles en minimisant le coût de cette mise à disposition. Pour cela, des aménagements techniques seront nécessaires ; si le fournisseur dispose d'API, le portail les référencera, dans le cas contraire, il pourra aider à la création d'API (« magasin numérique »). Une collaboration entre les administrateurs du portail et les fournisseurs sera nécessaire pour mettre en place les solutions d'accès aux données, décrites dans la partie "6. Points techniques critiques".

3.5 Les aspects juridiques de la donnée

Un portail web est un site Internet dédié à un domaine d'intérêt particulier ou construit pour une communauté spécifique. L'objet d'un portail est d'ouvrir une « porte d'entrée » sur un large choix de

ressources et de services accessibles par Internet. Plus précisément, ce portail a pour objet de mettre à disposition du public des données et services numériques pour une agriculture française compétitive, ouverte et souveraine.

Créer un portail d'une telle nature suppose, d'une part, de respecter un certain formalisme juridique, dont la mise en ligne de mentions légales sur le site et, d'autre part, de protéger juridiquement les données accessibles sur le site, notamment par la mise en place d'un cadre contractuel permettant de définir les modalités de la mise à disposition et de la réutilisation des données publiques ou privées qui alimenteront le portail, pour chaque utilisateur privé ou public.

Les problématiques juridiques figurent dans le schéma ci-après (Figure 2). Une étude détaillée figure en Annexe 6.

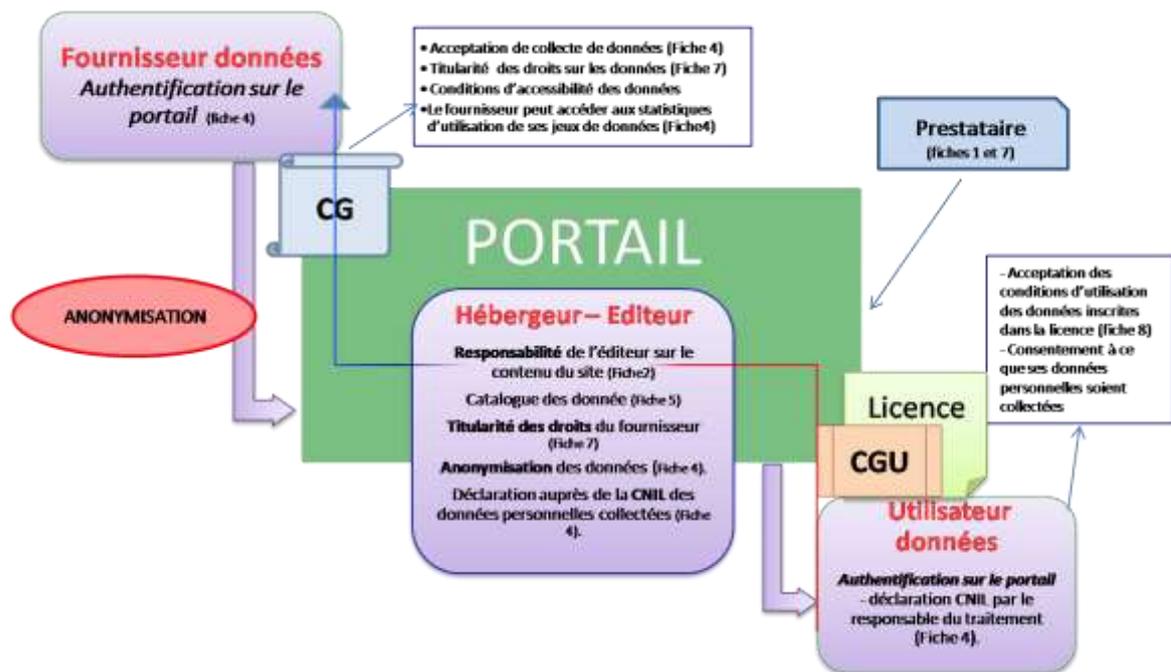


Figure 2 – Problématiques juridiques relatives au portail AgGate

De manière synthétique, il faut retenir les points suivants :

Lors de la collecte de la donnée :

- Il est indispensable d'identifier le titulaire des droits sur la donnée et d'avoir son accord pour rendre les données visibles ;
Lorsqu'il s'agit d'une donnée à caractère personnel, il est obligatoire d'obtenir l'accord des personnes pouvant être identifiées par ces données préalablement à leur collecte. La collecte fera l'objet d'une déclaration auprès de la CNIL. Leur accord doit notamment porter sur la finalité de la collecte. A ce titre, il sera indiqué à ces personnes pour quelle finalité leurs données à caractère personnel sont collectées, c'est-à-dire à des fins statistiques, de recherche.....
Ces données à caractère personnel ne seront en aucun cas rediffusées sans leur accord. En tout état de cause, elles seront anonymisées par l'administrateur du portail avant toute rediffusion.

Sur ce sujet des données à caractère personnel, il convient de souligner, au moment où est rédigé le présent rapport, que la CNIL est en train d'élaborer un pack de conformité sur « l'open data ». Elaborés en concertation avec les acteurs d'un secteur d'activité, les packs de conformité – il y en a trois à ce jour – sont, selon les propres termes de la CNIL, un nouveau mode de régulation. Ils visent à définir et diffuser les bonnes pratiques pour un secteur, tout en simplifiant les formalités administratives des acteurs qui s'y conforment. Ils peuvent ainsi contenir des mesures de simplification des formalités, des guides pratiques et pédagogiques, des tests de vérification de conformité à la loi.

Dans la mesure où le pack de conformité « open data » aura été effectivement publié avant la mise en œuvre du portail, il sera utile de s'y référer.

Gouvernance des données

Une attention particulière doit être portée à la gouvernance des données ; pour cela, deux dispositifs sont prévus: les Conditions Générales (CG), qui s'adressent aux fournisseurs de données et édictent les règles liées au dépôt, et les Conditions Générales d'Utilisation (CGU), qui s'adressent aux utilisateurs de données.

Pour le fournisseur, les CG devront notamment préciser :

- la nature des données mises à disposition ;
- Ses droits et obligations (notamment un engagement de titularité du jeu de donnée) ;
- Les droits de propriété intellectuelle afférant aux données ;
- Les droits qu'il confère à l'utilisateur en fonction des différents types de licence proposée (accès, utilisation, réutilisation...)
- Le degré de responsabilité de l'éditeur ;
- Le type d'anonymisation qui sera appliqué (si nécessaire)
- Les obligations relatives à la « loi Informatiques et libertés » (sur les données à caractère personnel)
- Le droit applicable.

Un fournisseur de données souhaitant rendre ses données accessibles devra préalablement déterminer les droits qu'il souhaite accorder sur ses données. Plusieurs types de licence seront proposés, par exemple :

- les licences fixées par décret conformément à la loi « pour une république numérique ».
- les licences spécifiques demandées par les fournisseurs eux-mêmes lorsqu'ils détiennent des droits d'auteur sur leurs données,
- les licences commerciales précisant les modalités financières.

Pour l'utilisateur, les CGU devront notamment préciser :

- la nature des données mises à disposition ;
- Les droits de propriété intellectuelle afférant aux données ;
- Les modalités d'accès aux données ;
- Les droits conférés à l'utilisateur en fonction des différents types de licence proposée (accès, utilisation, réutilisation...) et ses obligations, notamment l'obligation de citer les sources lorsqu'une donnée est mise en ligne sur le site ;
- Les droits et obligations des fournisseurs de données (notamment un engagement de titularité du jeu de donnée) ;
- Le degré de responsabilité de l'éditeur ;
- Les obligations relatives à la « loi Informatique et libertés » (sur les données à caractère personnel)
- Le droit applicable.

Au cœur du portail, le guichet unique



4. Au cœur du portail, le guichet unique

Conformément à la description des grandes fonctionnalités (1.3.3) la première brique fonctionnelle du portail sera le guichet unique, qui permettra de recenser et mettre à disposition les données publiques et privées stockées sur les plateformes de leurs propriétaires ou sur le « cloud agriculteurs » (une des fonctionnalités du portail décrite en 5. « Les autres fonctionnalités indispensables au portail »). Nous décrirons ci-dessous les utilisations et utilisateurs du portail/guichet, puis les fonctionnalités à mettre en œuvre pour le guichet.

4.1 Utilisateurs et usages du guichet

4.1.1 Les utilisateurs du portail

Trois types d'utilisateurs-acteurs du portail ont été identifiés :

- Le consommateur de données : il vient sur le guichet unique pour rechercher, visualiser, utiliser et télécharger des données.
- Le fournisseur de données : il propose l'accès à des données dont il est titulaire.
- L'administrateur : il modère l'utilisation du guichet, il anime les outils de communication auprès de la communauté d'utilisateurs (forum, newsletter,...), suit les statistiques, gère les droits, etc.

À savoir, ces rôles ne sont pas exclusifs : un fournisseur de données peut également être consommateur d'autres données et vice-versa.

Les utilisateurs potentiels du portail sont des coopératives, des entreprises de data science et de conseil, des institutions publiques, la recherche, des agriculteurs, des associations, des citoyens, etc. Chacun de ces acteurs pourra être intéressé par des données différentes et exprimera ses propres exigences quant à l'utilisation qu'il fera du portail, aux données qu'il mobilisera, et aux services qu'il souhaiterait voir développer. Les *data scientists*, chercheurs et informaticiens utiliseront davantage le guichet pour y trouver des données brutes et des API. Les autres utilisateurs (agriculteur, institutions) se dirigeront plus naturellement vers d'autres fonctionnalités qui seront décrites plus bas. La Figure 3 décrit les fonctionnalités spécifiques à chaque type d'utilisateur/acteur du portail.

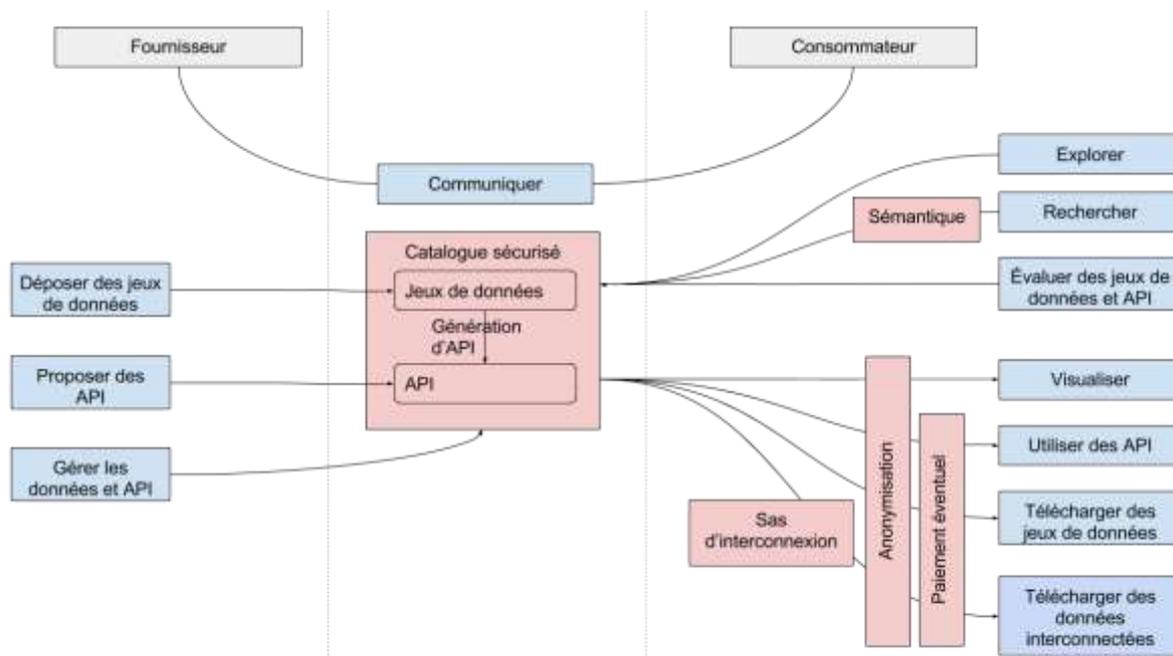


Figure 3 - Schéma simplifié des fonctionnalités disponibles sur le guichet pour le consommateur et le fournisseur

4.1.2 Vocation des données

Les données peuvent avoir deux vocations, en fonction de l'utilisation qu'en fait le consommateur :

- **"Product building"** : le consommateur va chercher des données diverses (par exemple itinéraires techniques, météo, données de sol, modèles numériques de terrain, etc.) qu'il pourra traiter dans une démarche de R&D, de "data analysis", dans le but de construire ou d'ajuster des modèles ou des applications. L'utilisation des données peut se faire soit via une API, soit en téléchargeant directement les jeux de données en local.
- **"Product running"** : le consommateur accède et télécharge les données dans le but de faire tourner des modèles déjà construits qui sont intégrés dans des produits en phase de commercialisation. Un exemple pourrait être l'accès à des données météo en temps réel dans un outil d'aide à la décision pour l'irrigation. Pour garantir des données toujours à jour et un accès en temps réel, l'utilisation des données ne peut se faire que via une API, en direct à partir du fournisseur.

Ces deux modes d'utilisation impliquent des fonctionnalités, une architecture et des contraintes techniques différentes qui seront abordées ultérieurement.

4.1.3 Scénarios d'utilisation

Ces scénarisations présentent deux utilisations possibles du portail pour opérer une requête via le moteur de recherche et télécharger des jeux de données ou déposer des jeux de données. Ils permettent d'illustrer les fonctions du portail mais ne présagent pas les solutions techniques qui pourront être réellement développées. Ces deux scénarios sont des exemples de l'utilisation du portail et ne sont pas exhaustifs des usages ou des modalités d'accès au portail.

Consommateur de données

Je suis le dirigeant d'une start-up de conseil. J'ai créé un prototype de modèle pour diagnostiquer le risque de mildiou en région bordelaise à partir de la météo, du type de sol, de l'itinéraire technique et des antécédents parasitaires de la parcelle sur 5 ans. Je veux extrapoler ce modèle à d'autres zones viticoles. Pour cela j'ai besoin de ces mêmes données dans d'autres régions viticoles.

Je cherche dans le portail de données les jeux combinant la météo, type de sol, itinéraire technique et antécédents parasitaires de la parcelle sur 5 ans

Je choisis de télécharger les jeux de données sous forme brute (avec coordonnées géographiques) ou interconnectées (jointes) par le portail et j'accepte de me conformer aux obligations inscrites dans le contrat de licence associé à ce jeu de données par « un double clic ».

Je télécharge les données.

Le portail de données répond "150 parcelles répondent à vos critères" :

- Région LR: 100 parcelles
- Région Bourgogne: 30 parcelles
- Région Alsace: 20 parcelles
- météo (résolution spatiale et temporelle)
- type de sol (résolution spatiale)

et donne les conditions d'accès pour chaque jeu de données (licence, coût, fournisseur ...)

Le portail accède aux données, et les interconnecte/ agrège / anonymise en fonction de l'option choisie et ouvre le téléchargement.

Fournisseur de jeux de données

Je suis un éditeur de logiciel de conseils. Je suis titulaire de données stockées sous format MS-Excel et je souhaite les rendre accessibles sur le portail.

Je crée un compte « fournisseur ». J'accepte les CG.

Je vais dans mon espace "fournisseur". Je dépose un premier jeu de données sous forme de fichier MS-Excel

Je renseigne les métadonnées de mon jeu de données (nom, description, projection, fréquence d'acquisition, licence, etc) et valide la demande d'ajout.

J'ai maintenant accès à l'espace de gestion de mon jeu de données (mise à jour, gestion des consommateurs, statistiques).

Le portail notifie de la bonne réception du fichier.

Le portail vérifie la complétude des métadonnées obligatoires et notifie de l'ajout du jeu de données au portail.

Ces 2 scénarios sont des exemples de l'utilisation du portail et ne sont pas exhaustifs des usages ou des modalités d'accès au portail.

4.2 Fonctionnalités communes à tous les utilisateurs

4.2.1 S'inscrire, créer un compte

Le portail sera ouvert à tous. Par défaut tout utilisateur aura accès en visualisation aux données disposant d'une licence de type "licence ouverte" sans authentification. La création d'un compte est nécessaire pour accéder aux autres données, télécharger ou utiliser des données, fournir des données.

La création d'un compte demande de renseigner un nom d'utilisateur, une organisation, un type d'organisation (une liste sera proposée), une adresse mail et un mot de passe. L'utilisateur doit s'engager à respecter les conditions générales d'utilisation. Il reçoit un mail d'activation pour valider

sa demande de création de compte. Le nom d'utilisateur et le mot de passe sont enregistrés pour faciliter les utilisations ultérieures.

Un accès restreint pourrait être éventuellement mis en place en conditionnant l'inscription de l'utilisateur aux respects de certains critères (objectif de l'inscription, utilisation des données etc...).

4.2.2 S'identifier

Un système de connexion permettra à chaque utilisateur inscrit (consommateur comme fournisseur) d'accéder au portail et à son compte personnel. L'utilisateur pourra ainsi modifier ses informations personnelles (organisation, mot de passe, nom, prénom) et gérer ses inscriptions aux newsletters du portail.

4.2.3 Mettre des commentaires sur le guichet

Le guichet comportera un espace de commentaires afin que les utilisateurs puissent donner leurs avis concernant l'utilisation du guichet et leurs besoins. Cela permettra de faire évoluer le guichet unique pour qu'il réponde au mieux à leurs attentes.

4.2.4 Choisir la langue

Le guichet devra à terme proposer une interface distincte en plusieurs langues : français, anglais, allemand, espagnol. Cela suppose une gestion des dictionnaires, des concepts reposant sur des outils sémantiques standards et ouverts (compatibilité vocabulaire agronomique-technique, GEMET...).

4.3 Fonctionnalités pour le consommateur de données

4.3.1 Explorer les jeux de données

Les jeux de données présents sur le guichet unique sont référencés et indexés en utilisant des métadonnées. Une mise en page permettra de ne pas surcharger l'écran et de garder un ensemble cohérent qui ne dépendra pas du nombre de jeux de données accessibles sur le guichet.

Le consommateur a la possibilité de filtrer le catalogue de jeux de données en utilisant les principales métadonnées (la liste des métadonnées est abordée dans la partie "6. Points techniques critiques"), par exemple :

- thèmes (génétique, assolement, météo, élevage...)
- sources et accès (publiques, privées, gratuites, payantes, licence, fournisseur)
- nature (géolocalisées, chiffrées, indicateurs, textes...)

La liste des standards, ontologies, schémas et référentiels relatifs est donnée en Annexe 5.

Des fonctionnalités de tri seront proposées au consommateur pour lui permettre de trier les résultats en fonction de ses attentes, par exemple :

- la notation attribuée par l'ensemble des consommateurs
- l'utilisation du jeu de données (nombre de téléchargement, nombre d'applications connectées).

4.3.2 Rechercher des jeux de données

En complément des filtres, l'utilisateur a la possibilité de rechercher des jeux de données suivant de multiples critères via le moteur de recherche :

- recherche par mot-clé (Ex : « céréales »)
- recherche par mots-clés et opérateurs booléens (ET, OU, mot à exclure, expression exacte), les mots-clés les plus fréquents pouvant être suggérés
- recherche géographique (par un encart cartographique ou les coordonnées d'une étendue)
- recherche temporelle
- enregistrement des recherches précédentes

Deux types de résultats seront proposés :

- Des jeux de données indépendants
- Des jeux de données interconnectées (par exemple, grâce à leurs composantes spatio-temporelles ou grâce à des numéros d'identification, pour les animaux)

Voici un exemple présentant les résultats de la recherche "parcelles + météo + sol" en précisant l'étendue géographique de l'Hérault et une fenêtre de temps correspondant à l'année 2013 :

- Jeux indépendants : "3 jeux de données trouvés"
 - Jeu n°1 : météo de la France entière sur l'année 2013
 - Jeu n°2 : données de sol dans l'Hérault
 - Jeu n°3 : parcelles et cultures dans l'Hérault en 2013
- Données interconnectées : "150 parcelles correspondent"
 - Un jeu de données interconnectées comportant des informations sur 150 parcelles de l'Hérault disposant de données météo et sol en 2013.

L'interconnexion apporte un service à haute valeur ajoutée à l'utilisateur. Les jeux de données qui en sont issus pourraient être proposés en service payant. L'interconnexion sera abordée plus techniquement dans la partie "6. Points techniques critiques".

4.3.3 Visualiser des données

Dans un souci d'accessibilité, le guichet unique proposera des outils de visualisation (tableau, graphiques, cartes, statistiques) avec a minima un extrait du jeu de données. Le consommateur aura accès aux principales métadonnées du jeu de données, y compris une description détaillée.

4.3.4 Consommer des données

Le guichet devra permettre à tout utilisateur d'accéder à des données selon les conditions fixées par leurs fournisseurs. Le consommateur pourra télécharger des données sous un format exploitable (xls, csv, json, rtf) afin de les valoriser ultérieurement (dans un cadre de "product building" par exemple). Il pourra également consommer les données via des API s'il souhaite utiliser les données directement dans un autre système informatique (dans un cadre de "product running").

4.3.5 Payer en ligne

L'ambition du portail est qu'un maximum de données soient mises à disposition librement et gratuitement. Cependant les fournisseurs souhaitant monnayer leurs données (ou une certaine quantité de données sur le mode du *freemium*) auront accès à un module permettant de gérer les paiements en ligne.

Le guichet devra donc disposer des garanties en matière de sécurisation des paiements et de modularité pour permettre la mise en place de différents modèles de paiement. Il devra gérer tous les échanges monétaires de façon transparente pour l'utilisateur.

4.3.6 Echanger avec le fournisseur de données

Le consommateur pourra échanger avec le fournisseur d'un jeu de données, par exemple pour demander des détails quant à l'origine du jeu, son utilisation, etc. Cela contribuera à améliorer la qualité de la documentation et des métadonnées des jeux présents sur le guichet.

4.3.7 Evaluer des jeux de données

Le consommateur pourra évaluer rapidement le jeu de données grâce à un formulaire en ligne comportant un espace de commentaires. Il pourra ainsi apporter son avis concernant par exemple la complétude ou la qualité du jeu de données.

4.4 Fonctionnalités pour le fournisseur de données

Le guichet unique devra permettre à tout utilisateur inscrit de mettre à disposition un jeu de données et de le gérer via une interface "fournisseur".

4.4.1 Ajout d'un jeu de données

Deux choix sont proposés au fournisseur pour mettre ses données à disposition sur le guichet :

- S'il dispose d'une API il peut la proposer sur le guichet. Concrètement il renseignera l'URL d'accès à ses données. Dans ce cas il aura accès à un espace de test de son API afin de vérifier qu'elle est effectivement accessible.
- S'il ne dispose pas d'API il peut déposer le jeu de données sur le guichet dans un format exploitable (json, csv, xls) ou ouvrir son schéma de données à l'administrateur du portail. Un service ou un module permettant de monter une API à partir de ce jeu de données sera proposé.

Dans les deux cas le fournisseur devra remplir les métadonnées caractérisant les données qu'il met à disposition. Il devra en particulier choisir la licence qui régira l'utilisation des données, parmi celles proposées par le portail. Les métadonnées renseignées, il pourra soumettre sa demande d'ajout. L'administrateur validera et le jeu de données et/ou l'API sera alors visible sur le catalogue.

4.4.2 Gestion des jeux de données

Le fournisseur pourra accéder facilement et de façon intuitive aux différents outils de gestion de ses jeux de données et API :

- Gestion des licences : le fournisseur pourra définir la licence à attribuer à chaque jeu de données ou aux données issues de son API parmi un choix de licences proposé sur le portail. Un descriptif détaillé de chacune des licences sera à disposition.
- Gestion des business model : le fournisseur pourra sélectionner pour chaque jeu de données ou API (niveau de détail de la requête) un business model existant parmi une liste prédéfinie ou concevoir ses propres business models à partir de paramètres disponibles (par exemple : nombre d'appels, fréquence d'appels, requêtes possible, etc.).
- Gestion des cycles de vie : le fournisseur pourra versionner ses jeux ou API. Un engagement sur la durée minimale de mise à disposition des API est indispensable (voir charte du fournisseur).
- Gestion de la documentation : le fournisseur pourra accéder à une interface de rédaction simple pour rédiger l'intégralité de la documentation et/ou décrire des exemples d'utilisation. Il pourra aussi télécharger des documents.
- Gestion de la facturation : le fournisseur pourra suivre les différentes factures par client et avoir un aperçu des rentrées d'argent par jeu et API. En fonction de la solution de monétisation retenue, cette interface pourra également être le lieu de déclenchement des différentes factures.
- Gestion des clients : La plateforme disposera d'un module de distribution de « tokens » personnels permettant de relier le consommateur à son identité. Chaque appel sera ainsi tracé et le fournisseur pourra consulter des statistiques (liste non exhaustive : nombre de consommateurs de ses données, nombre de requêtes sur tel ou tel jeu de données, nombre d'appels moyen/consommateur, nombre d'appels/requête, répartition horaire des appels, etc.). La plateforme permettra au fournisseur de contacter l'ensemble des consommateurs

de ses données par email pour les informer d'éventuelles évolutions à venir, avaries, ou pour tout autre échange (sondage, enquête d'utilisation, etc...).

Le consommateur de données accepte, à chaque utilisation d'un jeu de données, que ses données personnelles soient transmises au fournisseur de ce jeu de données via les CGU. Le fournisseur de données s'est engagé de son côté, par le biais des CG, à respecter la loi dite « Informatique et Libertés ».

- Gestion technique: La plateforme disposera d'un système de « ping » pour vérifier à intervalle régulier (à définir) la pleine disponibilité des API du fournisseur. En cas de défaillance de ces vérifications, en fonction de seuils à définir, que le fournisseur d'API pourra personnaliser, la plateforme enverra de façon automatique un email au fournisseur. Le fournisseur d'API pourra alors consulter les statistiques concernant les tests de disponibilité et de réactivité (temps de réponse) de ses API effectués par la plateforme.
- Gestion des droits: le fournisseur peut donner des droits de gestion à un ou plusieurs autres utilisateurs à l'échelle du jeu et de l'API (exemples de droit: éditer la documentation, accéder aux statistiques, etc.).

4.5 Fonctionnalités d'administration du portail et gestion des droits

L'espace administratif, accessible par identification, permet la gestion du back-office du portail, notamment le paramétrage des fonctions du portail et des différentes conditions d'accès aux données.

Les règles et les conditions d'accès aux données seront variables suivant les catégories d'utilisateurs. Voici les grandes catégories envisagées, et qui devront être affinées :

- Interne (administrateur du site)
- Acteurs publics (organismes de formation, de recherche, ministères)
- Entreprises et organismes privés
- Agriculteurs et professions agricoles
- Grand public

4.5.1 Monitoring et contrôle

L'administrateur validera l'inscription des jeux de données au catalogue. De la même manière, conformément aux dispositions des CGU, l'administrateur pourra bloquer les jeux de données des comptes utilisateurs ne respectant pas les Conditions Générales d'Utilisation (CGU) de la plateforme.

L'administrateur disposera d'un ensemble de statistiques pour visualiser l'utilisation des serveurs, et vérifier les engagements de qualité de service des bases fournissant les données. Dans le cas de jeux de données ou de fonctionnalités (services, API etc.) qui ne fonctionneraient plus ou n'auraient pas été mis à jour pendant une certaine durée, ceux-ci pourraient voir leur visibilité réduite en les transférant dans un module d'archivage ; cela permettrait de les laisser accessibles aux utilisateurs les ayant déjà utilisées et faciliterait une éventuelle réintégration.

4.5.2 Animation

L'administrateur disposera d'outils d'animation et de mise en page de certaines pages de la plateforme (modification bannières/news/etc. ...). Il modérera également les espaces d'échanges du portail le cas échéant. A ce titre, les modalités de contrôle des avis mis en ligne seront précisées dans les mentions légales ou les CGU (cf. Annexe 6). Il pourra enfin contacter un ou plusieurs utilisateurs (groupes) par email (newsletter par exemple).

4.5.3 Support utilisateur

L'administrateur devra avoir un accès aux contenus du portail et être en mesure d'intervenir sur la chaîne de traitement (création, modification, suppression de ressources pour le compte d'un fournisseur d'API).

Les autres fonctionnalités indispensables au portail



5. Les autres fonctionnalités indispensables au portail

Les fonctionnalités et les exigences techniques décrites ci-dessus correspondent au cœur du portail, à savoir le guichet unique. Cependant, d'autres fonctionnalités ont été identifiées, soit pour atteindre un niveau de trafic satisfaisant, soit pour faire évoluer le portail vers plus de services. La priorisation de ces projets sera faite par la gouvernance. Rappelons ces 4 fonctionnalités :

- Cloud agriculteurs
- Le magasin numérique
- L'espace échanges
- La vitrine de présentation et d'évaluation des OAD

5.1 « Cloud agriculteurs » : espace d'hébergement des données privées des agriculteurs

Les syndicats agricoles ont identifié le fait qu'il pourrait être utile de proposer aux agriculteurs un espace de stockage de données. En effet à l'heure actuelle l'agriculteur qui veut stocker les données relatives à son exploitation a deux possibilités :

- Stocker ses données chez le prestataire qui lui propose des services (éditeurs de logiciels de traçabilité, éditeurs d'OAD, agro-équipementiers, etc).
- Stocker ses données sur une solution d'hébergement individuelle (ordinateur, serveur externe, etc.), souvent pas ou peu adaptée à l'utilisation et la valorisation des données.

Le projet serait de proposer un cloud agricole afin que l'agriculteur puisse y déposer ses données et les valoriser en utilisant notamment les fonctionnalités présentes dans le guichet unique. Il faut rappeler que ce service est identique à celui qui sera proposé par l'AgDataCoalition à ses membres (cotisation annuelle prévue de 250 \$, voir 1.4.5). Le détail de la structuration et des fonctionnalités de ce cloud seront à définir ; elle pourra se baser sur celles proposées par l'AgDataCoalition aux USA, ou par le Seges (landmark.dk) et sur le dispositif AGROSYST (voir 1.4).

Sur ce sujet, il est à noter que, dans le cadre du plan « Cloud computing » (Nouvelle France Industrielle), l'Agence nationale pour la sécurité des systèmes d'information (ANSSI) est en train d'élaborer un label de sécurité visant à qualifier les opérateurs de Cloud. Ce label, dénommé « Secure cloud » comporte 2 niveaux de qualification : « Secure cloud » et « Secure cloud + ». Les prestataires de cloud certifiés respectent les recommandations émises par l'ANSSI non seulement en matière d'exigences techniques mais aussi en termes de localisation et d'hébergement. Le label « Secure cloud » impose des exigences en matière de : contrôles d'accès physiques, authentification forte avec mots de passe hachés et salés, chiffrement logiciel et hébergement des données en Europe. Le label « Secure cloud + » ira plus loin, imposant une authentification multi-facteurs, un chiffrement matériel (via HSM, Hardware Security Module) ou encore un hébergement en France ou dans l'Union Européenne.

Au regard des enjeux liés à la protection des données privées des agriculteurs, poser comme recommandation ou exigence que le prestataire choisi pour le cloud agricole soit labélisé « Secure cloud » voire « Secure cloud + » apparaît comme éminemment souhaitable.

Enfin, il est à noter que le « cloud agriculteurs » est un service offert par le portail pour garantir l'accès de l'agriculteur à sa donnée : les agriculteurs faisant appel à des gestionnaires de données pourront les dupliquer sur ce site, ceux n'ayant pas de système de gestion de leurs données en trouveront un via le cloud. Mais le cloud n'est pas le « passage obligé » par lequel les agriculteurs mettront leurs données à disposition ; en effet, les agriculteurs ayant stocké leurs données chez un gestionnaire de données pourront lui demander de les ouvrir au portail dans un approche de type « bases de données distribuées » (voir 6.1 – intégration de données).

5.2 Le « magasin numérique »

L'objectif de ce « magasin numérique » est d'attirer des utilisateurs sur le portail pour créer du trafic. Les utilisateurs à attirer en priorité sont les fournisseurs de données car les consommateurs de données devraient être prioritairement intéressés par le guichet unique.

Voici quelques idées de services qui permettraient d'attirer les fournisseurs (non exhaustif) :

- Mise en forme des données « fournisseur » sous forme de graphiques ou de cartes
- Widget graphique et cartographique à intégrer dans un site externe par une simple URL
- Convertisseurs de coordonnées géographiques
- Module d'aide à la création d'API
- Calculateurs simples, par exemple de doses à épandre (pesticides, fertilisants...)
- Accès à des API de calcul proposées par des prestataires extérieurs (Instituts techniques agricoles par exemple)
- Statistiques locales (sur la base des données du portail)
- Statistiques par secteur (par structure d'exploitation)
- Météo agricole
- Avertissements agricoles

La conception et le développement de ces fonctionnalités pourraient faire l'objet d'un hackathon. Dans un second temps, cet espace pourrait évoluer pour attirer davantage d'agriculteurs. Il pourrait donc aussi contenir des éléments de « teasing » envers les agriculteurs (voir chapitre « animation »), comme par exemple les statistiques de fourniture de données, par région, ou par secteur, ou bien encore des concours, des jeux, etc..

5.3 Espace vitrine des outils et services à l'agriculture et d'évaluation participative

Le portail pourra être un lieu de présentation (vitrine) et d'évaluation participative des outils et services de l'agriculture numérique.

Présentation : offrir aux entreprises l'opportunité de présenter et décrire leurs produits (objets connectés, capteurs, OAD...) dans une interface de type « vitrine » organisée par « grands thèmes » (fertilisation, irrigation, protection des cultures, cultures sous abris, monitoring des cultures alimentation animale, monitoring animal...).

Evaluation : il est remonté lors des entretiens que les OAD et services proposés par des tiers peuvent susciter la méfiance de la profession agricole. Les agriculteurs manquent de retours quant à l'efficacité de ces outils, leurs limites, leurs coûts, leur ergonomie, etc. Ils pourraient même être sensibles à une « certification de qualité ». Il est proposé à la communauté des utilisateurs de ces outils de faire office d'évaluateur, à l'image des évaluations du secteur hôtellerie-restauration via le site "TripAdvisor". Les agriculteurs pourraient commenter et noter les outils et services numériques qu'ils utilisent, indiquer ce qui a marché et ce qui n'a pas marché. Les éditeurs y trouveraient un avantage en identifiant les besoins et améliorations à apporter à leurs produits. Le tout serait modéré par le portail et donc impartialement.

5.4 Espace d'échanges entre acteurs de la profession agricole

Le portail peut servir de lieu d'échanges entre acteurs de la profession pour accélérer la diffusion du savoir, des pratiques et des idées en agriculture. Les différents acteurs de la profession agricole, au premier rang desquels les agriculteurs, pourraient partager leurs expériences, expertises, savoir-faire, idées, interrogations et besoins sur ce portail. Les besoins qui émergeraient ainsi pourraient nourrir l'animation de l'écosystème d'innovation. Cette fonctionnalité est l'ambition du projet Hub-Agri (<http://www.openagrifood-orleans.org/fr/les-projets/367-projet-hub-agri.html>)... Et de nombreux autres portails. Un espace de type forum/ wiki/ réseau social est envisageable, avec classement des interventions par grands thèmes.

5.5 Conclusion

En accompagnement du guichet unique, les services « Cloud » (offre de stockage) et « Magasin numérique » semblent tout-à-fait intéressants pour générer du trafic. Le service « Cloud » peut également générer des recettes pour conforter le modèle économique. La mise en place de cet espace d'hébergement se traduit par des contraintes en termes de responsabilités et de rôles comme décrites dans les figures 4 et 5.

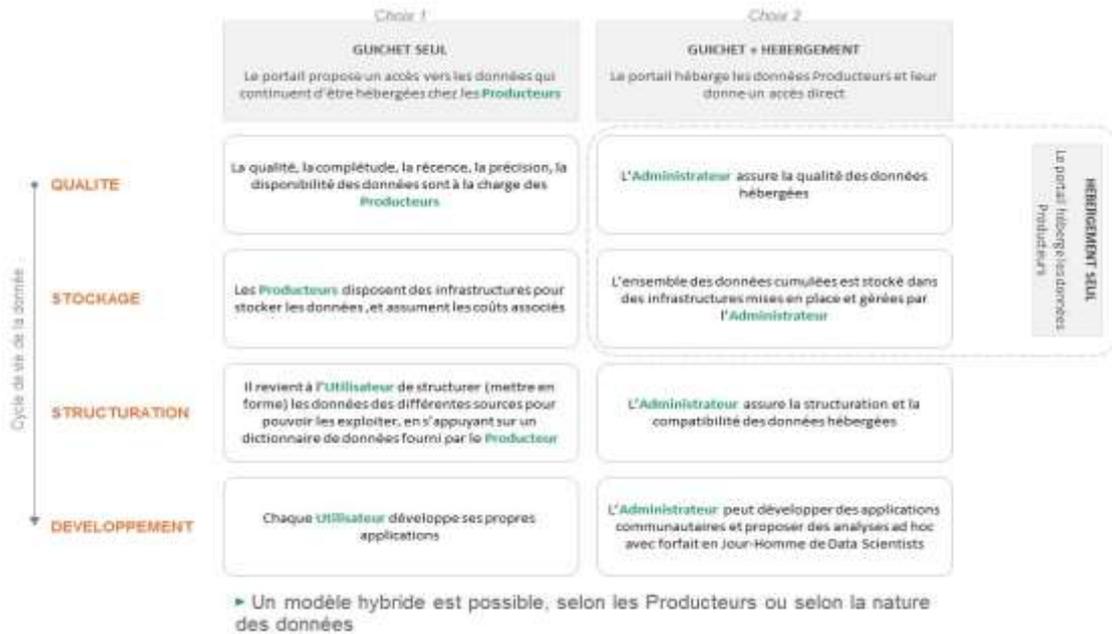


Figure 4 - Le partage des responsabilités en fonction des fonctionnalités choisies pour le portail (choix 1 : portail/guichet seul ; choix 2 : portail/guichet + hébergement).

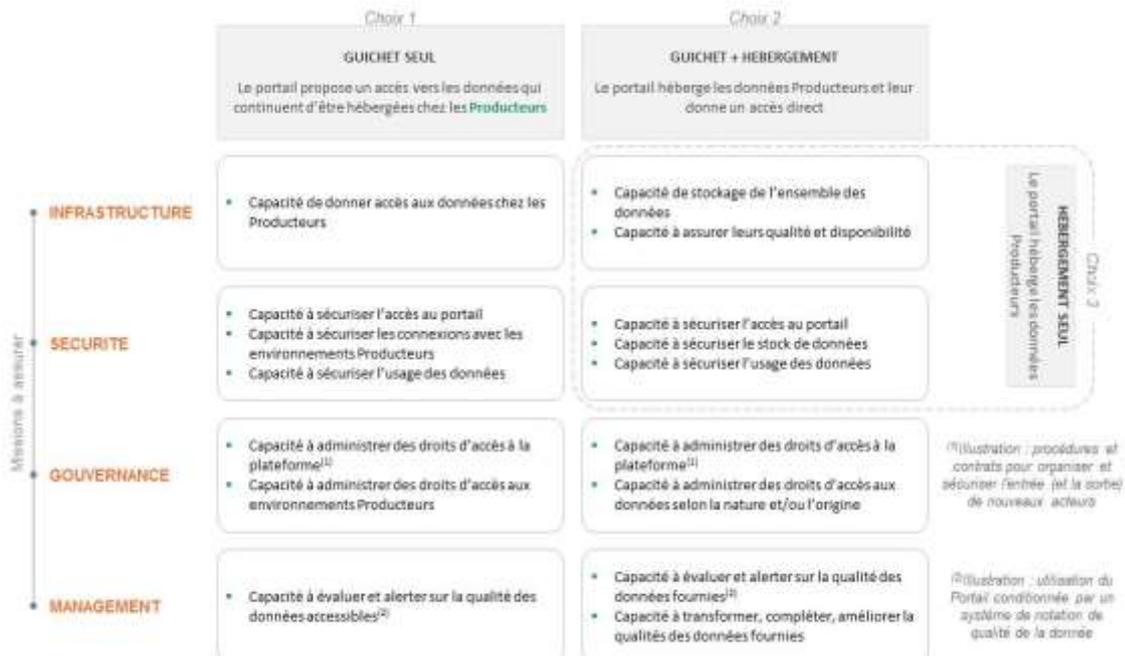


Figure 5 - Impact du choix du service rendu sur les missions de l'administrateur (choix 1 : portail/guichet seul ; choix 2 : portail/guichet + hébergement).

Points techniques critiques



6. Points techniques critiques

Le portail agricole AgGate sera une plateforme web accessible sur les navigateurs couramment utilisés (Internet Explorer, Firefox, Chrome, Safari et versions « mobiles »). La technologie utilisée devra respecter les standards du web. Il sera notamment nécessaire de préciser des mentions légales et d'insérer un bandeau relatif à l'utilisation des cookies. Les éléments exportables devront reposer sur des standards en vigueur et l'architecture principalement basée sur des technologies favorisant l'interopérabilité. Le portail devra donc être évolutif. Au-delà de ces considérations générales, plusieurs points sont apparus comme particulièrement critiques.

6.1 Recherche des données, intégration, métadonnées et sémantique

6.1.1 Intégration de données

Le moteur de recherche doit être capable de proposer les résultats d'une recherche en un minimum de temps, alors que les sources de données sont diverses et hétérogènes. Ainsi, une approche d'intégration de données doit être mise en place⁶. L'intégration de données doit fournir un accès uniforme (comme si c'était une seule base de données homogène) à des sources

- multiples
- autonomes (sans affecter leur comportement, indépendant des autres sources ou du système d'intégration)
- hétérogènes (différents modèles et schémas de données)
- structurées (ou semi-structurées).

L'intégration de données est une forte valeur ajoutée du portail. Elle permet de retrouver, de croiser des données et de les retourner vers l'utilisateur, afin que l'utilisateur « voie » le portail comme un guichet unique. La technologie offrant une **capacité d'intégration intelligente** et automatique constitue un accélérateur et un **facilitateur important (figure 6)**.

⁶ Özsü, M.T., Valduriez, P. (2011) Principles of Distributed Database Systems, Springer, 846 p. <http://www.springer.com/fr/book/9781441988331>

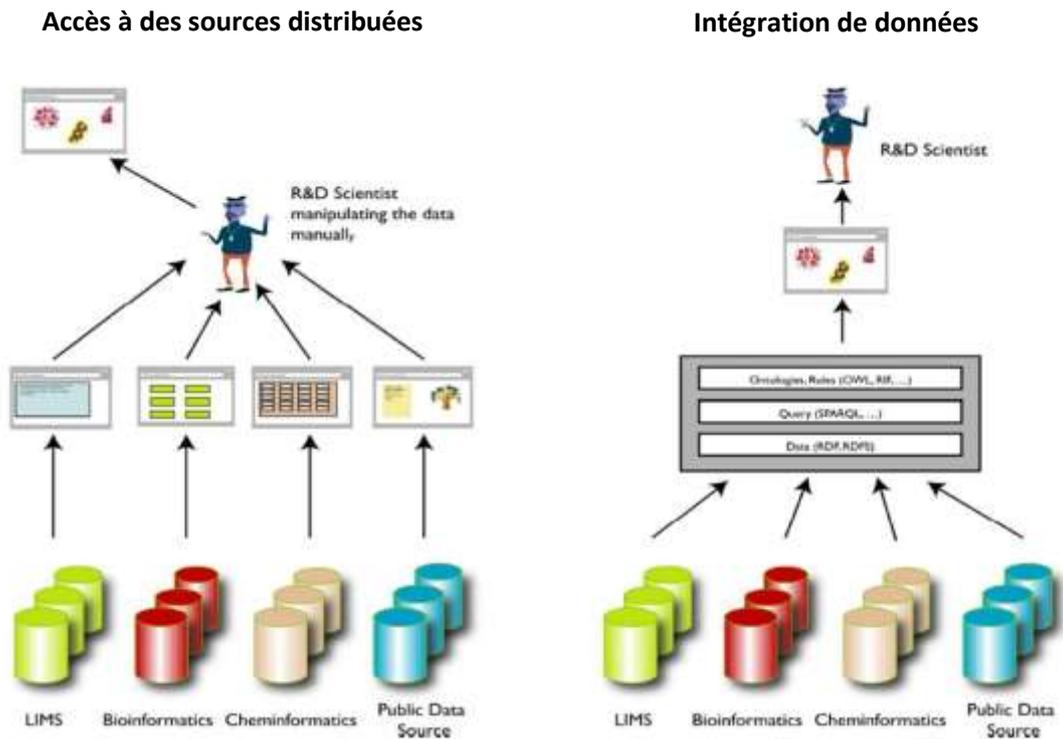


Figure 6 - Accès à des sources distribuées (à gauche) versus intégration de données (à droite) ; l'intégration rend transparent le mode de stockage des données et facilite l'accès aux données. (Source : D Vodislav, ENSEA)

Il y a globalement 2 logiques pour l'intégration de données:

- **Logique « entrepôt »** : les données sont moissonnées chez les fournisseurs, puis entreposées dans une base qui est gérée par le portail ; on parle de données « persistantes » car elles sont conservées par le portail ;
- **Logique « virtualisation des données »** : les données sont stockées dans des bases distribuées chez les fournisseurs de données et le portail joue le rôle d'un guichet unique ; l'objectif est d'utiliser ces données comme si elles constituaient un jeu de données homogènes : le guichet sera capable, à la demande, d'identifier les données d'intérêt dans les différentes bases, mêmes hétérogènes, et de les fédérer.

On trouve également des hybrides de ces 2 modes. A noter que si la fonctionnalité « espace de stockage/cloud agriculteurs » est retenue, le portail deviendra également fournisseur de données via ses propres bases de données.

Les avantages et inconvénients de ces 2 approches sont décrits dans le tableau suivant :

	Entrepôt	Bases de données Distribuées Fédérées
Avantages	<ul style="list-style-type: none"> • Maîtrise complète de la donnée • Accès ultrarapide 	<ul style="list-style-type: none"> • Adapté à un contexte « Big Data », à savoir où les données nouvelles ne sont pas connues a priori, et où les données sont hétérogènes • Sécurité et qualité des données de la responsabilité du fournisseur
Inconvénients	<ul style="list-style-type: none"> • Difficulté de faire évoluer la structure de l'entrepôt • Problème de manipulation si l'entrepôt contient de très grosses masses de données (moins adapté au big data) • Qualité et sécurité sont de la responsabilité du portail • Double stockage (chez le fournisseur et sur le portail), coût du stockage 	<ul style="list-style-type: none"> • Accès un peu moins rapide (mais reste temps réel) grâce aux technos Big Data) • Disponibilité de la donnée soumise à la politique et à la technologie du fournisseur

Il apparaît que la technologie la plus adaptée à l'objectif d'un guichet unique ayant accès à des bases hétérogènes est celle des bases de données fédérées, via l'approche « médiateur ». Dans l'approche « médiateur », les données restent dans les sources, les requêtes sont exprimées sur le schéma global, puis décomposées en sous-requêtes sur les sources et les résultats des sources sont combinés pour former le résultat final. Le moteur de recherche se base sur les métadonnées du schéma global (Figure 7).

Celle-ci exige la construction d'objets intermédiaires : un « médiateur » qui gère le catalogue global, et traduit les requêtes pour les sources données, et « l'adaptateur » (en général, un pour chaque source) qui reçoit les requêtes du médiateur et les traduit dans le langage de la source qu'il gère.

Chaque source (base) donne lieu à la construction d'un adaptateur spécifique, en utilisant le modèle de la base. La couche d'intégration/médiation est la seule interrogée par les utilisateurs, en mettant en œuvre un modèle global. Les métadonnées, qui contiennent les modèles de données, mais aussi des informations sur les licences, droits d'accès, de réutilisation et de propriété des données, par exemple, sont donc essentielles. Il est indispensable de s'assurer de la complétude de ces métadonnées pour assurer la performance du service. L'ensemble des métadonnées pressenties figure en Annexe 4.

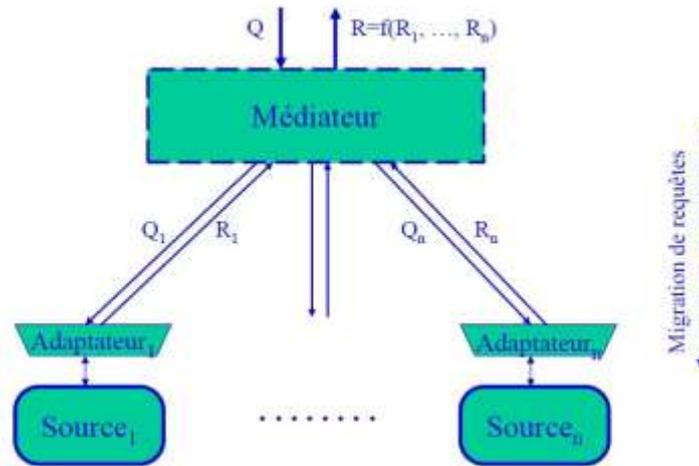


Figure 7 - Architecture d'intégration via un médiateur (intégration virtuelle) (source : D Vodislav, ENSEA)

Le guichet unique devra donc jouer le rôle d'un tiers de confiance avec les fournisseurs de données : si les API ne sont pas disponibles, il aura accès à leurs schémas de données et leurs référentiels (des valeurs prises par chacun des champs, par exemple les valeurs « blé », « maïs », « orge » pour le champ « culture »). La création du médiateur et des adaptateurs devra être prise en charge par l'administration du portail.

La mission ETALAB ayant beaucoup travaillé le sujet, il est intéressant de s'appuyer sur leurs résultats : le format choisi par la mission ETALAB est JSON-LD, avec l'utilisation du vocabulaire standard CSVW (du W3C).

6.1.2 Vers une ontologie des domaines « agriculture » et « environnement »

Si l'approche « médiateur » apparaît dans un premier temps comme la meilleure voie pour mettre en place rapidement un guichet unique, il peut être envisagé, pour améliorer le service de recherche d'intégrer des outils sémantiques et les standards W3C. Il serait notamment très utile de construire des ontologies car leur utilisation permet un requêtage plus précis (on peut préciser la sémantique) qu'avec uniquement le schéma du médiateur. Une ontologie est une formalisation logique, structurée et idéalement exhaustive⁷ des connaissances pour un domaine donné, réalisée en fonction d'un usage spécifique. Elle permet une communication entre hommes, entre machines ou hommes/machines. L'ontologie du domaine est utilisée pour représenter un domaine (par exemple, l'agriculture, l'environnement) sous forme de base de connaissances, en utilisant des concepts-clés, les attributs, les instances relatifs au domaine. Les ontologies, qui manipulent des concepts et leurs relations, peuvent être liées à une terminologie du domaine (définie par exemple par un thésaurus), ce qui permet de faire le lien avec la langue naturelle (on peut penser par exemple à passer de requêtes en langue naturelle dans un vocabulaire contrôlé construit sur la terminologie à des

⁷ Plus exactement « exhaustives vis-à-vis d'un usage spécifique »

requêtes informatiques construites sur l'ontologie). L'ontologie de domaine est donc un dispositif très efficace pour réutiliser les connaissances, ce qui est l'objectif du portail.

La création et l'utilisation d'une ontologie ad hoc passeraient par l'adaptation et la spécialisation d'ontologies et de ressources existantes (voir par exemple l'agroportail du LIRMM, en encadré). On propose donc un travail de recherche sur la mise au point d'ontologies en agriculture et en environnement, dans le cadre de la création de ce portail. Cette dernière peut s'appuyer sur des ontologies existantes afin d'en réduire le coût (à titre d'exemple, pour le projet Data-tourisme, le marché public visant à la création d'une ontologie spécifique au tourisme était de 75k€, la PME lauréate pouvant s'appuyer sur l'existant). Les données du portail et la communauté rassemblée par le portail seront mises à profit pour faire de la recherche sur les ontologies, par exemple, l'extraction automatique d'ontologies (cf chapitre « Animation », lien avec la recherche).

Focus sur l'Agroportail du LIRMM

L'agroportail (<http://agroportail.lirmm.fr/>) développé par le LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) met à disposition un ensemble de ressources pour les chercheurs en agriculture et agronomie voulant partager leurs données, dans les formats du web sémantique : thésaurus (SKOS), schéma de données (ontologie OWL), schéma avec données associées (jeux de données RDF). Un format CSV est aussi mis à disposition pour le téléchargement. Ce portail utilise une technologie déjà mise en œuvre dans le domaine biomédical (le bioportail de Stanford Stanford). Le portail permet de rechercher des ressources, de les visualiser, de les commenter, de les télécharger, de décrire les projets les utilisant. Le portail décrit chaque ressource en calculant un certain nombre d'indicateurs (nombre d'éléments dans la ressource, nombre de visites, etc). Chaque ressource peut être associée à un projet qui l'utilise afin de bien définir son usage dans un système d'information. Des outils sont mis à disposition sur le portail, par exemple un annotateur pour identifier les éléments d'une ressource apparaissant dans un texte fourni par l'utilisateur ou encore un outil pour visualiser les correspondances entre ressources. Le portail contient aussi une page pour présenter des indicateurs calculés sur l'ensemble des ressources mises à disposition. Pour faire évoluer leur portail et permettre à leurs utilisateurs de tester de nouvelles fonctionnalités le LIRMM a organisé un premier agrohackathon en juin 2016 à Montpellier.

6.2 Interconnexion

L'interconnexion envisagée dans le portail concerne la fusion de données de natures différentes provenant de plusieurs sources de données. Par exemple, une donnée de « qualité de sol » est fusionnée avec des données de rendement sur une parcelle. L'interconnexion ne concerne pas le regroupement d'individus de même nature (par exemple des parcelles) présents dans des sources différentes dans une même structure, mais la mise en relation de données de natures différentes.

Ces données doivent donc être mises en correspondance et interconnectées via un référentiel commun. Concernant les cultures, ce référentiel correspond aux coordonnées spatio-temporelles des données. Cela sous-entend que les données géographiques et temporelles sont interopérables d'un jeu de données à l'autre. Il est donc nécessaire de disposer des métadonnées caractérisant les

référentiels spatio-temporels utilisés (la projection et le fuseau horaire par exemple). Dans le cas de production animale, c'est l'identifiant de l'animal qui servira de référentiel.

Le "sas d'interconnexion" est un outil distinct du moteur de recherche, qui vient en complément. Les jeux de données et API identifiés par la recherche sont "envoyés" et utilisés dans ce sas pour réaliser l'interconnexion. L'interconnexion est une opération complexe car dans la plupart des cas, les références spatiales et/ ou temporelles des différentes bases de données à fusionner ne correspondront pas exactement ; on tombera alors sur un cas classique de fusion de données spatiales.

Suite à l'interconnexion, les coordonnées spatio-temporelles des jeux de données pourront être transformées (par translation, rotation ou autre méthode) pour les anonymiser (voir chapitre suivant). Il est important de bien valider le caractère anonyme des données suite à une interconnexion. En effet, le fait d'interconnecter des données peut accroître le risque de ré-identification : plus une position géographique sera documentée par le croisement des données, plus elle correspondra à un lieu unique.

6.3 Anonymisation

6.3.1 Pourquoi anonymiser ?

Les lois qui protègent la vie privée interdisent d'exploiter et de diffuser des données à caractère personnel, à moins que l'une des trois conditions suivantes soit remplie (cf. Annexe 6) :

- le **consentement des personnes concernées** a été recueilli après leur bonne information sur la finalité du traitement et les modalités de la communication ou de la réutilisation des données les concernant, sur l'identité du responsable du traitement, sur les droits qu'elles détiennent (ex : droit de rectification et d'opposition) et sur les destinataires ou catégories de destinataires des données,
- les **données sont anonymisées** (c'est-à-dire ne permettant pas, compte tenu de leur niveau d'agrégation, d'identifier les personnes concernées) ;
- la **réutilisation est autorisée par un texte** législatif ou réglementaire.

Certaines données d'une exploitation contiennent des données géographiques qui peuvent intégrer des données à caractère personnel et à ce titre, elles devront être anonymisées sauf à obtenir le consentement des personnes concernées sur la rediffusion de ces données.

Une exploitation agricole est une entreprise qui peut se prévaloir de la protection de ses secrets d'affaires. Définis ainsi, les « secrets d'affaires » sont des « informations dont non seulement la divulgation au public mais également la simple transmission à un sujet de droit différent de celui qui a fourni l'information peut gravement léser les intérêts de celui-ci ». ⁸ En se fondant sur cette jurisprudence, l'entreprise peut refuser la fourniture de données dont l'exploitation - et notamment le rapprochement avec des données d'autres sources - pourrait divulguer des informations

⁸ Tribunal de première instance des Communautés européennes dans son arrêt du 18 septembre 1996, Postbank c. Commission (affaire no T-353/94, Rec., II, p. 921, point 87)

susceptibles de nuire au détenteur/ producteur des données.

Le recours à l'anonymisation permet justement de réduire le risque d'une telle divulgation. L'anonymisation consiste à modifier le contenu ou la structure de données afin qu'il ne soit plus possible de ré-identifier les personnes physiques ou morales auxquelles elles sont rattachées. L'enjeu est de trouver le bon compromis entre anonymisation et utilité des données. En effet, retirer trop d'informations au cours de l'anonymisation générerait des données avec un intérêt plus limité. C'est le cas de l'agrégation : par exemple en agrégeant spatialement des grains supérieurs à la parcelle (ex : la commune) on perd une information qui peut être cruciale pour l'application mise en œuvre. Il est important de déterminer le seuil minimal d'anonymisation en évaluant le risque de ré-identification par une analyse préalable. Le niveau de risque acceptable sera fixé par la gouvernance.

Il faut néanmoins noter que toutes les données n'ont pas vocation à être anonymisées par le portail ; par exemple les données ne comportant aucune donnée à caractère personnel, ou les données que les fournisseurs accepteront de partager telles quelles (avec leurs données géographiques explicites) ne seront pas anonymisées.

6.3.2 Comment anonymiser ?

De nombreuses données agricoles sont référencées temporellement et géographiquement (par exemple les relevés d'opérations culturales dans les cahiers de cultures électroniques). Les données géographiques devraient faire l'objet d'une analyse spécifique en raison de leur fort potentiel de valeur et de leur fort potentiel de ré-identification. Pour les jeux de données dynamiques, leur potentiel de ré-identification est susceptible de varier dans le temps. Il est pertinent de le mesurer régulièrement.

L'analyse de l'état de l'art de l'anonymisation montre 3 familles de solutions :

- Des approches basées sur l'accroissement de granularité des données (suppression ou généralisation) ; au sens de la loi n° 51-711 du 7 juin 1951 modifiée sur le « secret statistique » (qui s'applique aux producteurs et diffuseurs de statistiques publiques) ;
- Des techniques de « confidentialité différentielle » basées sur la perturbation des données par exemple par l'injection de données fictives ;
- La réécriture des requêtes pour intégrer les contraintes d'anonymat (par exemple une requête qui demanderait la coordonnée géographique serait réécrite).

La 3^{ème} n'est pas forcément pertinente (cf. 6.3.3 « quand anonymiser »). Les 2 premières familles de solutions sont particulièrement adaptées, comme décrit ci-dessous :

Changement de granularité. Le « secret statistique » créé par un accroissement de granularité (agrégation par moyennage de données en changeant d'échelle spatiale) est le résultat de la loi n° 51-711 du 7 juin 1951 modifiée. Son objectif est d'assurer la confidentialité des informations des personnes physiques sur leur vie personnelle et familiale et des entreprises sur le secret commercial. Les informations publiées sur le site agreste (<http://agreste.agriculture.gouv.fr>) en sont un exemple.

Pour les données agricoles, le changement de granularité concernera l'espace (agréger les données sur plusieurs parcelles ou plusieurs communes), le temps (agréger des données sur la date de semis ou de traitement sur plusieurs jours) ou les individus (moyennes sur les animaux issus d'exploitations différentes)

Confidentialité différentielle. Cette méthode injecte une proportion connue de données fictives (mais réalistes) dans une population. On peut alors calculer les estimateurs de la moyenne ou de la médiane (mais les calculs d'extrêmes n'ont pas de sens). Cette technique s'applique surtout lorsqu'on cherche à protéger des données de géolocalisation, où il est facile de générer des données fausses « plausibles », et où les fonctions qu'on peut calculer avec cette technique d'anonymisation restent utiles (en particulier la densité et la distance).

Le référencement spatiotemporel de la donnée pose un véritable problème car un grand nombre de données de ce portail seront référencées géographiquement et temporellement. De manière générale, cette coordonnée ne sera pas révélée (sauf avec le consentement de son propriétaire) car elle peut amener à l'identification du propriétaire de la parcelle (voir risques identifiés en Annexe 6). Une des voies d'anonymisation pourra consister à changer les coordonnées géographiques, une fois l'interconnexion effectuée (transformation par translation, rotation ou autre).

En conclusion, rappelons que pour assurer une anonymisation efficace, le choix d'une technique, éventuellement combinée avec une autre doit s'appuyer sur une analyse du potentiel de ré-identification du jeu de données.

6.3.3 Quand anonymiser ?

En permettant de favoriser l'innovation basée sur des données d'intérêt agricole, ce guichet joue le rôle d'un "tiers de confiance" : dans les bases de données que les fournisseurs ont ouvertes, il a accès aux données non anonymisées afin d'être en capacité de les fédérer. Il retourne à l'utilisateur le résultat de la requête sous une forme d'inventaire des réponses (métadonnées), mais sans révéler de données à caractère personnel ou de données protégées par le secret des affaires. Les données seront anonymisées seulement lorsqu'elles seront délivrées à l'utilisateur, pour une visualisation ou un téléchargement, après une analyse du risque de ré-identification.

6.3.4 Conclusion – L'anonymat est un droit mais pas un objectif

L'anonymisation est un confort pour le producteur de données mais peut rendre difficile la tâche de l'utilisateur de données : en fonction des problèmes, elle rendra impossible la réutilisation de ces données conjointement avec une autre source de données que le consommateur de données pourrait avoir à sa disposition. En effet, le référentiel spatiotemporel qui permettrait de faire le lien entre ces données aura été supprimé ou modifié. L'intérêt pour le consommateur peut donc être fortement réduit. De plus, avec les problèmes d'interconnexions de données ayant des références géographiques non superposées, elle demandera un service supplémentaire au portail de mise en correspondance géographique (et qui pourra être payant). Un des enjeux du portail est donc de

collecter le plus fréquemment possible des données dans lesquelles la coordonnée géographique peut être ouverte. Cependant plusieurs niveaux sont possibles dans le brouillage du géoréférencement. Par exemple si on considère les données de l'itinéraire technique :

- Certaines données pourront être fournies avec le géo-référencement, par exemple, l'occupation du sol d'une parcelle, si elles ne permettent pas l'identification d'une personne physique ou ne contreviennent pas au « secret des affaires » ou si cela est autorisé par un texte réglementaire,
- D'autres données de production concernant la protection des cultures, la fertilisation etc) seront anonymisées vis-à-vis du géoréférencement, soit par agrégation, soit par brouillage du géoréférencement, en fonction de l'objectif d'usage de la donnée.

6.4 Sécurisation

Bien que l'approche privilégiée pour le stockage soit celle de bases de données distribuées fédérées via un médiateur, ce guichet unique pourra conserver des données (service cloud) contenant des informations personnelles. Il fera pour cela l'objet d'une déclaration à la CNIL (cf. Annexe 6).

L'accès au système devra être sécurisé. Le système devra être sûr et ne pas présenter de faille de sécurité connue. La mise à niveau régulière du dispositif de sécurité devra être intégrée. Les échanges monétaires devront être tout particulièrement sécurisés.

Il est à noter que l'approche « Médiateur » réduit le risque de sécurité par rapport à une approche « Entrepôt ». En revanche, l'hébergement d'un « cloud agriculteurs » rendra les problèmes de sécurité plus prégnants.

6.5 Ergonomie, design et charte graphique

Une attention particulière sera portée à l'ergonomie. Celle-ci devra être la plus simple et intuitive possible afin de ne pas rebuter les utilisateurs et de ne pas nécessiter de former les différents publics. Les interfaces pourront aussi être pensées et construites différemment selon les cas d'utilisation du guichet (utilisateur fournisseur, utilisateur consommateur de données). Le moteur de recherche sera immédiatement accessible et d'usage intuitif, de même que l'espace réservé (et soumis à identification) destiné à l'alimentation du portail.

Une visite virtuelle du portail pourra être proposée sur la page d'accueil.

Un système de FAQ pourra être mis en place en appui pour les utilisateurs. La facilité d'utilisation du portail et son ergonomie doivent limiter au maximum le recours à un module d'aide ou de guide d'utilisation du portail.

L'intégralité du contenu devra correspondre à une charte graphique personnalisée qui sera décidée en lien avec le prestataire.

Animation du projet



7. Animation du projet

7.1 L'équipe d'animation

Le projet sera animé par une équipe qui s'attachera à

- recenser les données accessibles et leur pertinence,
- négocier l'accès avec les fournisseurs,
- faire évoluer les capacités techniques du portail,
- communiquer autour du portail y compris en organisant des évènements ciblés.

Une des missions de l'équipe d'animation sera d'accompagner les fournisseurs dans la mise à disposition de leurs données, aussi bien d'un point de vue juridique (sur les problématiques de propriété/titularité des données) que d'un point de vue technique (pour accéder aux modèles de données et construire les adaptateurs).

7.2 Animer l'écosystème de l'innovation

La mise en place d'un écosystème ouvert d'innovation va nécessiter un plan d'action public-privé pour accompagner le développement du portail, sur la base d'actions dans la dynamique « French Tech » thématisée autour de l'agriculture : appels à projets, concours de startups, hackathons, compétition numérique, challenge numérique, renforcement des transferts technologiques depuis les organismes de recherche... Une panoplie d'actions à l'image de C3challenge ou de l'initiative greentech verte, lancée par le ministère de l'environnement en 2015-2016 (<http://c3challenge.com/>), mérite d'être mise en œuvre :

- 1- Vision Camp : Atelier d'intelligence et de créativité collectives pour faire émerger les besoins et les rêves partagés des parties prenantes ;
- 2- Challenge workshop : Atelier d'intelligence et de créativité collectives pour clarifier et enrichir les défis auxquels on veut répondre.
- 3- Innovation Jam : Marathon de l'innovation de 36 à 48h pour imaginer et développer des solutions innovantes aux défis posés.

On peut également envisager des financements plus conséquents de projets dans le cadre de Hackathon, sujets aux étapes 1 et 2.

Un premier hackathon pourrait être lancé avant l'ouverture du portail, à partir de bases mises à disposition, afin d'avoir des exemples de modèles ou de services créés à partir de données, dès l'ouverture du portail⁹. Les hackathons peuvent concerner (non exhaustif):

- la mise au point de modèles / services à partir des données du portail ;

⁹ Pour cela, une première équipe « ad hoc » sera constituée à partir de collaborateurs des sociétés qui portent le projet de création de la SAS ; elle sera rattachée opérationnellement au porteur de projet.

- la mise au point de petits applicatifs pratiques pour le « magasin numérique » ; la création d'ontologies et de référentiels communs.

Le portail valorisera les nouveaux services ainsi développés, notamment auprès des agriculteurs qui auront accepté d'y déposer leurs données.

7.3 Convaincre les producteurs et fournisseurs de données

Un point crucial est que les agriculteurs doivent comprendre l'intérêt d'une démarche d'ouverture de leurs données envers une multiplicité d'acteurs qui vont les valoriser sous forme d'outils d'aide à la décision. Cette approche, mise en œuvre au Danemark, est une alternative à la potentielle main-mise sur les données agricoles par des agro-fournisseurs dominants ou de grands acteurs de la donnée (les GAFA). L'objectif n°1 est :

- de construire la confiance en un système qui sera gouverné par la profession agricole, et qui ne dévoilera pas leurs données sensibles aux tiers non-autorisés ;
- de démontrer que la donnée n'a de valeur que lorsqu'elle est regroupée avec d'autres et partagée,

pour que les agriculteurs soient volontaires dans le partage de leur données.

Un plan de communication très solide est nécessaire. Des leviers divers doivent être mis en œuvre, parmi lesquels :

- La co-construction des activités du portail avec les agriculteurs, en particulier pour identifier les fonctionnalités prioritaires à implémenter, les outils attendus dans le magasin numérique, l'ergonomie... des brainstorming pourront être organisés dans différentes régions, pour toucher différentes filières et identifier des fonctionnalités nouvelles pour le portail dans une approche « design thinking » ;
- Des retours vers les agriculteurs sur l'utilisation des données du portail (par exemple, une chronique trimestrielle et bilan annuel), et de leur propres données (information systématique, statistiques annuelles etc), pour conserver le lien avec les agriculteurs ;
- Des retours vers les agriculteurs sous forme d'informations statistiques contextualisées, à savoir sur leur région, leur filière etc... (indicateurs, référentiels etc)
- Une information régulière sur les nouveaux outils du magasin numérique ou les OAD mis en ligne sur le portail (quand la fonctionnalité sera ouverte)
- Des approches plus ludiques : palmarès du dépôt de données, de la réutilisation de données, concours, etc

Un canal privilégié pour construire la confiance est celui des organisations agricoles, que le portail devra largement associer à l'animation.

La charte des fournisseurs de données et la charte des utilisateurs de données sont aussi des éléments cruciaux pour construire la confiance.

7.4 Liens avec la recherche

7.4.1 Une opportunité pour la recherche : trouver ou publier des données

Le portail est un outil exceptionnel dont la profession agricole va pouvoir se doter pour favoriser les innovations en agriculture numérique, avec les entreprises comme cible principale d'utilisateurs. La recherche est un autre utilisateur du portail qui est à la fois consommateur et fournisseur de données. Les données –à vocation agricole- produites pourront donc être déposées sur le portail et ouvertes, politique promue par l'action européenne, ce qui leur confèrera une visibilité exceptionnelle et une véritable facilité de réutilisation. Cette visibilité peut être accrue en publiant les données dans des data journals.

Comme indiqué plus haut, la création du portail devra être accompagnée d'une recherche spécifique sur les ontologies prioritairement dans le domaine de l'agriculture et ensuite dans les domaines de l'alimentaire et de l'environnement (thèmes à décider par la gouvernance). Une telle ontologie agricole est une attente forte de la profession agricole et des éditeurs de logiciels. Cette recherche préparera la nouvelle génération du portail, à savoir des performances de recherche accrue grâce à l'utilisation de l'ontologie et pourra servir de socle à une internationalisation du portail. L'autre intérêt, essentiel pour l'enjeu de l'appropriation du portail est que la création est une formidable opportunité de collaboration (1) avec la profession agricole, car une ontologie se travaille d'abord au sein de groupes de travail "métiers", ce qui renforce le lien avec les utilisateurs et les producteurs de données (2) avec d'autres projets en agriculture, d'autres acteurs, nationaux et internationaux. C'est aussi un important levier de compétitivité, car celui qui crée et gère l'ontologie est au cœur du système, y compris en termes de standardisation et de normalisation industrielle. Enfin, il faut bien noter qu'il y a un consensus au sein de la communauté scientifique pour que la construction d'une ontologie soit un processus "ouvert", et pas propriétaire, ce qui est tout-à-fait cohérent avec les objectifs du portail.

Au-delà de cette action de recherche spécifique, qui serait à mener en parallèle avec la création du portail, les outils classiques de financement de la recherche (ANR, FUI, H2020...) peuvent être mobilisés. Sans chercher à être exhaustifs, nous présentons ci-dessous trois dispositifs de recherche structurants - en partie issus des recommandations de la mission AI2025- qui peuvent établir un lien fort avec AgGate, le portail de données.

7.4.2 Le programme CASDAR

Le CASDAR est un programme alimenté par la taxe sur les exploitations agricoles porté par le ministère de l'agriculture (DGER) avec l'objectif d'innover pour faire évoluer les pratiques des agriculteurs en vue de répondre aux enjeux liés à la transition agro-écologique de l'agriculture française. Il a par exemple servi à financer le projet API-AGRO (cf 1.4.3). La mission AI2025 a recommandé de lancer un programme sur les capteurs, en particulier pour la détection précoce des

maladies. Privilégiant les TRL¹⁰ élevés et la collaboration « instituts techniques / recherche », le CASDAR peut être utilisé pour encourager les projets ayant vocation à :

- Développer des systèmes d'acquisition de données (capteurs, crowdsourcing...), ces données étant réutilisées dans le portail ;
- Développer des applicatifs à partir des données du portail ;
- Développer des méthodologies et des dispositifs, améliorant l'utilisation des données du portail, par exemple des ontologies, des systèmes de sécurisation/ anonymisation etc. ;

7.4.3 L'institut Convergences en agriculture numérique #DigitAg

Suite à la recommandation AI 2025 de créer un institut structurant la recherche en agriculture numérique, le projet #DigitAg (www.HDigitag.fr) a été lauréat de l'appel d'offre « instituts Convergences » de février 2016. Cet « institut Convergences en agriculture numérique », localisé à Montpellier, propose un programme de recherche et un programme de formation pour favoriser la diffusion de l'agriculture numérique. Les verrous à lever sont autant techniques qu'organisationnels ou sociaux, c'est pourquoi une équipe pluridisciplinaire de 200 ETP¹¹ a été créée (mêlant sciences humaines et sociales, sciences du numérique, sciences pour l'ingénieur, sciences de la vie) et le lien avec les entreprises est établi, en particulier via la formation. Les flux entre le portail et #DigitAg seront bidirectionnels : le portail fournira aux chercheurs des données pour faire avancer des applicatifs et des modèles ... mais également des méthodologies génériques, par exemple pour améliorer l'accès à la donnée agricole (ontologies...), son traitement (technologies big data) ou calculer la valeur des bases de données. Les avancées méthodologiques seront intégrées dans le portail.

7.4.4 Les Living Labs (ou laboratoires vivants)

Les Living Labs sont de nouveaux dispositifs de recherche, préconisés par la mission AI2025 dans lesquels une approche intégrée est développée à l'échelle territoriale de façon à tester en vraie grandeur des usages, des services et des outils. De nouvelles technologies, de nouvelles organisations, de nouveaux modes de production agricole peuvent ainsi être mis en œuvre et validés au regard de leurs performances économiques tout en tenant compte des spécificités. Les living Labs seront l'occasion de déployer de façon expérimentale des TICs d'acquisition de données (smartphones, objets connectés) chez les agriculteurs d'un territoire pour encourager leur mise en œuvre et avoir les retours directs sur leur facilité d'usage. Les données ainsi collectées pourront alimenter le portail. Il serait souhaitable que les Living Labs agricoles incluent tous une dimension de collecte et mise à disposition de données, à l'instar du réseau de fermes DEPHY.

Conclusion : Le portail peut alimenter la recherche ou être alimenté par la recherche, aussi bien en méthodes qu'en données. Il peut ainsi devenir un vecteur de l'ouverture de données de recherche, en couplage éventuel avec des publications de données dans des « data journals », en particulier pour les données issues des Living Labs.

¹⁰ Technologies Readiness Level

¹¹ Emploi Temps Plein

Mise en œuvre



8. Mise en œuvre

8.1 Calendrier

La mise en place du portail va s'effectuer graduellement pendant 3 ans, pour mettre en œuvre les 5 fonctionnalités identifiées (phases 1 et 2 ci-dessous). On distingue trois phases :

Phase 1 : Construction du guichet, cœur du système, s'achevant par l'ouverture du portail au public (12-15 mois)

Phase 2 : Lancement (construction des modules complémentaires, acquisition du trafic) et une phase de « croisière » avec un passage de versions nouvelles et une animation (2 ans)

Des recherches sur les ontologies et sur la chaîne de valeur sont menées en parallèle des phases 1 et 2.

Phase 3 : Croisière, phase d'utilisation du portail

Des échanges avec le cabinet de conseil Bearing Point ont permis de proposer un planning estimatif du travail de mise en œuvre de cette première phase.

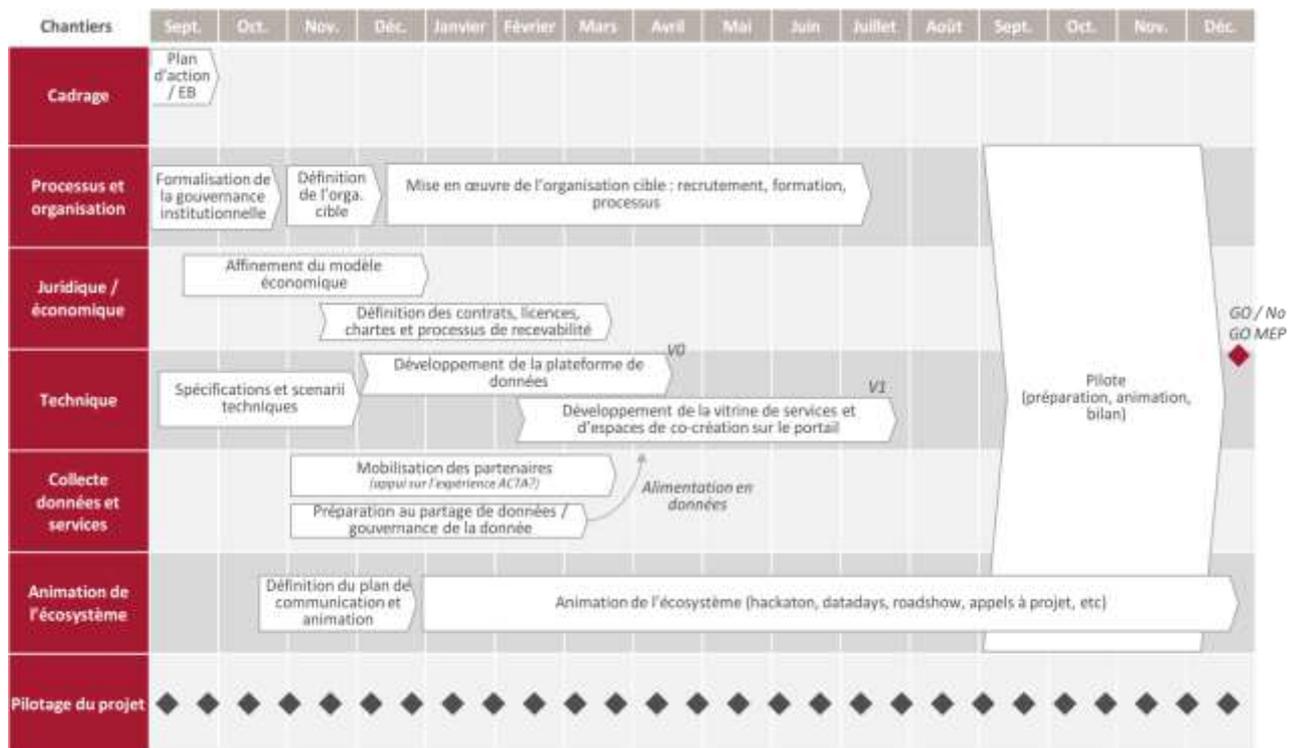


Figure 8 - Planning de mise en œuvre du portail de données agricoles AgGate à partir de sept 2016 (Source : Bearing Point).

La phase la plus cruciale est la première, à savoir la construction du guichet (12-15 mois). Différents chantiers ont été identifiés, avec l'aide de la société de conseil Bearing Point :

- **Chantier Juridique / économique**

Benchmark, cadrage du modèle économique, modélisation

Mise en place des contrats, licences, chartes, audit de recevabilité

- **Chantier Organisation / gouvernance**

Définition de la gouvernance institutionnelle

Définition de l'organisation cible

Définition des processus de fonctionnement / modes ops

- **Chantier Technique**

Spécifications et mise en œuvre du lot 1 (guichet et fonctionnalités pour l'acquisition de trafic)

AMO et développement de la plateforme v1

- **Chantier Pilote**

Préparation, animation et bilan du pilote

- **Chantier collecte données et services**

Mobilisation des partenaires et préparation au partage de données

Achats de certaines données / ressources / apps

- **Chantier Communication / animation**

Référencement initial (SEO)

Actions de communication et animation pour le lancement (conférence, hackaton, appels à projet, data days, publicité, emailing)

8.2 Evaluation économique préliminaire

8.2.1 Dépenses

Personnel, l'équipe projet

L'équipe projet doit être recrutée le plus rapidement possible en 2017. Elle doit intégrer entre 4 et 6 personnes, sur les profils suivants :

- un directeur, bien intégré dans le milieu agricole gestionnaire, connaissances juridiques
- un data management officer,
- un marketeur /communication
- un chef de projet informatique,
- un gestionnaire

Le **coût annuel** de cette équipe-projet (salaires chargés, environnés) est estimé à **750-800 K€** pour une implantation parisienne. Il peut y avoir du sens à installer cette équipe en province à proximité d'universités ou de centres de recherche travaillant sur le sujet des données en agriculture (ce qui se traduirait aussi par une réduction de l'ordre de 20% des coûts salariaux).

Phase 1- Construction (2017)

Pendant la phase 1, les coûts additionnels sont estimés comme suit

- Action de mise en place 700 – 900k€

(chantier juridique, économique, organisation, gouvernance, technique AMOA)

- Actions d'accompagnement pour l'animation (vision camp, challenge, hackathons...): 400 K€

En ajoutant les salaires, **la phase de construction est donc estimée à 1,85 – 2,1 M€ dont 400K€ d'accompagnement.**

Phase 2 -Lancement (2018-2019)

Pendant la phase 2, les coûts sont estimés comme suit :

- Prestation fournisseur de la plateforme (maintenance, hébergement, ...) 150-250 K€

- Assistance juridique récurrente 50 K€

- Actions d'accompagnement pour l'animation (vision camp, challenge, hackathons...): 300 K€

En ajoutant les salaires, **la phase de lancement est donc estimée à 1,25-1,4 M€/an soit 2,5 à 2,8 M€.**

Phase 1 et 2- Recherches ciblées (2017-2019) 500K€

Développement d'ontologies et de référentiels

Recherche sur la chaîne de valeur des données agricoles

L'accompagnement recherche des Phases 1 et 2 est donc estimé à 500 K€.

Au total le coût de la construction du portail (avec accompagnement recherche sur les ontologies) est de 4 850 – 5 400 K€

Phase 3- Croisière (à partir de 2020)

Pendant la phase 3, la prestation fournisseur devra être adaptée au trafic, et les actions d'accompagnement seront réduits, une fois le trafic acquis, et financés par le privé.

Pendant la phase 3, les coûts annuels sont estimés comme suit :

- Prestation fournisseur de la plateforme (maintenance, hébergement, ...)	150-250 K€
- Assistance juridique récurrente	50 K€
- Actions d'accompagnement pour l'animation (sponsoring):	150 K€

En ajoutant les salaires, **la phase de croisière est donc estimée à 1,1- 1,25 K€/ an**

8.2.2 Recettes

Construction du portail (2017-2019)

Phase 1 (2017)	
Contributions des actionnaires	1 400-1 800 K€
Subvention pour les challenges (coût challenges + salaires)	600 K€
SOUS- TOTAL Phase 1	2 000 - 2 400 K€
Phase 2 (2018-2019)	
Contribution des actionnaires	1 800 – 2 000 K€
Subvention pour les challenges (coût challenges + salaires)	900 K€
Publicité	100 K€
SOUS- TOTAL Phase 2	2 800 – 3000 K€
TOTAL RECETTES - CONSTRUCTION (2017-2019)	4 800 - 5 400 K€

Cette phase de construction peut être accompagnée par des financements publics.

L'accompagnement (animation, hackathons...) pourrait être financé par le ministère de l'agriculture et le secrétariat d'état pour le numérique en mobilisant notamment les outils mis en place dans le cadre de France numérique par le CGI (1,5 M€ sur 3 ans).

Le dispositif, qui a vocation à mutualiser les données pour favoriser l'innovation ouverte, peut être assimilé à un « pôle d'innovation »¹² au sens du CGI, et bénéficier d'avances remboursables dans le cadre du PIA, à hauteur de 50% des dépenses. Un projet pourra être déposé à France AgriMer

Dans ce dernier cas, le besoin en apport par les actionnaires (capital, cash et in kind) est de 3300-3900 K€ sur 3 ans, dont 50% d'aide sous forme avance remboursable.

Croisière (2021-) Budget annuel (Recettes)

Sponsoring privé des challenges (incluant les salaires)	200 K€
Paiement pour téléchargement de données intégrées et interconnectées	100 K€
Cloud des données agriculteurs : 100€/ an * 5000 agriculteurs (a)	500 K€
50% de réduction si 100% ouverture donnée : 50€/ an * 3000 agriculteurs	150 K€
Publicités sur page d'accueil (CPC, CPA, CPM, CPL, CPV)	50 K€
Pages pub dans la vitrine technologique	50 K€
TOTAL RECETTES PHASE CROISIERE	1 150 K€

(a) Ces chiffres ne proviennent pas d'une étude de marché mais d'une évaluation à dire d'experts.

En phase de croisière, l'équilibre budgétaire est assuré par la vente d'un service de stockage sécurisé de la donnée, à l'instar de celui proposé par la AgDataCoalition, par la vente d'espaces publicitaires et de données, et par du sponsoring privé (voire public/privé) pour assurer l'animation. Ce modèle économique n'est pas unique ; il repose sur la mise en place rapide de l'espace de stockage. En conclusion sont proposées d'autres approches pour contribuer au modèle économique.

¹² Pour bénéficier de ces niveaux d'aides, des conditions sont applicables sur les entreprises actionnaires de la SA ou SAS ; en particulier, aucun groupe ou ETI ne doit détenir plus de 25% du capital.

Conclusion



9. Conclusion

9.1 Les enjeux du portail

Les enjeux du portail sont résumés dans le tableau ci-dessous. Les points clés en sont 1) l'excellence, à savoir la capacité à retrouver des données pertinentes, 2) la maîtrise des risques techniques (anonymat, sécurité) et d'image, 3) la maîtrise des coûts et 4) la maîtrise du modèle économique. Pour celui-ci, plusieurs approches sont proposées, qui peuvent éventuellement être mises en œuvre de manière hybridée.

EXCELLENCE	<ul style="list-style-type: none">Proposer un catalogue de produits et services à jour et adapté aux besoinsDisposer des capacités et compétences pour offrir les meilleures performancesSécuriser les données avec les méthodes de cryptologie les plus avancées																								
RISQUES	<ul style="list-style-type: none">Instruire les responsabilités juridiques concernant la gouvernance et la sécurité des données à partir d'un socle réglementaire dense mais complexe et peu matureAnticiper et prévenir le risque d'image																								
POSTES DE COÛTS	<table border="1"><thead><tr><th>IMPLEMENTATION</th><th></th><th>FONCTIONNEMENT</th><th></th></tr></thead><tbody><tr><td>Capacité de stockage</td><td>●○○</td><td>Run des services</td><td>●○○</td></tr><tr><td>Logiciels exploitation, sécurisation</td><td>●●○</td><td>Ouverture d'un flux</td><td>●●○</td></tr><tr><td>Développement des services</td><td>●●○</td><td>Nouveaux services et produits</td><td>●●○</td></tr><tr><td>Ressources</td><td>●●●</td><td>Ressources</td><td>●●○</td></tr><tr><td>Acquisition de données externes</td><td>●●○</td><td>Nouvelles données externes</td><td>●●○</td></tr></tbody></table>	IMPLEMENTATION		FONCTIONNEMENT		Capacité de stockage	●○○	Run des services	●○○	Logiciels exploitation, sécurisation	●●○	Ouverture d'un flux	●●○	Développement des services	●●○	Nouveaux services et produits	●●○	Ressources	●●●	Ressources	●●○	Acquisition de données externes	●●○	Nouvelles données externes	●●○
IMPLEMENTATION		FONCTIONNEMENT																							
Capacité de stockage	●○○	Run des services	●○○																						
Logiciels exploitation, sécurisation	●●○	Ouverture d'un flux	●●○																						
Développement des services	●●○	Nouveaux services et produits	●●○																						
Ressources	●●●	Ressources	●●○																						
Acquisition de données externes	●●○	Nouvelles données externes	●●○																						
MODELE ECONOMIQUE	<p>Approche « consortium » : partage des coûts</p> <ul style="list-style-type: none">Les coûts sont partagés entre les acteurs : plus ils sont nombreux, plus le prix baisseCela nécessite de réunir une masse critique d'acteurs pour maîtriser les coûtsCela nécessite de réunir suffisamment de sources de données complémentaires pour générer de la valeurUn financement initial par subvention peut résoudre une partie du problème <p>Approche « App-store » : place de marché</p> <ul style="list-style-type: none">Les Développeurs et Producteurs paient pour référencer leurs applications et leurs donnéesLibre à eux ensuite de monétiser l'audience des Utilisateurs (bénéficiaires)L'enjeu est d'ouvrir le Portail à suffisamment d'Utilisateurs pour attirer les Producteurs et Développeurs, et générer un cercle vertueux de valeur <p>Approche « Amazon Web Services » : paiement à l'utilisation</p> <ul style="list-style-type: none">Les prix sont définis en fonction de l'usage que font les acteurs du Portail (calcul, transfert, stockage)Cela revient à une approche de partage de coût proportionnel à l'usage <p>Approche « Netflix » : abonnement</p> <ul style="list-style-type: none">L'Administrateur achète des droits d'utilisation des données et applications développées et les refacture aux Utilisateurs (bénéficiaires) sous forme d'abonnement																								

Figure 9 - Résumé des enjeux liés à l'ouverture du portail

9.2 Facteurs de succès et risques principaux

9.2.1 Appropriation et confiance

Le succès du projet réside tout d'abord dans le partage d'un maximum de données. Il conviendra notamment que les détenteurs de données publiques qu'il s'agisse des ministères, des organismes publics et para publics ou des entreprises détenant des données ayant fait l'objet de financements publics, au sens de la réglementation en vigueur, rendent leurs données librement accessibles via le portail.

De plus un producteur de données aura un intérêt à déposer ses données s'il trouve en retour une information qui l'intéresse. D'où l'importance de proposer dès la version 1 du portail AgGate des services offrant des retours rapides vers les producteurs (agriculteurs) et les fournisseurs (éditeurs). Ces services seront identifiés via des actions de « design thinking » avec les agriculteurs et avec les éditeurs. Enfin, la confiance des producteurs de données est un élément clé pour qu'ils demandent à

leurs prestataires gestionnaires de données (éditeurs de logiciels, agro-équipementiers) d'ouvrir leurs données au portail. Le rôle des acteurs de ce dispositif devra être clairement exposé pour être bien compris, approprié, et conduire à un jeu collectif, d'autant que certains (par exemple la recherche, les éditeurs de logiciels) peuvent être à la fois fournisseurs et utilisateurs de données. La figure 10 décrit les rôles des différents acteurs.



Figure 10 - une réussite fondée sur une bonne répartition des rôles entre acteurs, l'appropriation par chacun des acteurs et un jeu collectif

Enfin, les agriculteurs ne sont pas les seuls concernés par la production de données, les acteurs de la recherche ou des territoires génèrent également des données d'intérêt. La mise en place du portail peut être une opportunité pour ces acteurs publics de mieux valoriser les données.

Au-delà des agriculteurs, il est indispensable que les fournisseurs potentiels de données (éditeurs, agro-équipements) trouvent un intérêt dans la démarche. Les outils du magasin numérique sont une première réponse, mais ces acteurs peuvent être particulièrement intéressés par des actions de standardisation¹³. La recherche d'une ontologie de domaine (agriculture voire environnement ou alimentaire) s'inscrit parfaitement dans cette optique et sera une excellente occasion de fédérer les acteurs et de faciliter l'appropriation.

¹³ Rappelons que la société DKE-data créée par des agro-équipementiers en Allemagne décrites dans le benchmarking (1.4.5), qui sont précisément de créer un hub avec une standardisation des données pour faciliter les échanges.

9.2.2 Performance technique

Un autre critère de succès est technique et a trait aux utilisateurs de données ; il porte sur la capacité du moteur de recherche et des procédures d'interconnexion des données et des outils de traitement de données, permettant d'utiliser de manière transparente des données issues de sources différentes.

9.2.3 Les enjeux du partage de la valeur

L'objectif du portail est de mettre les données à disposition pour les valoriser, sous forme d'OAD ou de services nouveaux. Les données vont donc être à l'origine d'une création de valeur. Sans données, pas de création de valeur ! C'est sur cette base que des questions se posent aujourd'hui quant à la rémunération de la donnée, et ce dans tous les secteurs économiques. Concernant les données agricoles, certains collectifs d'agriculteurs, comme la COPA-COGECA (association des agriculteurs et des coopératives européennes)¹⁴, ont pris position sur le fait qu'il était « important de veiller à ce que les agriculteurs obtiennent une part équitable de la valeur générée par les données agricoles ». Ce questionnement rejoint celui de la création de la valeur sur la chaîne de la donnée. La chaîne de valeur de la donnée est la suivante¹⁵ :

- 1- création de données
- 2- stockage de la donnée
- 3- analyse de la donnée pour en tirer une information (ou un service)
- 4-usage de l'information par un système automatique (ou du service par un utilisateur).

La valeur est créée à toutes les étapes de la chaîne, mais de manière non linéaire. La question de la quantification de la valeur (ajoutée) à chaque étape de la chaîne est complexe, ou plus simplement de la valeur de la donnée de base est complexe et peut être abordée de diverses manières. Les différents facteurs qui font la valeur de la donnée sont décrits dans la Figure 11. L'équation qui modélise l'ensemble n'existe pas encore. Des recherches sont nécessaires sur ce point, elles pourront être menées dans le cadre du montage du portail, en association avec l'institut Convergences #DigitAg.

¹⁴ COPA-COGECA (2016) Grands principes régissant la collecte, l'utilisation et l'échange de données agricoles, QJ(16)2689:5 – DA/FG/mvs, position paper publié le 28-09-2016 sur <http://www.copa-cogeca.be/Main.aspx?page=Papers>

¹⁵ Gustafson T., Fink D. (2013) Winning within the Data Value Chain, <http://www.innosight.com/insight/winning-within-the-data-value-chain/>



Figure 11 : Les facteurs qui construisent la valeur d'un jeu de données (source : digital ethics)

Un premier point d'attention est que le coût de la donnée ne soit pas prohibitif pour certains acteurs, et qu'il soit équitable pour chacun (à savoir qu'il garantisse le même accès à tous, start-up ou grand groupe).

Cependant, si la question du juste retour financier vers le producteur de données se pose, le coût d'opportunité de ne pas entrer dans une démarche d'ouverture de données serait beaucoup plus élevé : il présenterait le risque de laisser de grosses structures s'emparer du monopole des données et de l'ensemble de la chaîne de valeur. L'enjeu de la juste rémunération est donc moindre que celui de l'indépendance vis-à-vis des fournisseurs.

9.3 Recommandations

Créer un portail unique au monde

- 1- Compléter la fonctionnalité « guichet des données » par d'autres services, utiles aux agriculteurs (cloud, magasin numérique, espace d'échanges, espace d'évaluation des outils d'aide à la décision);
- 2- Référencer des données publiques, des données de la recherche et des données privées ;
- 3- Faire du guichet un hub de données distribuées chez leurs fournisseurs, en travaillant en mode « réservoir de données/ médiateur » pour interconnecter les données issues de sources différentes ;
- 4- Etablir un plan d'évolution du portail, focalisé sur la performance des recherches, via la construction d'ontologies ;
- 5- Essaimer à l'international, en particulier au niveau européen. Un débat organisé à l'initiative de la Commission européenne sur la numérisation de l'agriculture, partant d'un bilan de l'existant permettrait de mettre en valeur la démarche originale d'un portail numérique visant à développer en Europe l'innovation dans ce secteur.

**Gagner la confiance de la profession agricole et
la sensibiliser aux enjeux de l'agriculture à base de données :**

- 6- Créer une société indépendante, gouvernée par la profession agricole, pour gérer le portail ;
- 7- Développer des services offrant un retour rapide vers le fournisseur de données (cloud, magasin numérique), en impliquant la profession agricole dans une démarche de design thinking;
- 8- Assurer l'anonymat des données sensibles lorsque celui-ci est demandé ou requis ;
- 9- Assurer la sécurité du portail ;
- 10- Sensibiliser la profession agricole à l'agriculture numérique et à l'intérêt de la démarche pour qu'ils demandent aux entreprises qui gèrent leurs données de les ouvrir au portail ;
- 11- Mener des recherches complémentaires pour créer des ontologies et évaluer la valeur le long de la chaîne de valeur de la donnée agricole.

Animer le portail pour assurer sa visibilité, son utilité et sa pérennité

- 12- Accompagner le portail d'actions de valorisation de la donnée mobilisant les utilisateurs (vision camp, compétitions, challenges, hackathons) et communiquer autour des résultats ;
- 13- Communiquer régulièrement vers les fournisseurs de données, en particulier les agriculteurs ;
- 14- Encourager les liens entre le portail et les dispositifs de recherche pour une fertilisation croisée ;
- 15- Encourager la réutilisation entre le portail et les autres initiatives proches (Agrosyst, API AGRO)
- 16- Choisir un modèle économique et travailler sur le partage de la valeur (recherche)

ANNEXES



ANNEXES

10.1 Annexe 1 - Lettre de mission pour JM Bournigal



**MINISTÈRE DE L'ÉDUCATION NATIONALE,
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE**

LA MINISTRE

LE SECRÉTAIRE D'ÉTAT
CHARGÉ DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE

**MINISTÈRE DE L'AGRICULTURE
DE L'AGROALIMENTAIRE ET DE LA FORÊT**

LE MINISTRE
PORTE-PAROLE DU GOUVERNEMENT

**MINISTÈRE DE L'ÉCONOMIE
DE L'INDUSTRIE ET DU NUMÉRIQUE**

LE MINISTRE

LA SECRÉTAIRE D'ÉTAT CHARGÉE DU NUMÉRIQUE

Paris, le **15 AVR. 2016**

Monsieur le Président de l'Institut
national de recherche en sciences et
technologies pour l'environnement
et l'agriculture (IRSTEA)
1, rue Pierre-Gilles de Gennes
92761 Antony Cedex

Objet : mission relative à la mise en place d'un portail de données pour l'innovation en agriculture

Vous avez bien voulu participer à la mission « Agriculture-Innovation 2025 », dont les conclusions nous ont été remises le 22 octobre 2015. Nous vous renouvelons nos remerciements pour les travaux conduits dans le cadre de cette mission, qui ont permis de clairement identifier les défis que l'agriculture a à relever pour rester compétitive et respectueuse de l'environnement.

Le rapport souligne l'enjeu important que représente l'exploitation des données numériques pour le monde agricole. La multiplication des sources de données, combinée à des capacités renforcées pour leur stockage et leur traitement, rend possible le développement de nouveaux services innovants pour l'ensemble des acteurs de la chaîne de valeur : agriculteurs, organisations professionnelles, industriels du secteur...

Afin de répondre à cet enjeu, il a été proposé la mise en place d'un portail de données pour l'innovation en agriculture. L'objectif de ce portail est de mettre à disposition

de l'écosystème agricole un ensemble important de données et d'informations, issues de diverses sources : données publiques, données des entreprises, données des agriculteurs (qui en conserveront la maîtrise)...

En exploitant les données et informations accessibles depuis ce portail, des acteurs économiques innovants, notamment des startups, pourront développer de nouveaux cas d'usage et des services à destination du monde agricole, dans des démarches d'innovation ouverte. La mise en place du portail pourra ainsi permettre aux agriculteurs de conserver la maîtrise de leurs données et constituera une opportunité pour faire émerger des solutions innovantes pour tous les acteurs de la filière.

L'ensemble des consultations menées depuis la remise du rapport a confirmé le vif intérêt des acteurs du monde agricole pour le portail de données qui a été proposé. Le comité de pilotage de la mission « Agriculture-Innovation 2025 » du 4 février 2016 en a dès lors validé le principe. Le comité a néanmoins souligné la nécessité de conduire des travaux complémentaires en vue de la mise en œuvre opérationnelle du portail. Il a en outre souhaité qu'un plan d'actions soit défini afin de stimuler l'émergence de nouveaux services numériques pour le monde agricole.

Dans ce contexte, nous vous confions une nouvelle mission visant d'une part, à préparer la mise en place opérationnelle du portail de données ; d'autre part, à définir un plan d'actions pour favoriser l'émergence de nouveaux services portés par des acteurs innovants, à partir des données mises sur le portail.

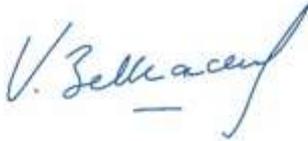
Vos travaux viseront notamment à formuler des propositions concrètes, en liaison avec la profession agricole, sur les questions suivantes :

- la structure juridique du portail et le modèle économique associé ;
- la gouvernance du portail. Vous veillerez à ce que celle-ci associe des représentants de l'ensemble des composantes du monde agricole et l'Etat ;
- le plan de financement du portail. Vous définirez notamment les contributions attendues des membres de la gouvernance, ainsi que le besoin de financement public (en liaison avec le Commissariat Général à l'Investissement) ;
- les jeux de données nécessaires pour le développement de services innovants. En liaison avec les entités représentatives de la profession agricole, vous engagerez notamment des discussions avec les détenteurs de données privées (agriculteurs, agro-équipementiers...) afin que celles-ci soient mises sur le portail. Ces discussions pourront notamment aboutir sur la définition de contrats-type ;
- la sécurité des données mises sur le portail, ainsi que leur anonymisation. Vous vous appuierez notamment sur les recommandations que pourront formuler la CNIL et l'ANSSI ;
- les conditions de développement d'un écosystème industriel innovant autour des données du portail ;
- la définition d'un calendrier prévisionnel de mise en place du portail.

Pour la conduite de la mission, vous pourrez vous appuyer sur les services concernés de nos Ministères respectifs. Vous veillerez en outre à ce que les conditions de mise en place du portail permettent à terme son extension au niveau européen.

Vous voudrez bien nous faire part de votre analyse en juin 2016.

Najat VALLAUD-BELKACEM



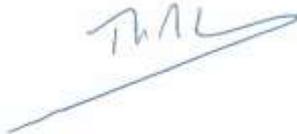
Stéphane LE FOLL



Emmanuel MACRON



Thierry MANDON



Axelle LEMAIRE



10.2 Annexe 2 – Questionnaire pour la caractérisation des bases de données agricoles

Portail de données à vocation agricole

*** OBJECTIF ***

Bonjour,

Un des axes de la mission "Mise en place d'un portail de données à vocation agricole" consiste à élaborer un cahier des charges de ce futur portail, en particulier en formulant des propositions concrètes sur les « jeux de données nécessaires au développement de services innovants ». Dans cette optique, nous dressons un inventaire des données publiques et privées disponibles et pouvant être mobilisées : type de données, propriétaire, objectif du recueil, qualification des données (validation), fréquence de mise à jour...

*** COMMENT REpondre A CE QUESTIONNAIRE ? ***

Le questionnaire ci-dessous vous guidera afin de renseigner un maximum d'informations concernant les bases/jeux de données à vocation agricole dont vous disposez. Nous reprendrons contact avec vous par la suite si nous avons des questions complémentaires.

Si vous avez plusieurs bases/jeux de données merci de dupliquer ce fichier et de le renseigner plusieurs fois.

Pour les questions à choix multiples merci de compléter le champ "Autres possibilités, précisions, commentaires" avec le plus de détails possibles.

Je vous laisse mes coordonnées. N'hésitez pas à me contacter si besoin.

Merci d'avance pour votre participation.

Charles MOSZKOWICZ - société ENEO

06 79 92 50 79

charles@eneo.fr

1. CONTACT (plusieurs contacts possibles)

Nom - Prénom - Structure - Fonction *

Email *

Téléphone

2. BASE DE DONNEES

Nom complet de la base de données (tel qu'il apparait dans les textes officiels pour les administrations)

Acronyme de la base de données s'il y en a un

Description de la base : contenu, taille et couverture de la base de données

ex : La base regroupe des données de traçabilité de l'itinéraire technique et rendement de 1000 exploitations uniquement dans le Sud-Ouest de la France de 2006 à maintenant.

Qui est le gestionnaire de la base de données? (pour les administrations merci d'utiliser le nom tel qu'il apparait dans [l'annuaire des services publics](#))

Qui est/sont le(s) fournisseur(s) de données? Quels organismes alimentent la base?

La base contient-elle des données personnelles?

oui

non

Si oui, Fait-elle l'objet d'une déclaration CNIL ?

oui

non

Autres possibilités, précisions, commentaires

Textes : textes officiels publiés par des administrations faisant référence à la base de données.

décret, arrêté, délibération, rapport, code, loi, réponse, convention, circulaire, article...

Quels sont les outils de collecte?

- Logiciel
- Capteur
- Questionnaire
- Guide d'entretien

Autres possibilités, précisions, commentaires

3. DONNEES

Données brutes présentes dans la base (le plus exhaustif possible). Merci de mettre ces informations les unes en dessous des autres.

ex : interventions effectuées dans chaque parcelle avec

- date
- numéro de parcelle
- type d'intervention
- durée
- type de produit
- dose

Granularité/finesse des données brutes. Les données brutes concernent quelle(s) échelle(s)?

- individu (plante ou animal)
- intraparcellaire
- parcellaire
- cheptel
- lot
- atelier
- exploitation
- groupement d'exploitations
- échelle départementale ou régionale
- échelle nationale

Autres possibilités, précisions, commentaires

La donnée est-elle géolocalisée? A-t-elle des coordonnées géographiques?

Si oui, merci de préciser la projection utilisée. Si non pouvez-vous préciser le référentiel géographique que vous utilisez éventuellement (cadastre, Registre Parcellaire Graphique...)?

oui

non

Autres possibilités, précisions, commentaires

Fréquence d'acquisition des données brutes

continu

horaire

quotidien

hebdomadaire

mensuel

annuel

Autres possibilités, précisions, commentaires

Fréquence de mise à jour des données dans la base

continu

horaire

quotidien

hebdomadaire

mensuel

annuel

Autres possibilités, précisions, commentaires

Délai moyen entre l'acquisition et la disponibilité de la donnée dans la base

Existe-t-il un système de vérification d'erreurs (de saisie, d'incohérence, de manques)?

Disposez-vous de métadonnées?

ex : précision de la mesure, précision de la géolocalisation, source de la donnée (capteur, manuelle), référentiel, ontologies

Avez-vous des données traitées, analysées, agrégées à partir de ces données brutes?
Statistiques, calculs d'indices... ? Si oui, lesquels? Précisez l'échelle géographique et l'échelle de temps.

4. ACCES

Liberté d'accès aux données

- Accès à tous : données accessibles au public
- Accès restreint (dans ce cas merci de préciser la restriction)
- Accès fermé : données non accessibles

Autres possibilités, précisions, commentaires

ex : accès autorisé seulement à certaines catégories de personnes, par exemple l'administration, les chercheurs, certains organismes professionnels

Coût d'accès aux données (pour un futur utilisateur du portail). Souhaitez-vous fournir ces données gratuitement ou contre rémunération?

- Gratuit pour tous
- Payant pour tous
- Tarif selon le profil d'utilisateur (dans ce cas merci de préciser votre réponse)

Autres possibilités, précisions, commentaires

ex : gratuit seulement pour la recherche, prix plus accessible pour instituts techniques...

Format de mise à disposition des données

- Exploitable : par exemple fichier tableur (CSV, Excel), JSON
- Non exploitable : par exemple fichier PDF, consultation en ligne

Autres possibilités, précisions, commentaires

Si les données sont déjà accessibles sur un portail, les conditions de réutilisation sont elles explicitées?

- Explicitées sans restriction :

conditions générales d'utilisation du portail d'accès aux données autorisant leur réutilisation y compris à des fins commerciales, à la condition éventuelle de mentionner la source.

- Explicitées avec restriction :

conditions générales d'utilisation du portail d'accès interdisant certaines réutilisations des données

- Non explicitées :

absence de conditions générales d'utilisation sur le portail d'accès aux données

Autres possibilités, précisions, commentaires

Adresse URL d'accès aux données (lien profond si disponible)

Datasets : Il s'agit de la dernière partie de l'url du jeu, pouvant être trouvée en haut du navigateur.

Ex : <https://www.data.gouv.fr/fr/datasets/balances-comptables-des-communes/>

Références : les éventuelles adresses web évoquant la base de données.

Remarque : Il ne s'agit pas du portail officiel de la base, ce n'est pas l'URL d'accès aux données.

5. AUTRES

Donnez un extrait de la base à titre d'exemple (10 à 20 lignes) et/ou le schéma de la base si possible.

Commentaires et remarques éventuelles

ex : données manquantes mois de Juillet 2008 sur la Beauce

En tant qu'utilisateur (et non plus fournisseur de données), de quelles données souhaiteriez-vous disposer sur le portail?

10.3 Annexe 3 - Extrait de l'inventaire des bases de données à caractère agricole

Comme précisé dans le paragraphe "3.2 Origine des données", les informations récoltées dans le cadre de l'inventaire sont issues de 4 sources :

- Les entretiens menés en présentiel ou par téléphone par le groupe de travail de l'Irstea en charge du projet AgGate en utilisant le questionnaire présenté en Annexe 2. Ce travail a été réalisé entre mai et septembre 2016.
- Une démarche d'identification et de caractérisation des bases de données publiques menée par le MAAF en s'appuyant sur ce même questionnaire. Ce travail a été réalisé entre mai et septembre 2016.
- Les entretiens menés dans le cadre du groupe "Connaissance des pratiques" du GIS GC HP2E, réalisés en 2016 par Agnès SALAÜN, chargée de mission « Connaissance des pratiques » du GIS GC HP2E (a.salaun@arvalisinstitutduvegetal.fr). Ce groupe est animé par Valérie LEVEAU, Responsable de l'équipe Economie et Systèmes de production, Service Agronomie - Economie - Environnement, Arvalis Institut du végétal (v.leveau@arvalisinstitutduvegetal.fr) et Eric CAHUZAC, INRA Toulouse, US-Observatoire du Développement Rural (eric.cahuzac@toulouse.inra.fr). Les entretiens ont été menés en utilisant un questionnaire semblable à celui présenté en Annexe 2. Il est à noter que la personne enquêtée ne fait pas nécessairement partie de la structure qui possède la base de données.
- Le document « GISEDSEmParis120424-Recueil fiches.pdf » du GIS « Elevage Demain" qui regroupe les résultats d'un inventaire réalisé en 2012 par le Groupement d'Intérêt Scientifique sur l'Elevage en utilisant leur propre protocole.

La diversité des sources expliquent la possible hétérogénéité dans la manière dont les bases sont décrites.

Propriétaire(s) de la base	Personne enquêtée	Structure - Fonction de la personne enquêtée	Nom de la base	Description, objectifs de la base	Source des informations
ACTA, le réseau des instituts techniques agricoles ;	Baudet Alice	ACTA	WebPhyto ACTA	Données à l'origine des index phytosanitaires ACTA paraissant chaque année : informations sur les caractéristiques des substances actives, des produits phytosanitaires, les conditions d'usages des produits.	Groupe «Connaissance des pratiques » du GIS GC HP2E
ACTOVI (et IDELE)			AT-Ov : Appui Technique Ovin viande	Données sur les appuis techniques individuels et pouvant faire l'objet d'analyses retournées aux techniciens. En 2011, 7 371 éleveurs sont suivis par le dispositif d'Appui Technique Individuel.	GIS Elevage Demain

Agence Bio	Guhl Florent	Directeur de l'Agence Bio	Base nationale des opérateurs de l'agriculture biologique	Données de notification et de certification des opérateurs de la bio en France	Inventaire MAAF
Agence de l'Eau Artois-Picardie	Thépaut Nolwenn	Agence de l'Eau Artois-Picardie	Observatoire des pratiques agricoles du bassin Artois-Picardie	Suivi de l'évolution des pratiques agricoles, en lien avec la ressource en eau dans le bassin Artois-Picardie : collecte annuelle de données (fertilisation et interculture) auprès d'un échantillon représentatif d'exploitations agricoles en GC, Polyculture-Elevage, Maraîchage.	Groupe «Connaissance des pratiques » du GIS GC HP2E
Agence de l'Eau Artois-Picardie Collectivités	Thépaut Nolwenn	Agence de l'Eau Artois-Picardie	Observatoire ORQUE : Observatoire pour les Opérations de la Reconquête de la Qualité de l'Eau	Enquête annuelle sur les pratiques agricoles d'exploitants du bassin Artois-Picardie. Données sur la fertilisation, l'interculture, le travail du sol, la protection des cultures et sur l'état environnant des parcelles.	Groupe «Connaissance des pratiques » du GIS GC HP2E
Agence de l'Eau Seine-Normandie	Ratiarson Jérôme	Agence de l'Eau Seine-Normandie	Observatoire des reliquats azotés Seine-Normandie	Description des pratiques des agriculteurs sur des parcelles de référence, de façon continue sur une période pluriannuelle, plutôt orientée fertilisation et pratiques phytosanitaires. Description globale des Systèmes de culture sur l'ensemble de l'exploitation (diagnostics agricoles au niveau des AAC du bassin Seine Aval)	Groupe «Connaissance des pratiques » du GIS GC HP2E
Agence Européenne de l'Environnement SOeS			CLC : Corine Land Cover	Occupation biophysique des sols, échelle européenne, mise à jour tous les 5 à 10 ans, dans le cadre du programme européen d'observation de la terre Copernicus.	Groupe «Connaissance des pratiques » du GIS GC HP2E
Agriculteur APCA	Masseran André	APCA	Mes p@rcelles	Données d'exploitation agricole : pratiques, déclarations PAC, localisation des parcelles, plan d'épandage... Aujourd'hui, plus de 30 000 abonnés en France mais tous n'utilisent pas l'intégralité des fonctionnalités de l'outil.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Agriscope	Fernandez Guillaume	Président	Données des stations météo	23 Millions de données brutes 12000 capteurs	Entretien Irstea

ANSES	Vigneron Julien	Chef de projet Internet	Ephy : Catalogue des produits phyto-pharmaceutiques et de leurs usages, des matières fertilisantes et des supports de culture autorisés en France	Informations sur les produits (produits phytopharmaceutiques, matières fertilisantes et supports de cultures, adjuvants, produits mixtes et mélanges) couverts par une Autorisation de Mise sur le Marché (AMM).	Entretien Irstea
ANSES ANMV	Le Letty Laetitia	Chef de l'unité enregistrement (dpt autorisation de mise sur le marché)	iRCP : Index des médicaments vétérinaires autorisés en France	Données sur les médicaments vétérinaires autorisés en France : résumés des caractéristiques du produit (RCP) et de son utilisation.	Entretien Irstea
APCA	Pérès Jean-Luc		Matériels Agricoles	Référentiel sur les coûts de matériels agricoles en filières GC, viticulture, maraîchage et arboriculture, avec des hypothèses sur 3 niveaux d'utilisation différents : charges, amortissement, coût de réparation par ha.	Groupe «Connaissance des pratiques» du GIS GC HP2E
ARAA	Vandijk Paul		Agri-Mieux	Données sur les pratiques des agriculteurs sur un échantillon de parcelles, de façon discontinue (tous les 4-5ans), sur une période pluriannuelle, des systèmes de production.	Groupe «Connaissance des pratiques» du GIS GC HP2E
AREAS	Richet Jean-Baptiste		Pesticeros : Observatoire des usages et des transferts de produits phytosanitaires dans les eaux de ruissellement du BV érosif de Bourville	Suivi exhaustif des pratiques, notamment phytosanitaires, depuis 2007 sur le BV de Bourville, en GC et polyculture-Elevage pour identifier les conditions d'un fort risque de transfert de matière active par ruissellement en contexte érosif	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal	Dubois Baptiste	Arvalis - Institut du végétal SAE	BASE MEQ	Caractéristiques du matériel agricole (prix du matériel, débit de chantier, consommation en carburant en fonction du sol, de la parcelle).	Groupe «Connaissance des pratiques» du GIS GC HP2E

Arvalis - Institut du végétal	Siné Mehdi	Arvalis - Institut du végétal SSIM	BASE SOL	Fiches descriptives de sols agronomiquement importants, rattachés à des petites régions agricoles, au niveau national. Caractéristiques du sol décrites pour chaque sol-type : texture, calcaire, profondeur, hydromorphie, CEC, pH...	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal	Verdier Jean-Luc	Arvalis - Institut du végétal Actions Régionales	ECOHERMIP Grandes Cultures	Données expérimentales au niveau Systèmes de culture du réseau ECOHERMIP GC qui comprend 5 sites expérimentaux.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal	Deudon Olivier	Arvalis - Institut du végétal SSIM	EOLE	Données météorologiques de station météo d'Arvalis et de l'INRA (journalières ou hebdomadaires).	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal	Berrodier Marc	Arvalis - Institut du végétal SAEF	Fermothèque GC	Données structurelles de pratiques et de performances de 30 fermes-types réparties sur les grandes zones de production céréalière française. Deux fermes types complémentaires pour représenter le même système de production : une ferme-type moyenne (MAJ tous les 3 ans) et une ferme-type annuelle (MAJ annuelle), sur les cultures suivantes : BT; BD, orge; pomme de terre; maïs.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal	Lauga Bruno	Arvalis - Institut du végétal SSIM	FERTICOM	Informations légales et de composition de nombreux produits fertilisants et amendements sur GC.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal	Toqué Clotilde	Arvalis - Institut du végétal SAEF	Micro-fermes de Boigneville	Essais de Systèmes de culture menés sur la station expérimentale d'Arvalis à Boigneville (91) de 1990 à 2015 ; données de pratiques sur chaque parcelle et données d'observations. Cinq micro-fermes: production intégrée, production monoculture de blé, production raisonnée, augmentation de la productivité du travail, Agriculture Biologique (en place depuis 2008).	Groupe «Connaissance des pratiques» du GIS GC HP2E

Arvalis - Institut du végétal	Lauga Bruno	Arvalis - Institut du végétal SSIM	PHYTCOM	Référentiel des caractéristiques AMM (composition, doses, usages, toxicité, efficacité,...) sur les fongicides, les herbicides, les ravageurs, les traitements de semences, les régulateurs de croissance et les adjuvants. Preprend en partie les données du référentiel LEXAGRI.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal	Le Souder Christine	Arvalis - Institut du végétal SAEE	Pratiques de fertilisation azotée du blé tendre	Enquêtes Arvalis annuelles sur les pratiques de fertilisation azotée (et, depuis 2009, soufrée) sur BT avec 2 angles : pratiques et stratégie de raisonnement de la fertilisation.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal	Lauga Bruno	Arvalis - Institut du végétal SSIM	VARCOM	Caractéristiques des variétés de céréales à paille, pommes de terre, maïs et lin expérimentées à ARVALIS : identité de la variété, qualité technologique, note de résistance et agronomie. Mise à jour confiée aux experts métiers du pôle Variétés du SG3P	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal FranceAgriMer	Streiff Adeline	Arvalis - Institut du végétal SQV	Enquête variétale au champ sur Blé	Enquêtes annuelles auprès d'agriculteurs sur la qualité, sanitaire et technologique de la récolte nationale de BT et BD, depuis 2010 : données de pratiques et analyse de la qualité technologique de la récolte. Environ 500 parcelles (BT et BD) enquêtées chaque année	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal ITB Terres Inovia			Vigicultures®	Epidémiologie-surveillance des GC sur le territoire français : données d'observations de bioagresseurs (maladies et les ravageurs et depuis peu les adventices) et des pratiques sur la parcelle suivie, pour rédiger les Bulletins de Santé du Végétal (BSV). Depuis 2008. L'outil est utilisé au niveau de 16 régions françaises (anciennes régions avant 2016).	Groupe «Connaissance des pratiques» du GIS GC HP2E

Arvalis - Institut du végétal Terres Innovia	L'Herbier Crystel	Arvalis - Institut du végétal Actions Régionales	SYPPRE, volet Observatoire : SYstèmes de Production Productifs et Respectueux de l'Environnement , volet Observatoire	Enquête annuelle sur les pratiques et la structure des exploitations : données sur l'ITINÉRAIRE TECHNIQUE du BT, colza et orge d'hiver, sur les stratégies de raisonnement des pratiques et les caractéristiques structurelles de l'exploitation. Développé depuis 2014 par les ITA GC afin de suivre de façon pérenne l'évolution des pratiques agricoles et de leurs performances.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Arvalis - Institut du végétal UIPP	Marks Perreau Jonathan	Arvalis - Institut du végétal SAEE	Données PCQE Fontaine du Theil	Données très variées (hydrologie, analyses chimiques des eaux, audits d'exploitations, pratiques culturales) au niveau du BV de la Fontaine du Theil dans le cadre de son suivi entre 1998 et 2006 par Arvalis et l'UIPP. L'ITINÉRAIRE TECHNIQUE de chacune des parcelles du BV a été répertorié (cultures, date de semis, interventions phytosanitaires avec produit, dose, date, interventions mécaniques, fertilisation, rendement).	Groupe «Connaissance des pratiques» du GIS GC HP2E
ASP	Gendre Cédric	INRA US ODR	Données ASP de paiements sur le PDRH	Données sur les engagements et paiements par dispositif de la PAC sur le PDRH, réalisés par l'ASP entre 2007-2013.	Groupe «Connaissance des pratiques» du GIS GC HP2E
ASP	Gendre Cédric	INRA US ODR	Données ASP de paiements sur le PRDN	Données sur les engagements et paiements par dispositif de la PAC sur le PRDN, réalisés par l'ASP entre 2000-2006.	Groupe «Connaissance des pratiques» du GIS GC HP2E
ASP MAAF	Guilain Stéphanie	ASP	RPG : Registre Parcelaire Graphique	Système de déclaration graphique annuelle des surfaces agricoles dans le cadre des aides PAC.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Axérial	Bouchie Jean-Michel	Axérial Service Agronomie	EXPE Système GC Berry	Données expérimentales sur 3 Systèmes de culture du réseau EXPE Système GC Berry (1 site expérimental)	Groupe «Connaissance des pratiques» du GIS GC HP2E
BASF	Guillaume Vigneron	BASF, Chargé de mission marketing	Atlas Désherbage : risques résistance aux herbicides	Données de risques de résistance aux herbicides. Modèle de calcul basé sur les pratiques agriculteurs et ses traitements herbicides : délivre un niveau de risque de résistance et des leviers d'action agro et chimiques. Tout le territoire français.	Entretien Irstea

BASF	Guillaume Vigneron	BASF, Chargé de mission marketing	Atlas MB : risques maladies blé et orge	Données de risques maladies sur le blé tendre, blé dur et orge. Modèle de calcul de risque agronomique en morte saison et risque maladies en saison.	Entretien Irstea
BASF	Guillaume Vigneron	BASF, Chargé de mission marketing	Observ'Online	Données d'observations des clients dans différentes régions : maladies et ravageurs.	Entretien Irstea
BASF	Guillaume Vigneron	BASF, Chargé de mission marketing	Référentiel phyto	Données produits phytosanitaires / voir Phytodata pour avoir l'ensemble des firmes.	Entretien Irstea
BD PORCS			BD PORCS : Système d'information et de traçabilité chez les porcins	Données : - d'identification des porcins ; - de surveillance sanitaire.	GIS Elevage Demain
Cap2020	Denis Boigontier	Gérant Cap2020	Données de prévision	Données de prévision météo à l'échelle horaire à 6-7jours avec un maillage spatial de 3kmx3km	Entretien Irstea
Cap2020	Denis Boigontier	Gérant Cap2020	Données des stations météo	Données météo (rayonnement global, pluie / données horaires, température, humidité relative, vent, texture du sol) fournies par l'agriculteur	Entretien Irstea
Centre National de la Propriété Forestière	Thomas Formery	CNPF, Directeur Général	ILEX : Référentiel des dispositifs expérimentaux installés et suivis par le CNPF	Données de traçabilité des dispositifs expérimentaux (essais, placettes de références...) installés et suivis par le CNPF (plus de 6 000 références au 31/12/2015)	Inventaire MAAF
Centre National de la Propriété Forestière	Thomas Formery	CNPF, Directeur Général	MERLIN : Statistiques annuelles sur les documents de gestion forestières durables	Données de traçabilité des propriétés forestières (suivi en nombre et en surface) concernées par les documents de gestion forestière durable (DGD) : plans simples de gestion (PSG) et codes de bonnes pratiques sylvicoles (CBPS)	Inventaire MAAF
CER France	Varchavsky Marc	CER France	Observatoire coûts de production de Grandes Cultures	Coûts de production de GC basé sur des données comptables. Créé en 1990.	Groupe «Connaissance des pratiques» du GIS GC HP2E

CER Meuse			Données CER Meuse	Données de comptabilité analytique d'exploitations agricoles de la Meuse.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre d'Agriculture de la Drôme	Castel Laurie	Chambre d'Agriculture Drôme	VERTICAL	Données expérimentales de Systèmes de culture de 2 sites expérimentaux en AB.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre d'Agriculture de Vendée	Rattier Thierry		Service Auxil 2	Relevés biologiques sur certains ravageurs (limaces, pucerons des épis du blé) et auxiliaires du maïs et du blé (carabes, syrphes et chrysopes au stade larvaire et parasitoïdes), données de pratiques des parcelles enquêtées et description du paysage environnant la parcelle. Projet débuté à la campagne 2014 dans 5 départements (Vendée, Loire-Atlantique, Sarthe, Mayenne et Maine-et-Loire).	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre Régionale d'Agriculture Aquitaine-Limousin-Poitou-Charentes	Minette Sébastien	Chambre Régionale d'Agriculture Centre-Val de Loire	Observatoire des pratiques et des systèmes en Poitou-Charentes	Systèmes de culture de la région Poitou-Charentes, permettant notamment de constituer des références adossées au RMT Systèmes de culture.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre Régionale d'Agriculture d'Alsace-Champagne-Ardenne-Lorraine	Schaub Anne	ARAA	INNOViPEST	Données expérimentales des Systèmes de culture sur 8 sites dont 6 chez des producteurs, 1 en station expérimentale et 1 en établissement agricole.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre Régionale d'Agriculture de Bourgogne-Franche-Comté	Petit Marie-Sophie		Réseau PI	Données expérimentales des 9 Systèmes de culture sur 6 sites expérimentaux.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre Régionale d'Agriculture de Bretagne	Giteau Jean-Luc		SGC Bretagne	Données expérimentales des Systèmes de culture sur 2 sites expérimentaux.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre Régionale d'Agriculture de Lorraine	Leclech Nathaël		EXPE Ecophyto Lorrain	Données expérimentales des 7 Systèmes de culture sur 4 sites expérimentaux.	Groupe «Connaissance des pratiques» du GIS GC HP2E

Chambre Régionale d'Agriculture des Hauts-de-France	Halama Patrice	ISA Lille Agriculture / Laboratoire BioGAP	DEPHY EXPE NPDC	Données expérimentales des Systèmes de culture sur 2 sites expérimentaux (Tillay et Lorgies).	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre Régionale d'Agriculture des Hauts-de-France	Tosser Véronique	Arvalis - Institut du végétal SAE	Entomophages	Données de suivi d'auxiliaires et de pratiques agricoles en GC au niveau de 3 régions (Picardie, Rhône-Alpes et Centre) pendant 3 ans.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre Régionale d'Agriculture des Hauts-de-France	Preud'homme Rose-Line	MNHN	Auximore	Données de suivi d'auxiliaires des cultures et de pratiques (travail du sol, nombre de traitements par catégorie) sur des parcelles de GC (et autres) sur 15 départements à travers toute la France pendant 2 ans.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambre Régionale d'Agriculture du Centre-Val de Loire ; CRA des Hauts-de-France CRA d'Auvergne-Rhône-Alpes ; CRA Nouvelle Aquitaine	Agenis-Nevers Aude	APCA	INOSYS GC : INNOvation SYStèmes Grandes Cultures	Observatoire des systèmes agricoles performants en GC : 140 fermes en GC qui sont suivies annuellement au niveau national. Édite des cas-types pour aboutir à des référentiels de performances techniques et économiques calculées grâce aux données de suivi de fermes et à l'expertise métier.	Groupe «Connaissance des pratiques» du GIS GC HP2E
Chambres d'agriculture du Grand Ouest			Enq-Avi : Enquête avicole des Chambres d'Agriculture de l'Ouest	Données techniques et économiques sur les élevages avicoles.	GIS Elevage Demain
CLAAS	Louis Jean-Noël	Directeur marketing CLAAS	Utilisation du matériel	Données des capteurs embarqués sur tracteurs.	Entretien Irstea
CNIEL - Centre National Interprofessionnel de l'Economie Laitière			INFOLABO	Résultats d'analyses de lait (bovins, ovins, caprins) pour les laboratoires, les laiteries et les producteurs.	GIS Elevage Demain
CNRS ITSAP	Bretagnolle Vincent	CNRS	Données Pollinov-Risq'API	Données de suivi des pratiques agricoles sur la zone atelier de Chizé, (notamment projets Pollinov et DEPHY Abeilles) sur des systèmes en GC et Polyculture-Elevage: travail du sol, fertilisation, protection des cultures, semis.	Groupe «Connaissance des pratiques» du GIS GC HP2E

DDT(M)	Poupeau Philippe		SYLVA : Système de Gestion administrative des dossiers forestiers (arrêté du 27/02/2001)	Données cadastrales des propriétés forestières et les régimes d'autorisation ou droits afférents.	Inventaire MAAF
DGAL	Rémy Sébastien	Chef du bureau d'appui à la surveillance de la chaîne alimentaire	Publication des résultats des contrôles officiels en sécurité sanitaire des aliments	Résultats des contrôles officiels en sécurité sanitaire des aliments : contrôles réalisés par les agents du MAAF et bientôt contrôles des autres administrations.	Inventaire MAAF
DGAL	Commeau Natalie	Chef du bureau de l'évaluation scientifique, de la recherche et des laboratoires	Laboratoire agréés pour les analyses visées au livre II du CRPM	Listes des laboratoires agréés pour une analyse donnée.	Inventaire MAAF
DGAL / SDPRAT / BMOSIA -	Pascal Deriu Frédéric Masson	Adjoint au chef de bureau Directeur de projet USAGERS	RESYTAL - USAGERS	Gestion des USAGERS (entreprises et établissements) et des activités détaillées qu'ils exercent, au regard de la réglementation liée aux contrôles sanitaires dans les domaines de la Santé et protection animale, sécurité sanitaire des aliments et santé qualité et protection des végétaux. Volumétrie : env. 2 millions d'usagers.	Inventaire MAAF
Ekylibre	Sebastien Hemono	Responsable produit	Itinéraire technique et comptabilité	Données agriculteurs : de la comptabilité jusqu'à l'itinéraire technique	Entretien Irstea
FDSEA51- Conseil CDER	Carpentier Alain	INRA SMART	Données comptables de centres de gestion de la Marne	Données de comptabilité analytique d'exploitations agricoles de la Marne (en cours de construction).	Groupe «Connaissance des pratiques» du GIS GC HP2E
FNCUMA	Stéphane Chapuis	Responsable BDD	Diagnostic de performance des tracteurs	Données de diagnostic sur la performance environnementale des tracteurs (10 000 tests).	Entretien Irstea
FNCUMA	Stéphane Chapuis	Responsable BDD	Fiabilité du matériel	Données qualitatives sur la fiabilité des matériels (issues d'une enquête auprès de 200 CUMA et agriculteurs).	Entretien Irstea

FNCUMA	Stéphane Chapuis	Responsable BDD	Matériel des exploitations	Données sur les adhérents, le matériel, la comptabilité liée au matériel, des données analytiques.	Entretien Irstea
France Conseil Elevage et France Bovins Croissance			CPC-B-C : Bases contrôle de performances et de conseil bovins et caprins	Données du contrôle de performances des bovins et caprins.	GIS Elevage Demain
FranceAgriMer /MEP/BIE –	Meizels Monique – Seyer Philippe	Chef du service des bases d'information économique Chef de l'unité structuration de données	Cotations 2 : Cotations animales	Données de volumes, de prix, d'activité d'abattages hebdomadaires des principaux abatteurs en France (veaux, bovins, ovins). Données de volumes, de prix, d'animaux commercialisés sur un panel d'environ 30 marchés aux bestiaux. Couverture : France métropolitaine Taille de la base de données individuelles : 6 Go.	Inventaire MAAF
FranceAgriMer	Meizels Monique – Seyer Philippe	Chef du service des bases d'information économique Chef de l'unité structuration de données	Etats 2 : VISIOGrains (BdD) et VISIOBase (entrepôt de données décisionnel) NB2 : ce formulaire concerne les données françaises sur la collecte de céréales, d'oléagineux et de protéagineux ainsi que celles sur les stocks de ces produits chez les collecteurs.	Données de recensement (tous les collectés enquêtés) sur les quantités de grains de céréales et de graines d'oléagineux et de protéagineux collectées par les collecteurs déclarés auprès de FranceAgriMer , et les stocks de ces produits chez ces collecteurs. Couverture nationale. Depuis juillet 1998. Collecte : 1 500 000 enregistrements Stocks : 1 800 000 enregistrements VISIOGrains – Etats 2 : 2,5 Go VISIOBase - Etats 2 : 7 Go (contient les historiques)	Inventaire MAAF
FranceAgriMer	Meizels Monique – Seyer Philippe		Coûts de production blé tendre	Enquêtes annuelles entre 1995 et 2009 sur les ITINÉRAIRES TECHNIQUES de 100 parcelles dans 5 départements du grand Bassin Parisien.	Groupe «Connaissance des pratiques» du GIS GC HP2E

Fruition	Sébastien Payen	CEO Fruition	Données météo et irrigation vigne	Données météo depuis 2006. Données vigne uniquement (flux de sève et analyses labo sur baies) en Bordeaux, LR, Alsace, Champagne. Données ponctuelles	Entretien Irstea
GEOSUD et pôle Theia	Pierre Maurel	Irstea, Coordinateur projet GEOSUD + membre bureau exécutif pôle national Theia	Plateforme imagerie de la Terre – GeoSud-Theia	Images satellites très haute résolution spatiale (50 cm à 6 m) : couvertures nationales annuelles depuis 2010, acquisitions ad-hoc monde entier depuis 2015. Licence tout acteur public français + acteur privé pour R&D après accord préalable. Archives SPOT 1-5 1986-2014. Images Sentinel 2 prétraitées depuis 2016. Produits bio-physiques dérivés des images.	Entretien Irstea
Geosys	Chloé Cochais	Geosys, responsable produit	Imagerie basse résolution	Images observation de la terre en basse résolution 30 – 50 m public USGS + 5 - 30m.	Entretien Irstea
Geosys	Chloé Cochais	Geosys, responsable produit	Itinéraire technique	Données d'itinéraire technique + carte de rendement (fournies par l'agriculteur) + référentiel variété.	Entretien Irstea
Geosys	Chloé Cochais	Geosys, resp produit	Météo	Météo.	Entretien Irstea
IFIP			GTE-P : Gestion Technico-Economique	Données techniques et économiques en élevage porcin : coût alimentaire et du renouvellement.	GIS Elevage Demain
IFIP			GTE-TB : Gestion Technico-Economique – Tableau de Bord	Données pour prolonger GTE en gérant les autres postes du coût de production.	GIS Elevage Demain
IFIP			GTTT : Gestion Technique des Troupeaux de Truies	Données de performances de reproduction de l'ensemble de l'élevage et de chaque animal reproducteur (truies et verrats) : résultats de performances de reproduction globaux et critères permettant de les analyser. Méthode standardisée permettant la comparaison entre élevages. Couverture : tous les élevages détenteurs de truies, naisseurs et naisseurs-engraisseurs.	GIS Elevage Demain

IGN /Direction des programmes civils	Elisabeth Leblanc – Bruno Goupil	Service d'Appui aux Politiques Publiques, Chargée de relation avec les ministères Service Aménagement du Territoire, Directeur de programme	RGE® ALTI, BD ORTHO®, BD PARCELLAIRE®, SCANS 25®	RGE® ALTI http://professionnel.ign.fr/sites/default/files/D_L_RGE-ALTI_2-0.pdf BD ORTHO® http://professionnel.ign.fr/sites/default/files/DL_raster_ORTHO.pdf BD PARCELLAIRE®: http://professionnel.ign.fr/sites/default/files/DL_BDPARCELLAIRE_1_2.pdf SCANS 25®: http://professionnel.ign.fr/sites/default/files/DL_raster_SCAN.pdf	Entretien Irstea
INERIS			SIRIS Pesticides 2012	Caractéristiques chimiques et données toxicologiques de substances actives.	Groupe «Connaissance des pratiques» du GIS GC HP2E
INRA	Bertrand Christine – Milan Denis - INRA -	Directrice du CTIG (centre de traitement de l'information génétique) – Chef de Département Génétique Animale,	SNIG : Systèmes nationaux d'information génétique des espèces bovine, ovine, caprine et porcine. SIG (bovins), SIEOL et OVALL (ovins), SIECL (caprins) et BANAPOG (porcins).	SNIG : Ensemble des données zootechniques et de généalogie gérées en application de l'arrêté du 24 mars 2015, pour l'évaluation génétique officielle, la recherche et le suivi de l'évolution des races et populations animales sélectionnées des espèces bovine, caprine, ovine et porcine.	Entretien Irstea
INRA	Cellier Vincent		RésOPest	Données expérimentales des 8 Systèmes de culture du réseau PI sur 8 sites expérimentaux.	Groupe «Connaissance des pratiques » du GIS GC HP2E
INRA -CNRS			Réseau SEBIOPAG : Réseau de sites d'observation à long terme des Services Ecosystémiques assurés par la Biodiversité dans les Paysages Agricoles	Données de suivis de pratiques sur 20 parcelles au niveau de 5 sites (dont 4 en GC et 1 en verger) en France sur 3 ans, en parallèle d'observations de prédateurs des cultures, de flore et de carabes (2 fois par an en 3 séries pour chaque session) sur les mêmes parcelles.	Groupe «Connaissance des pratiques» du GIS GC HP2E

INRA SMART	Cariou Sylvain		Séries nationales de prix intrants	Données de prix sur l'énergie, les carburants, l'engrais et les aliments de 1990 à 2009.	Groupe «Connaissance des pratiques» du GIS GC HP2E
INRA SAD Paysage	Joannon Alexandre		Suivi des pratiques agricoles sur la Zone Atelier Armorique	Données de suivi sur 4 années des pratiques agricoles en GC et Polyculture-élevage pour environ 240 parcelles de cultures (6 exploitations les 2 premières années et 5 les 2 dernières) dans la Plaine de Fougères.	Groupe «Connaissance des pratiques» du GIS GC HP2E
INRA UMR Agroécologie	Munier-Jolain Nicolas	INRA	System-Eco-Puissance4	Données expérimentales des 19 Systèmes de culture sur 4 sites expérimentaux.	Groupe «Connaissance des pratiques» du GIS GC HP2E
INRA Agence de l'Eau Seine-Normandie	Schott Céline	INRA SAD-ASTER	APOCA : Agricultural Practices of the Orgeval Catchement Area	Données sur les pratiques d'une vingtaine d'agriculteurs de GC du BV de l'Orgeval entre 1990 et 2011. Donnant naissance à la description de l'assolement parcellaire par exploitation de 1990 à 2011, l'estimation de l'assolement du BV de 1990 à 2011 et les ITINÉRAIRES TECHNIQUES moyens par culture et par an.	Groupe «Connaissance des pratiques» du GIS GC HP2E
INRA Agence de l'Eau Seine-Normandie	Puech Thomas	INRA SAD-ASTER	BdD ARSEINE : Base de Données Agricole Régionalisée sur le bassin SEIne-NormandiE	Données de pratiques représentatives (successions et ITINÉRAIRES TECHNIQUES) de régions agricoles homogènes en GC, polyculture-élevage et viticulture.	Groupe «Connaissance des pratiques » du GIS GC HP2E
INRA-URH (Unité de Recherche sur les Herbivores), Equipe EGEE (Economie et Gestion de l'Exploitation d'Elevage)			R-INRA-BO : Suivi technico-économique Bovins Viande et Ovins Viande	Données Bovins Viande : démarrage en 1971 avec 50 exploitations (Nièvre), intégration de 18 (Creuse) en 1980... Aujourd'hui : 78 exploitations dont 10 en Agriculture Biologique et 1 exploitation de lycée agricole. Données Ovins Viande : démarrage en 1987 avec 50 exploitations (Montmorillonnais, Plateau de Millevaches, Margeride et Dôme). En 2010, 48 exploitations dont 12 en Agriculture Biologique.	GIS Elevage Demain

INSEE			IPAMPA : Indice des prix d'achat des moyens de production agricole	Données sur les variations de prix des intrants (base 2010) pour: Produits de protection des cultures, Aliments des animaux, Matériel et petit outillage, Entretien et réparation, Produits et services vétérinaires, Semences et plants, Energie et lubrifiants, Engrais et amendements.	Groupe « Connaissance des pratiques » du GIS GC HP2E
INSEE			IPPAP : Indice des prix des produits agricoles à la production	Données sur les variations des prix de production agricoles.	Groupe « Connaissance des pratiques » du GIS GC HP2E
Institut de l'élevage			RECP : Réseaux d'Élevage pour le Conseil et la Prospective	Données sur le fonctionnement technique et économique des principaux systèmes d'élevage en France et de leurs évolutions : 1420 fermes suivies en 2010 dans le cadre du socle national, complétées de 500 fermes suivies dans le cadre régional.	GIS Elevage Demain
Institut français du cheval et de l'équitation	de Cadolle Hubert	Adjoint au directeur général chargé du pilotage stratégique du système d'information	SIRE (système d'information relatif aux équidés).	Données relatives à la propriété, la détention et à l'identification des équidés nés ou détenus en France, et données sanitaires et zotechniques relatives à ces équidés : Nb de chevaux : 3 millions Nb de personnes : 1 million 1/ Base de données centrale (Règlement d'exécution (UE) 2015/262 de la Commission du 17 février 2015) 2/ Fichier national des équidés (code rural L212-9) 3/ Fichier central zotechnique des équidés (code rural D212-46)	Inventaire MAAF
Invivo Agrosolutions	Guillaume Py	Data scientist	FermesEcophyto	Données de pratiques et parcellaires des agriculteurs engagés dans le réseau FERMEcophyto, indicateurs d'impact environnemental et de performance économique, des récoltes 2011 à 2015. En tout 2900 fermes dans le réseau, 150 à 320 fermes sont tracées chaque année.	Entretien Irstea

INVS ACTA Université Claude Bernard, Lyon 1			CIPA : Compilations des Index Phytoprotecteurs ACTA	Ensemble des données relatives aux substances actives et à leurs usages depuis 1961 et disponibles au sein des index phytoprotecteurs ACTA.	Groupe « Connaissance des pratiques » du GIS GC HP2E
Irstea	Vernier Françoise	Irstea	Enquête Bassin versant du Né	Données d'enquête (2014) auprès d'agriculteurs et d'experts agricoles : description des pratiques agricoles dans le BV du Né (viticulture+ GC sur le BV) en Charente. Données sur leur perception des évolutions de leurs pratiques.	Groupe « Connaissance des pratiques » du GIS GC HP2E
Isagri	Alexandre Diaz	Isagri, responsable Innovation	Gestion globale de l'exploitation	Données de gestion de l'exploitation. 110 000 licences.	Entretien Irstea
ITAVI	Guillaume Coutelet		CUNIMIEUX : Réseau de fermes de références cunicoles	Données technico économiques sur les élevages cunicoles.	GIS Elevage Demain
ITAVI			Ob-Avi : Observatoires économiques avicoles – suivi des coûts de production	Données économiques sur les marchés avicoles.	GIS Elevage Demain
ITAVI	Guillaume Coutelet		RENACEB – RENALAP : Gestion Technico- Economique cunicole	Références nationales techniques et économiques annuelles sur la production cunicole.	GIS Elevage Demain
ITAVI			RENAPALM : Gestion technico- économique des éleveurs et gaveurs de palmipèdes à foie gras	Données sur les bandes d'élevage et de gavage en filière longue, sur les performances zootechniques et les résultats économiques.	GIS Elevage Demain
ITAVI			Res-PALM : Réseau de fermes de références palmipèdes à foie gras	Données de caractérisation des systèmes de production d'ateliers représentatifs de la production nationale (main d'œuvre, SAU, production annexe,...).	GIS Elevage Demain

ITB	Gouwie Céline		Enquête SITE (Suivi des Itinéraires Techniques)	Enquêtes annuelles sur les pratiques culturales en betterave; de la récolte du précédent à la récolte de betterave. Depuis 1997.	Groupe « Connaissance des pratiques » du GIS GC HP2E
ITB	Gouwie Céline		Amendements ITB	Caractéristiques générales des amendements utilisés par les betteraviers enquêtés dans le cadre de l'enquête SITE : fournisseur, type d'amendement, composition,... Données d'expérimentation.	Groupe « Connaissance des pratiques » du GIS GC HP2E
ITB	Gouwie Céline		Machinisme ITB	Liste de marques, modèles et caractéristiques techniques (taille, puissance,...) de matériels agricoles utilisés par les betteraviers enquêtés dans le cadre de l'enquête SITE de l'ITB.	Groupe « Connaissance des pratiques » du GIS GC HP2E
ITB	Gouwie Céline		Produits phytosanitaires ITB	Caractéristiques (n°AMM, composition, fournisseur, usage, ...) des produits phytosanitaires utilisés par les betteraviers enquêtés dans le cadre de l'enquête SITE.	Groupe « Connaissance des pratiques » du GIS GC HP2E
ITB	Gouwie Céline		Variétés Betterave	Données descriptives sur toutes les variétés testées par l'ITB depuis une vingtaine d'années(classification d'inscription, distributeur, date d'inscription, lieux d'essai, évolution des classements dans le temps,code GEVES, obtenteur, prix standard,...).	Groupe « Connaissance des pratiques » du GIS GC HP2E
ITB IRSTEA FNCUMA FNEDT	Blocaille Suzanne	ITB Service Agro- équipement et conservation	Perfbet	Caractéristiques techniques des chantiers de récolte de betteraves classées dans différentes catégories : pneumatique, effeuillage, nettoyage... Plus d'une vingtaine de chantiers de récolte référencés.	Groupe « Connaissance des pratiques » du GIS GC HP2E
ITK	Philippe Stoop	ITK, directeur R&D	Assolement	Données sur les cultures en place sur la parcelle.	Entretien Irstea
ITK	Philippe Stoop	ITK, directeur R&D	Données de sol	- texture et profondeur de sol fournis par agriculteur - carte de sol USDA	Entretien Irstea
ITK	Philippe Stoop	ITK, directeur R&D	Etat hydrique et sanitaire	- déficit hydrique - état sanitaire (mildiou, oïdium, blé, septo, rouille) - potentiel de rendement - scénarios prospectifs	Entretien Irstea

ITK	Philippe Stoop	ITK, directeur R&D	Météo	Données météo.	Entretien Irstea
Lasalle Beauvais	Grandgirard David	Lasalle Beauvais	SCA 0 pest	Données expérimentales des Systèmes de culture sur 1 site (GC combinées à l'agroforesterie).	Groupe « Connaissance des pratiques » du GIS GC HP2E
MAAF			CALAMNAT : Outil d'instruction des demandes d'indemnisation au titre des calamités agricoles	Données nécessaires aux calculs des indemnisations et résultat de ces calculs. Lots de paiement.	Inventaire MAAF
MAAF	Eveillard Philippe	UNIFA	Livraisons d'engrais et d'amendements minéraux basiques	Statistique des livraisons correspondant aux déclarations faites à l'UNIFA par l'ensemble des producteurs et des importateurs de fertilisants minéraux et organo-minéraux.	Groupe « Connaissance des pratiques » du GIS GC HP2E
MAAF INRA	Fried Guillaume	ANSES Laboratoire de la santé des végétaux, Unité d'Entomologie et Plantes invasives	Réseau Biovigilance Flore	Données de suivi des pratiques agricoles (l'ITINÉRAIRE TECHNIQUE, la topographie, le sol, la présence de haies et les cultures jouxtant la parcelle) et d'observation de la flore adventice au niveau national sur environ 1000 parcelles de 2002 à 2010. Mis en place par le Service de la Protection des Végétaux et l'INRA aujourd'hui géré par l'ANSES. Accompagnée d'une BdD de calculs d'indicateurs (IFT, ...).	Groupe « Connaissance des pratiques » du GIS GC HP2E
MAAF/DGAL			BDNI : Système d'information de l'identification et de la traçabilité des bovins	Données d'identification des bovins et suivi des mouvements (traçabilité d'ensemble de la viande jusqu'au consommateur final).	GIS Elevage Demain
MAAF/DGAL			OVINFOS : Système d'information de l'identification et de la traçabilité des ovins et des caprins	Données d'identification des ovins et caprins et suivi des mouvements dans l'objectif de la surveillance sanitaire des cheptels face à un risque d'épizootie.	GIS Elevage Demain

MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	ESEA : Enquête sur la Structure des exploitations agricoles	Données individuelles sur un échantillon de 53 000 exploitations agricoles: cultures, cheptels, main-d'œuvre, activités de diversification, production sous signe de qualité, modes de production (irrigation, couvert), gestion (statut, mode de faire valoir).	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	ESEA-tableaux : Enquête sur la Structure des exploitations agricoles	Données régionales concernant les cultures, les cheptels, la main-d'œuvre, les activités de diversification, la production sous signe de qualité, les modes de production (irrigation, couvert), la gestion (statut, mode de faire valoir) des exploitations agricoles.	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	IVER : Inventaire des vergers - Enquête sur la structure des vergers	2013 : Enquête quasiment exhaustive sur les exploitations métropolitaines exploitant une surface >= un hectare en vergers de pommiers, pêchers, abricotiers, pruniers, kiwis, noyers, agrumes, ou d'au moins 50 ares en vergers de poiriers, cerisiers, raisin de table : données sur les structures, les superficies, les circuits de commercialisation. Enquête 2007 : enquête sur un échantillon de 6500 exploitations agricoles ayant des vergers de plus de 50 ares : structure des exploitations, variétés, pratiques culturales, commercialisation.	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	IVER-tableaux : Inventaire des vergers - Enquête sur la structure des vergers - Tableaux	Données tabulées par région issues de l'enquête IVER 2013 (Enquête quasiment exhaustive sur les exploitations métropolitaines >=1 hectare en vergers de pommiers, pêchers, abricotiers, pruniers, kiwis, noyers, agrumes ou > 50 ares en vergers de poiriers, cerisiers, raisin de table : données sur les structures, les superficies, les circuits de commercialisation). Enquêtes par sondage de 1997, 2002 et 2007 (échantillon d'environ 6500 exploitations agricoles)	Inventaire MAAF

MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	PKGC : Enquête sur les pratiques culturales en grandes cultures	Relevé des itinéraires techniques sur un échantillon de parcelles (Blé tendre 3000 ; Blé dur 900 ; Orge 2100 ; Triticale 2500 ; Colza 2100 ; Tournesol 1500 ; Pois protéagineux 1900 ; Maïs fourrage 2500 ; Maïs grain 2200 ; Betterave sucrière 800 ; Pomme de terre 900 ; Canne à sucre)	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	PKGC-tableaux 2011 : Enquête sur les pratiques culturales en grandes cultures – tableaux 2011	Données tabulées par extrapolation des données de l'enquête relevant les itinéraires techniques sur un échantillon de parcelles (Blé tendre 3000 ; Blé dur 900 ; Orge 2100 ; Triticale 2500 ; Colza 2100 ; Tournesol 1500 ; Pois protéagineux 1900 ; Maïs fourrage 2500 ; Maïs grain 2200 ; Betterave sucrière 800 ; Pomme de terre 900 ; canne à sucre)	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	PKViti : Enquête sur les pratiques culturales en viticulture	Enquête portant sur un échantillon d'environ 6200 parcelles : Itinéraire technique, fertilisation, pratiques phytosanitaires.	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	RA : Recensement agricole	Données individuelles sur les exploitations agricoles : cultures, cheptels, main-d'œuvre, activités de diversification, production sous signe de qualité, modes de production (irrigation, couvert), gestion (statut, mode de faire valoir).	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	RA-données chiffrées : Recensement agricole – données chiffrées	Données agrégées, par commune ou zonage plus large, sur les exploitations agricoles : cultures, cheptels, main-d'œuvre, activités de diversification, production sous signe de qualité, modes de production (irrigation, couvert), gestion (statut, mode de faire valoir).	Inventaire MAAF

MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	Rica : réseau d'information comptable agricole	Données de comptabilité d'un échantillon d'exploitations agricoles représentatif de l'ensemble des exploitations agricoles de France métropolitaine, Etendu aux Dom depuis 2013 (Guadeloupe, Martinique, La Réunion). Sont disponibles les micro-données et également des données de synthèse.	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	SAA : Statistique agricole annuelle	Séries longues historiques annuelles sur les productions en volume physique : végétales (surfaces, production, rendement) et animales (cheptels, production de viande, production de lait).	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	Teruti-Lucas : Enquête sur l'utilisation du territoire : TERritoire UTILisation Land Use/Cover Area frame Survey	Enquêtes annuelles associant photographies aériennes (base de sondage) et relevés de terrain effectués par des enquêteurs. A partir de 1982, enquête basée sur un échantillon national obligatoire constant. Occupation et utilisation du sol relevé sur 322 000 points en France métropolitaine, en Guadeloupe, à la Martinique, à La Réunion.	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	VVT : Valeur vénale des terres	Données d'estimation du prix des terres, prés et vignes par petites régions agricoles ou regroupement de petites régions agricoles.	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	Bilan d'approvisionnement	Séries longues historiques annuelles sur les bilans exprimés en quantités : équilibre des ressources (stocks de début, production utilisable, importations) et des emplois (exportations, utilisations intérieures, stocks finals) pour un produit donné : Céréales - Riz - Pomme de terre (bilan interrompu en 2011) - Protéagineux ; Sucres et produits sucrés ; Oléagineux - Huiles et graisses ; Viandes - Œufs - Lait et produits laitiers ; Légumes - Fruits ; Ressources fourragères.	Inventaire MAAF

MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	Bulletin mensuel de conjoncture	Séries longues historiques mensuelles depuis 1993.	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	Comptes régionaux de l'agriculture	Comptes de l'agriculture, en valeur (prix courants) et en volume (prix de l'année précédente).	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	Enquête sur la structure de la production légumière	Enquête par échantillon auprès de 13000 exploitations >= 50 ares (20 ares et plus dans les DOM) de légumes frais ou pour la transformation, en plein air et sous abri bas ou ayant un abri haut : Données de SAU, superficies légumières, UTA, modes de culture, surfaces en abris hauts ou serres, mode de succession ou d'association des cultures, commercialisation.	Inventaire MAAF
MAAF/SG/SSP	Giroux Gisèle	Cheffe du bureau de l'information statistique	Enquête sur la structure de la production légumière - tableaux	Tableaux obtenus après extrapolation des données de l'enquête par échantillon auprès de 13000 exploitations >= 50 ares et plus (20 ares et plus dans les DOM) de légumes frais ou pour la transformation, en plein air et sous abri bas ou ayant un abri haut consacré aux légumes frais : données de SAU, superficies légumières, UTA, modes de culture, surfaces en abris hauts ou serres, mode de succession ou d'association des cultures, commercialisation.	Inventaire MAAF
MAAF DGPE SGPAC SDGP SG/SM/SDSI/BS IP	Justin Willy	Chargé de projet informatique	LOGICS : L'outil de gestion et d'instruction des dossiers de contrôle des structures	Données relatives aux biens agricoles, objets des demandes d'autorisation d'exploiter (et des déclarations) effectuées via Logics (outil de télédéclaration des demandes). En cours de développement (généralisation de l'outil fin 2016).	Inventaire MAAF

MAAF / Sous-direction de la Performance environnementale et de la valorisation des territoires (SDPE)	Karine Belna	Chargée de mission stratégie environnementale dans les politiques agricoles ; Bureau du développement agricole et des chambres d'agriculture (BDA)	Suivi GIEE	Données de suivi des groupements d'intérêt économique et environnemental créés à l'issue de chaque appel à projet régional.	Inventaire MAAF
MNHN APCA	Preud'homme Rose-Line	MNHN	Observatoire agricole de la biodiversité	Données de suivi au niveau de parcelles en GC, cultures pérennes et maraîchage, de différentes faunes (abeilles, papillons, invertébrés du sol), données paysagères sur l'environnement proche de la parcelle (occupation du sol, mode de conduite la parcelle).	Groupe « Connaissance des pratiques » du GIS GC HP2E
Multi-propriétaire, chaque organisme rentrant les données dans la BdD	Munier-Jolain Nicolas	INRA UMR Agroécologie	DEPHY Ferme GC	Données de pratiques des Systèmes de culture d'exploitations engagées dans le réseau DEPHY en GC, collecte des informations dans l'outil Agrosyst développé par l'INRA.	Groupe « Connaissance des pratiques » du GIS GC HP2E
NORMABEV			NORMABEV	Données de classement et de marquage des gros bovins, vérification de la présentation des carcasses et de la pesée fiscale des gros bovins dans les entreprises d'abattage.	GIS Elevage Demain
ODEADOM	Bouhnik Norbert	RSI	Plusieurs bases de données par mesure d'aide : Banane, Canne, Végétal, Certificats d'imports	Données : coordonnées du bénéficiaire, montant de l'aide perçue, date du paiement, département concerné, produit imputé.	Inventaire MAAF
ONF			BDN : données d'observation naturalistes		Inventaire MAAF

ONF			Contour des forêts et des parcelles forestières des terrains relevant du régime forestier		Inventaire MAAF
ONF			Contour des réserves biologiques		Inventaire MAAF
Organisme Stockeur	Simonin Pascal	Terres Inovia	DDP Colza Diester : Données Démarche de Progrès Colza Diester	Données des enquêtes annuelles sur les pratiques culturales menées par les OS depuis la campagne 2007-2008 dans le cadre de la Démarche de Progrès (DDP) de la filière Colza Diester : type de sol, travail du sol, fertilisation, rendement, mode de récolte,...	Groupe « Connaissance des pratiques » du GIS GC HP2E
SIABAVEChambre d'Agriculture Marne	Négri Ophélie	SIABAVE	Observox : Observatoire orienté Xénobiotiques	Données de description des pratiques réelles sur toutes les parcelles des exploitations suivies en GC, viticultures et maraîchage, de façon continue sur une période pluriannuelle (questions orientées plus fertilisation et usages des produits phytosanitaires) au niveau du BV de la Vesle (Nord-Est du bassin de la Seine)	Groupe « Connaissance des pratiques » du GIS GC HP2E
SMAG	Anthony Clenet	SMAG, directeur R&D	Itinéraire technique	Données d'itinéraires techniques depuis 2006-2007. 60 000 agriculteurs.	Entretien Irstea
SMAG	Anthony Clenet	SMAG, directeur R&D	Référentiel phyto	Base phyto.	Entretien Irstea
SPIE (association)			SPIE : Système Professionnel d'Information en Elevage	Base de données nationales d'informations sur les bovins et les exploitations ; mise en place et gestion d'un service de mise à disposition de données auprès des partenaires des filières adhérents de l'association ; transmission de données réglementaires I.P.G. à la base de données nationale du ministère de l'agriculture ; toute action visant à permettre le recueil et la transmission de données de traçabilité et de valorisation des produits animaux.	GIS Elevage Demain

Terres Inovia	Cadoux Stéphane	Terres Inovia	PHYTO-SOL	Données expérimentales des 3 Systèmes de culture sur 3 sites expérimentaux.	Groupe « Connaissance des pratiques » du GIS GC HP2E
Terres Inovia	Lecomte Vincent	Terres Inovia	Pratiques culturales Chanvre	Données d'enquêtes annuelles sur les pratiques culturales en chanvre dans les principaux départements producteurs.	Groupe « Connaissance des pratiques » du GIS GC HP2E
Terres Inovia	Lecomte Vincent	Terres Inovia	Pratiques culturales Colza	Données d'enquêtes sur les pratiques culturales en colza réalisées tous les 2 ans dans les principaux départements producteurs	Groupe « Connaissance des pratiques » du GIS GC HP2E
Terres Inovia	Lecomte Vincent	Terres Inovia	Pratiques culturales Lin	Données d'enquêtes sur les pratiques culturales en lin oléagineux dans les principaux départements producteurs.	Groupe « Connaissance des pratiques » du GIS GC HP2E
Terres Inovia	Lecomte Vincent	Terres Inovia	Pratiques culturales Soja	Données d'enquêtes sur les pratiques culturales en soja réalisées annuellement dans les principaux départements producteurs.	Groupe « Connaissance des pratiques » du GIS GC HP2E
Terres Inovia	Lecomte Vincent	Terres Inovia	Pratiques culturales Tournesol	Données d'enquêtes sur les pratiques culturales en tournesol réalisées tous les 2 ans dans les principaux départements producteurs	Groupe « Connaissance des pratiques » du GIS GC HP2E

10.4 Annexe 4 - Liste des métadonnées pressenties

Nom	Type	À remplir	Description - Commentaire	Compatibilité Open Data - W3C
Titre	Texte	X	Texte écrit par le fournisseur	dct:title
Description	Texte	X	Texte écrit par le fournisseur pour décrire le jeu de données	dct:description
Thème	Liste	X	Libellés parmi une liste définie	dcat:theme
Mots clés	Texte avec proposition de mots	X	Libellés donnés par le fournisseur (choix libre)	dcat:keywords
Licence pressentie	Liste	X	Ouverte, OpenData,...	dct:licence
Business Model pressenti	Liste	Auto ?	Gratuit, Premium, Traffic, Quota, ...	
Suivi des modifications	Texte	X	Ex : "Le fichier de suivi des modifications est disponible à l'url http://www.blabla.fr/data/xxxx/suivi_modifs "	
Périodicité MàJ	Liste	X	Pluriannuel (Tri-annuel, Biannuel), Annuel, semestriel, trimestriel, bimestriel, mensuel, bihebdomadaire, hebdomadaire, ponctuel (mise à jour irrégulière et non prévisible), pas de mise à jour	dct:accrualPeriodicity
Fournisseur	Texte avec proposition des personnes déjà inscrites	X	Personnes morales ou physiques	dct:creator
Distributeur	Texte avec proposition des personnes déjà inscrites	X	Si différent du fournisseur	dct:publisher
Mail personne référente	email	X		
Emprise (couverture spatiale)	Liste ou coordonnées géographiques ?	X	France métropolitaine, Région administrative, département, Petite Région Agricole (PRA), commune	dct:spatial
Couverture temporelle	Dates	X	Début de l'évènement (campagne culture, salon thématique par exemple) ou de la durée administrative (année, mois, semaine) - Fin de l'évènement	dct:temporal
Granularité géographique (précision)	Liste ou coordonnées géographiques ?	X	France, région, départements, Petite Région Agricole (PRA), commune, canton, BV, exploitation, parcelle, POI (points distincts)	dcat:granularity
Granularité temporelle (précision)	Liste	X	Fréquence d'enregistrement de la donnée : seconde, minute, heure, jour, semaine, mois, année, etc...	
Langue du jeu de données	Liste	X	Français, Anglais, Allemand, Espagnol.	dct:language
Modalité de diffusion des données	Choix multiples	X	API, téléchargement des données, les deux	
Formats disponibles	Choix multiples	Auto ?	format tableur (.csv, .xls, .xlsx, .odf, .dbf), géoréférencés (.shp, .tab, .Gjson + .prj et autres fichiers connexes (.lyr)), standard d'échanges (.xml, .json) et autres (.txt, .pdf, .rtf)	dct:format

Description des libellés des champs	Texte	X	Ex : "Le guide de descriptions des champs est disponible à l'url http://www.blabla.fr/data/xxxx/descriptions_champs"	dct:references
Encodage	Liste	X	Est-ce que l'encodage des données est connu et référencé ? Si oui, précisez l'encodage/le référentiel utilisé ? UTF-8, RGF 93 Lambert 93 (référentiel géographique), fuseau horaire (référentiel temporaire)	
Conformité à certains standards	Texte	X	Compatible AGRO-EDI Europe, références utilisés dans un OAD etc.	
Utilisation du jeu de donnée ou de l'API	Texte	X	La ressource est-elle déjà utilisée ? Si oui, par quels utilisateurs/applications ?	
Identifiant unique		Auto	Sera créé automatiquement lors de la création du jeu de données	dct:identifier
Date de publication du jeu de données	Date	Auto	01/01/2011	dct:issued
Dernière date de mise à jour	Date	Auto	07/04/2014	dct:modified
Version du fichier	Texte	X	v1.2, v2.4, version unique (1.0)	
Date de création du jeu de données	Date	X	01/01/1990	dct:created
Note de qualité des données	Nombre		Défini par la communauté des consommateurs	dcat:dataQuality
accessURL	url	Auto	Pour accéder à la fiche du jeu de données	dcat:accessUrl
downloadURL	url	Auto	Pour télécharger le jeu de données	dcat:downloadURL
webserviceURL	url	Auto	Pour accéder au jeu de données par Web Service	dcat:webServiceURL
Volumétrie		Auto	Nombre de lignes, nombre d'entités géographiques distinctes	dcat:size

10.5 Annexe 5 - Listes des standards, ontologies, schémas et référentiels recommandés.

Liste des sites d'accès aux schémas :

- Agro EDI qui organise et normalise les échanges de données informatisées dans les secteurs agricoles et agro-industriels
- Agroportal du LIRMM (incluant schémas + référentiels) (<http://agroportal.lirmm.fr/>)
- AgroXML, AgroRDF schéma de données d'observation agricole proposé par KTBL <http://agroxml.com/about/index.html>
- Linked Open Vocabulary (LOV) (général)
- World Wide Web consortium W3C qui propose des groupes de travail thématique, des ontologies et des bonnes pratiques de publication de données sur le web.

Des listes de schémas relativement généraux qui pourront être utiles :

- Représentations des données de capteurs dans Semantic Sensor Network (W3C) ; l'équipe COPAIN a réalisé la première publication de données utilisant ce standard en France. Cf Irstea.
- Représentation des données d'observation environnementale, OBOE,
- PROV qui décrit la provenance dans les processus scientifiques (W3C)

Des listes de schémas plus directement liés à l'agronomie existent :

- Crop ontology, <http://www.croponontology.org/> , voir <http://www.biodiversityinternational.org/news/detail/harmonizing-crop-trait-data-crop-ontology/>
- Plant ontology , <http://plantontology.org/>

Concernant la description des données spatiales, plusieurs ontologies sont en cours, par exemple stSPARQL (<http://www.strabon.di.uoa.gr/stSPARQL>), il faudra suivre leur évolution et la position du W3C et de l'OGC sur ce point.

Des référentiels :

- taxref (GBIF) : référentiel plutôt scientifique des êtres vivants
- Le Sandre en partie
- Le référentiel des produits phytopharmaceutiques de l'ACTA (API-AGRO)
- GeoNames (www.geoname.org) jeu de données spatiales du monde

10.6 Annexe 6 - La structure juridique d'un portail de données

Sommaire

- I. Formalisme juridique lié à la création d'un site Internet**
 - A. La propriété du site Internet et du nom de domaine**
 - B. La responsabilité de l'hébergeur et de l'éditeur du site Internet**
 - C. Les mentions légales obligatoires d'un site Internet**

- II. Cadre juridique des données accessibles**
 - A. Principe de libre accès des données publiques et ses exceptions**
 - B. Les données personnelles et les obligations de déclaration auprès de la CNIL**
 - C. Cadre des conditions générales d'utilisation des données mises en ligne**

Introduction

Cette note fait état du droit à ce jour et ne tient pas compte du règlement (UE) 2016/679 du parlement européen et du conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) qui devra être mis en application d'ici deux ans.

Un portail web est un site Internet dédié à un domaine d'intérêt particulier ou construit pour une communauté spécifique. L'objet d'un portail est d'ouvrir une « porte d'entrée » sur un large choix de ressources et de services accessibles par internet.

Plus précisément, ce portail a pour objet de mettre à disposition du public des données et services numériques pour une agriculture française compétitive, ouverte et souveraine.

Créer un portail d'une telle nature suppose d'une part de respecter un certain formalisme juridique, dont la mise en ligne de mentions légales sur le site (I) et, d'autre part, de protéger juridiquement les données accessibles sur le site, notamment par la mise en place d'un cadre contractuel permettant de définir les modalités de la mise à disposition et de la réutilisation des données qui alimenteront le portail pour chaque utilisateur (II).

Le cahier des charges devra donc être adapté à ces obligations.

I- Formalisme juridique lié à la création d'un site internet

Lorsqu'un site Internet est créé, il est nécessaire de s'interroger sur la propriété intellectuelle (droits d'auteur) de ce site et la propriété de son nom de domaine afin d'être en mesure de l'exploiter librement (A), sur les obligations légales incombant à l'hébergeur et à l'éditeur de ce site (B) et enfin, de s'assurer qu'il pourra intégrer les mentions obligatoires devant figurer sur tout site Internet (C).

A. La propriété du site Internet et du nom de domaine

1. Le site Internet

Un site Internet est protégé par le droit d'auteur dès lors qu'il est original (une création originale porte « l'empreinte de la personnalité de son auteur » selon la jurisprudence).

L'article L. 111-1 du code de la propriété intellectuelle dispose que « l'auteur d'une œuvre de l'esprit jouit sur cette œuvre, du seul fait de sa création, d'un droit de propriété incorporelle exclusif et opposable à tous ». Ainsi, la personne qui réalise le site internet en est propriétaire (personne physique ou personne morale).

Si la réalisation du site est confiée à un prestataire, celui-ci en aura la propriété.

Par conséquent, **il convient d'intégrer, dans le contrat de prestation, une clause de cession des droits patrimoniaux du prestataire sur le site au profit du bénéficiaire afin que ce dernier puisse en disposer librement.** Le prix de cette cession doit être clairement identifié dans le contrat.

A noter que la maintenance du site peut être faite par le prestataire, par le bénéficiaire ou par un tiers.

Il sera par ailleurs nécessaire que le prestataire respecte la « Charte Internet de l'Etat », publiée par la circulaire du Premier ministre relative à l'Internet de l'État, n° 5574, du 16 février 2012. Cette charte constitue un guide pratique de l'Internet de l'Etat.

(http://references.modernisation.gouv.fr/sites/default/files/Charte_Internet_de_l'Etat_v1.0.pdf).

2. Le nom de domaine

Lorsqu'un site internet est créé, quel que soit son objet, il est nécessaire de donner à ce site un nom de domaine. Le nom de domaine permet la localisation et l'accès aux pages web du site.

Aux termes de l'article L45-1 du code des postes et des communications électroniques : « Les noms de domaine sont attribués et gérés dans l'intérêt général selon des règles non discriminatoires et transparentes, garantissant le respect de la liberté de communication, de la liberté d'entreprendre et des droits de propriété intellectuelle ».

Le nom de domaine est constitué de plusieurs éléments :

- Le nom : Ex : portail1
- L'extension. Ex : .fr

Concernant l'extension, il est possible de choisir dans le cas de ce portail :

- un nom de domaine générique de premier niveau : .com, .net, .org et .info
- un nom de domaine qui est un code de pays : .fr pour la France

Il est également possible de choisir un sous-domaine afin de permettre l'identification d'une activité dans un secteur particulier (ex : .museum.fr pour un musée, .aero.fr pour un organisme du secteur de l'aéronautique).

Une entité ne pourra pas prétendre à sous-domaine ne correspondant pas à son secteur d'activité.

Le sous-domaine « .gouv.fr » est strictement réservé à l'Etat français. Il permet de garantir qu'il s'agit d'un site officiel de l'administration française. Par principe, l'utilisation de cette extension est obligatoire pour les sites créés par des services de l'Etat, à moins que le service d'information du Gouvernement (SIG) accepte expressément une demande de dérogation.

L'enregistrement du nom de domaine :

Les noms de domaine sont attribués pour une durée limitée, d'un an à dix ans, et renouvelable (article L45-1 du code des postes et des communications électroniques). Le dépôt d'un nom de domaine est payant.

La gestion technique des noms de domaine est confiée à l'ICANN (Internet Corporation for Assigned Names and Numbers). L'ICANN a agréé un certain nombre de bureaux d'enregistrement pour traiter les enregistrements. Ainsi, l'AFNIC est l'office d'enregistrement des noms de domaine en .fr (renouvelé pour un mandat de 5 ans le 28 juin 2012), l'Internic gère les adresses internationales.

En France, il n'est pas possible de s'adresser directement à l'Afnic. Il convient de passer par un bureau d'enregistrement accrédité par l'Afnic. L'annuaire des bureaux d'enregistrement est disponible sur le site de l'Afnic¹⁶.

Si le portail est porté par un service de l'Etat, l'extension choisie devra obligatoirement faire l'objet d'une demande d'agrément auprès du SIG, quelle que soit l'extension du nom de domaine choisie (agrementweb@premier-ministre.gouv.fr).

Avant tout enregistrement, il est nécessaire de vérifier si le nom de domaine existe déjà en effectuant une recherche par le biais du site d'une unité d'enregistrement ou en effectuant une recherche par un programme "Whois"¹⁷.

Si l'enregistrement est effectué par un prestataire, il est primordial que celui-ci le fasse au nom et pour le compte du bénéficiaire afin qu'il ne lui appartienne pas.

B. La responsabilité de l'éditeur et de l'hébergeur du site internet

La loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (LCEN) et la jurisprudence qui a suivi définissent le statut de l'hébergeur d'un site internet et celui de l'éditeur. L'hébergeur est défini comme un intermédiaire technique.

L'éditeur, au sens juridique du terme, est celui qui contrôle le contenu de son site internet. En ce sens, pour un site internet, son rôle se confond généralement avec celui du directeur de publication chargé de rendre public le contenu éditorial. Tous deux sont responsables de ce qui est publié.

L'hébergeur et l'éditeur peuvent également être la même personne (physique ou morale).

L'hébergeur est civilement et pénalement irresponsable du contenu des sites qu'il héberge, à moins qu'il n'ait été averti du contenu manifestement illicite d'un site qu'il héberge et qu'il n'interrompe pas promptement sa diffusion. Il bénéficie d'un régime de responsabilité allégé.

Un contenu manifestement illicite peut concerner des propos diffamatoires, une atteinte au droit à l'image ou encore la violation d'un droit d'auteur.

En revanche, l'éditeur est responsable du contenu qu'il publie sur internet, quel que soit le genre de ce contenu. Article 1382 du Code civil : « tout fait quelconque de l'homme, qui cause à autrui un dommage, oblige celui par la faute duquel il est arrivé à le réparer ».

Il existe un principe de responsabilité en cascade : le directeur de publication est responsable en 1^{er} ressort, ou, à défaut, l'éditeur (si les deux fonctions se distinguent) ou, à défaut, l'auteur du propos fautif (ex : si le propos fautif a été mis en ligne par un internaute).

¹⁶ <https://www.afnic.fr/fr/produits-et-services/services/whois/>

¹⁷ Exemple pour les noms de domaine en .fr : <https://www.afnic.fr/fr/produits-et-services/services/whois/>.

Si l'éditeur collecte, modère ou diffuse des avis en ligne provenant de consommateurs, il est tenu de délivrer à ces consommateurs une information loyale, claire et transparente sur les modalités de contrôle des avis mis en ligne.

Aux termes de l'article L. 111-5-3 du code de la consommation ajoutée par la loi n°2016-1321 du 7 octobre 2016 pour une « République numérique », l'opérateur du portail doit préciser si ces avis font ou non l'objet d'un contrôle et, si tel est le cas, il doit indiquer les caractéristiques principales du contrôle mis en œuvre.

Il doit également indiquer aux consommateurs, dont l'avis en ligne a été rejeté, les raisons qui justifient ce rejet, et mettre en place une fonctionnalité gratuite qui permet aux responsables des données ou des services faisant l'objet d'un avis en ligne, de lui signaler un doute sur l'authenticité d'un avis, à condition que ce signalement soit motivé.

Les obligations précitées ainsi que les rapports et les conflits pouvant naître entre l'éditeur du site et le visiteur peuvent être encadrés dans des conditions générales d'utilisation (cf. infra).

C. Les mentions légales obligatoires d'un site internet

D'une part, selon le site officiel de l'administration française¹⁸ : « tous les sites internet édités à titre professionnel, qu'ils proposent des ventes en ligne ou non, doivent obligatoirement indiquer les mentions légales suivantes » :

- pour une société : raison sociale, forme juridique, adresse de l'établissement ou du siège social (et non pas une simple boîte postale), montant du capital social ;
- adresse de courrier électronique et numéro de téléphone ;
- pour une activité commerciale : numéro d'inscription au registre du commerce et des sociétés (RCS) ;: numéro individuel d'identification fiscale numéro de TVA intracommunautaire ;
- nom et adresse de l'autorité ayant délivré l'autorisation d'exercer quand celle-ci est nécessaire ;
- nom du responsable de la publication ;
- coordonnées de l'hébergeur du site : nom, dénomination ou raison sociale, adresse et numéro de téléphone ;
- pour un site marchand, conditions générales de vente (CGV) : prix (exprimé en euros et TTC), frais et date de livraison, modalité de paiement, service après-vente, droit de rétractation, durée de l'offre, coût de la technique de communication à distance ;
- numéro de déclaration simplifiée Cnil, dans le cas de collecte de données sur les « clients » (non obligatoire, mais recommandé) (cf. infra).

D'autre part, le code de la consommation modifié par la loi « pour une république numérique » dispose que « tout opérateur de plateforme en ligne est tenu de **délivrer au consommateur une information loyale, claire et transparente sur les conditions générales d'utilisation du service d'intermédiation** qu'il propose et sur les **modalités de référencement, de classement et de**

¹⁸ <https://www.service-public.fr/professionnels-entreprises/vosdroits/F31228>

déréférencement des contenus, des biens ou des services auxquels ce service permet d'accéder » (article L111-5-1 du code de la consommation).

Aux termes de ce même article, l'opérateur de la plateforme doit fournir « une information loyale, claire et transparente sur la qualité de l'annonceur et les droits et obligations des parties en matière civile et fiscale ». Un décret précisera le contenu que doivent revêtir ces informations et si cela est applicable au présent portail.

L'opérateur du portail doit également mettre à disposition des consommateurs un espace leur permettant de communiquer avec les fournisseurs de la donnée sur les informations suivantes :

- Les caractéristiques essentielles de la donnée ou du service ;
- Le prix de la donnée ou du service
- Les informations relatives à son identité, à ses coordonnées postales, téléphoniques et électroniques et à ses activités, pour autant qu'elles ne ressortent pas du contexte, ainsi que, s'il y a lieu, celles relatives aux garanties légales, aux fonctionnalités du contenu numérique et, le cas échéant, à son interopérabilité, à l'existence et aux modalités de mise en œuvre des garanties et aux autres conditions contractuelles.

La liste et le contenu précis de ces informations seront fixés par décret en Conseil d'Etat.

Les cookies

Selon les termes de la CNIL, un cookie est « une information déposée sur le disque dur d'un internaute par le serveur du site qu'il visite. Il contient plusieurs données : le nom du serveur qui l'a déposé, un identifiant sous forme de numéro unique, éventuellement une date d'expiration. Ces informations sont parfois stockées sur l'ordinateur dans un simple fichier texte auquel un serveur accède pour lire et enregistrer des informations ».

Si le site internet prévoit l'utilisation de cookies, il faudra indiquer la finalité de ces cookies dans les mentions légales. L'objet d'un cookie peut être de reconnaître un internaute d'une visite à l'autre sur le site, de faire de la publicité ciblée etc.

L'utilisation de cookies ne peut se faire qu'avec le consentement de l'internaute (article 5(3) de la directive 2002/58/CE modifié par la directive 2009/136/CE - L'article 32-II de la loi du 6 janvier 1978 reprend ce principe).

La CNIL recommande de recueillir ce consentement en deux étapes :

- par un bandeau d'information indiquant que la poursuite de navigation sur le site vaut accord pour l'installation et la lecture de cookies. La finalité de ces cookies doit également être précisée sur ce bandeau, ainsi que les modalités d'opposition à l'installation du cookie. Ainsi, si l'internaute poursuit sa navigation, il accepte l'installation de cookies sur son serveur.
- par une information simple et lisible des moyens mis à sa disposition pour accepter ou refuser tout ou partie des cookies, notamment via les mentions légales.

La durée de validité de ce consentement est de treize mois maximum.

Certains cookies sont cependant dispensés du recueil de ce consentement, notamment :

- les cookies " identifiants de session ", pour la durée d'une session, ou les traceurs persistants limités à quelques heures dans certains cas ;
- les cookies d'authentification de l'internaute ;
- les cookies de session créés par un lecteur multimédia ;
- les cookies de session d'équilibrage de charge ;

Le manquement à l'une de ces obligations peut être sanctionné jusqu'à un an d'emprisonnement, 75 000 € d'amende pour les personnes physiques et 375 000 € pour les personnes morales.

Aussi, il est nécessaire de prévoir dans le cahier des charges que le site internet présente, en page d'accueil, un lien visible sur les mentions légales répertoriant les points énumérés ci-avant ainsi qu'un bandeau d'information sur la présence de cookies.

II- Le cadre juridique des données collectées

La donnée est communément définie comme une information sur laquelle on peut fonder un raisonnement, ce qui est connu ou admis comme tel, qui sert de point de départ pour une recherche.

La donnée brute est une donnée non traitée ou contextualisée, elle n'est pas juridiquement protégeable. La donnée brute peut ainsi être diffusée ou utilisée librement.

La base de données est « un recueil d'œuvres, de données, ou d'autres éléments indépendants, disposés de manière systématique ou méthodique (carte scannée, imagerie, etc.), et individuellement accessibles par des moyens électroniques ou par tout autre moyen » (art. L. 112-3 du code de la propriété intellectuelle). Elle est protégée par le droit d'auteur et/ou par le droit sui generis.

La nature des données doit être identifiée de manière précise car elle détermine les conditions d'accès et de réutilisation de la donnée. En effet, par principe, les données publiques doivent être rendues accessibles au plus grand nombre (A) alors que les données personnelles ont un régime spécifique, précisé par la loi « Informatique et Libertés », et ne peuvent être mises à disposition du public librement (B). Les conditions d'utilisation de ces données doivent donc être précisées de manière claire et précise au sein du portail (C).

A. Le principe de libre accès des données publiques et ses exceptions

La Commission d'accès aux documents administratifs (CADA) assimile la donnée à un document administratif. L'art. L300-2 du CRPA dispose que « sont considérés comme documents administratifs, au sens des titres Ier, III et IV du présent livre, quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions, codes sources et décisions. ».

La loi pour « une république numérique » ajoute que « constituent également de tels documents, les documents relatifs à la gestion du domaine privé de l'État et des collectivités territoriales ».

L'article 311-2 du même code précise que cela ne concerne pas les documents préparatoires et non définitifs de l'administration en vue de ses délibérations.

Ainsi il convient de s'interroger sur la mission de service public dont pourrait être investie la SAS qui porterait le portail.

Selon la CADA, les organismes privés chargés de la gestion d'un service public entrent dans le champ de la loi du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal.

Elles sont identifiées à partir d'un certain nombre de critères, parmi lesquels :

- l'objet statutaire ;
- la nature des missions ;
- l'origine des ressources, notamment la part des subventions publiques ;
- la composition de l'organe de direction ;

- le degré de contrôle de l'administration ;
- les modalités de fonctionnement et notamment l'exercice de prérogatives de puissance publique.

La mise à disposition et la publication de données de référence en vue de faciliter leur réutilisation constitue une mission de service public relevant de l'État (art. L. 321-1. – I du code des relations entre le public et l'administration (CRPA)). Si le portail met à disposition des données de référence, ce critère favorisera la qualification de la SAS en organisme privé chargé de la gestion d'un service public.

Sont des données de références au sens de l'article L. 321-1. – I du CRPA ajouté par la loi pour une « république numérique », les données publiques qui :

- « constituent une référence commune pour nommer ou identifier des produits, des services, des territoires ou des personnes ;
- sont réutilisées fréquemment par des personnes publiques ou privées autres que l'administration qui les détient ;
- nécessitent qu'elles soient mises à disposition avec un niveau élevé de qualité, notamment en termes de précision, de disponibilité ou de fréquence de mise à jour ».

Les modalités d'application de cet article seront définies par un décret en Conseil d'État.

Un organisme privé chargé de la gestion d'un service public sera soumis aux mêmes obligations qu'un organisme public. Toutes les données qu'il aura produites ou reçues seront considérées comme des documents publics et soumises aux règles y afférentes (cf. infra) dès lors que ces données sont liées à l'exécution de la mission de service public qu'il exerce, à l'exclusion des autres données qu'il détient ou élabore.

Pour les données non liées à sa mission de service public, l'organisme privé est libre de choisir son mode d'accès et de rediffusion sous réserve du respect du droit des tiers et notamment des droits de propriété intellectuelle attachés aux données (L122-4 du CPI : toute représentation ou reproduction faite sans le consentement de l'auteur est illicite).

Concernant les données à caractère personnel, l'organisme privé est soumis aux mêmes règles qu'un organisme public (cf. règles ci-après).

Aussi, si le portail est gouverné par une personne morale de droit public ou une personne morale de droit privé chargée d'une mission de service public, les données qu'elle élabore ou reçoit sont considérées comme des documents administratifs et sont soumis aux règles y afférentes.

1- Le principe de libre accès des données publiques

Pour toutes les données publiques, c'est-à-dire les données produites par l'Etat, les collectivités territoriales, un établissement public ou une structure privée ayant une mission de service public, **le principe général est l'ouverture et l'accès aux données**. A ce titre, si la structure privée en charge de la gestion du portail est qualifiée de personne privée en charge d'une mission de service public, ce principe lui sera applicable.

Inversement, les personnes morales privées et les personnes physiques ne sont pas tenues par les obligations ci-après si elles ne sont pas investies d'une mission de service public.

L'ouverture et l'accès aux données se fondent sur les textes juridiques suivants :

- Le code des relations entre le public et l'administration (livre III) dispose que **les administrations** (Etat, collectivités territoriales, établissements publics ou privés ayant une mission de service public) **sont tenues de publier en ligne dans un format ouvert et aisément réutilisable ou de communiquer les documents administratifs qu'elles détiennent aux personnes qui en font la demande** (art. L.311-1 modifié par la loi « pour une république numérique »).

En outre, Les administrations précitées **peuvent rendre publique** les documents administratifs qu'elles produisent ou reçoivent (art. L.312-1 du même code).

Excepté les personnes morales dont le nombre d'agents/salariés est inférieur à 50 (ou moins si un décret le prévoit expressément), les administrations précitées doivent publier en ligne les documents suivants lorsqu'ils sont disponibles sous format électronique (art. L.312-1-1 ajouté par la loi « pour une république numérique ») :

- Les documents qu'elles communiquent aux personnes qui en font la demande ainsi que leurs versions mises à jour,
 - Un répertoire des principaux documents dans lesquels les informations publiques figurent. Ce répertoire doit être mis à jour chaque année (L.322-6 ajouté par la loi « pour une république numérique »),
 - Les bases de données qui ne font pas par ailleurs l'objet d'une diffusion publique,
 - Les données, mises à jour de façon régulière, présentant un intérêt économique, social, sanitaire ou environnemental.
- La directive INSPIRE (transposée dans le code de l'environnement par l'ordonnance n° 2010-1232 du 21 octobre 2010 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière d'environnement) **impose aux autorités publiques de publier sur internet leurs données sous format électronique à référence géographique.**

Ces données doivent être mises à disposition du public et partagées librement avec toute autre autorité publique.

Plus précisément, la directive INSPIRE impose de :

- Recenser toutes les données géographiques régies par la directive ;
- Créer des métadonnées et les maintenir à jour ;
- Mettre les métadonnées et les données sous une forme interopérable ;
- Publier ces données sur internet avec :
 - un service de recherche
 - un service de consultation
 - un service de téléchargement
 - un service de transformation (au vu de l'interopérabilité)
 - un service permettant d'appeler des services de données géographiques
- Référencer les métadonnées dans le Géocatalogue
- Partager les données avec les autres autorités publiques

Ces deux régimes - le régime général d'accès aux données (CRPA) et le régime spécial relatif aux données environnementales (code de l'environnement) - s'articulent naturellement. Ils diffèrent principalement sur des points opérationnels.

Dans ce cas, le code de l'environnement prévaudra s'il est plus favorable, tel qu'indiqué ci-dessous.

2- Les exceptions au principe de libre accès des données publiques

Selon la CADA, « si le texte instituant le régime particulier est d'origine législative, joue alors le principe selon lequel les lois particulières dérogent aux dispositions des lois de portée générale (loi de 1978). Le principe étant la liberté d'accès aux documents, les dispositions législatives qui instituent un régime d'accès moins favorable que la loi du 17 juillet 1978 doivent être interprétées de manière restrictive, en particulier dans leur champ d'application ».

Ainsi, ne sont pas communicables :

1- Les avis du Conseil d'Etat, les documents de la Cour des comptes etc. tels que mentionnés à l'article L.331-5 du CRPA (modifié par la loi « pour une république numérique »).

2- Les documents administratifs dont la consultation ou la communication porterait atteinte (art. L.331-5 du même code) :

- a) Au secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif ;
- b) Au secret de la défense nationale ;
- c) A la conduite de la politique extérieure de la France ;
- d) A la sûreté de l'Etat, à la sécurité publique ou à la sécurité des personnes ;
- e) A la monnaie et au crédit public ;
- f) Au déroulement des procédures engagées devant les juridictions ou d'opérations préliminaires à de telles procédures, sauf autorisation donnée par l'autorité compétente ;
- g) A la recherche, par les services compétents, des infractions fiscales et douanières ;
- h) Ou sous réserve de l'article L. 124-4 du code de l'environnement, aux autres secrets protégés par la loi.

3- Les données couvertes par un droit de propriété littéraire et artistique (L311-4 du CRPA). **L'accord du titulaire des droits est alors nécessaire.** Il se fera impérativement par écrit (Article L131-3 du code de la propriété intellectuelle) indiquant les droits de représentation et de reproduction concédés, les supports de diffusion autorisés, l'étendue, la destination, la durée et le territoire couverts.

Le consentement du titulaire des droits d'auteur de la donnée est également obligatoire lors de la collecte de la donnée couverte par un droit de propriété intellectuelle (art. L111-1 du code de la propriété intellectuelle (CPI) - l'auteur jouit d'un droit exclusif sur son œuvre).

4- Les données à caractère personnel et notamment les données « sensibles » (art. L312-1-2 modifié par la loi « pour une république numérique » et L 322-2 du CRPA) (cf. supra).

Pour être communicables, ces documents doivent avoir fait l'objet d'un traitement permettant d'occulter ces mentions.

B. Les données personnelles et les obligations de déclaration auprès de la CNIL

1. Les données à caractère personnel

La loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (dite loi « Informatique et Libertés ») régit les traitements de données à caractère personnel.

Selon l'article 2 de cette loi « constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres » (ex : adresse mail, n° de sécurité sociale, adresse IP etc.).

Il s'agit de toute donnée portant sur une personne physique **identifiée ou identifiable**.

Selon ce même article « pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne ».

Un **traitement de données à caractère personnel** est toute « toute opération ou tout ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé ».

Dans le cadre d'une collecte de données à caractère personnel, **il est obligatoire de déclarer auprès de la CNIL ladite collecte**. La déclaration reprend toutes les informations permettant de connaître les données traitées, la finalité et leur durée de conservation. Il s'agit d'une formalité obligatoire sous peine de sanction pénale, à moins que ce traitement ne soit soumis à une dispense.

Les formalités à respecter concernant le traitement d'une donnée à caractère personnel diffèrent en fonction de la nature des données traitées et de la finalité poursuivie par le traitement.

Cependant, les informations publiques comprenant des données à caractère personnel ne sont communicables que si au moins l'une des trois conditions suivante est remplie (repris par la loi « pour une république numérique » à l'article L.312-1-2 du CRPA) :

- le **consentement des personnes concernées** a été recueilli après leur bonne information sur la finalité et les modalités de la communication ou de la réutilisation des données les concernant ;
- la **réutilisation est autorisée par un texte** législatif ou réglementaire
- les **données sont anonymisées**¹⁹ (c'est-à-dire ne permettant pas, compte tenu de leur niveau d'agrégation, d'identifier les personnes concernées) ;

Si les données sont anonymisées, elles ont préalablement été collectées. Une déclaration de collecte auprès de la CNIL est donc quand même nécessaire.

¹⁹ Le processus d'anonymisation est en principe à la charge de l'administration qui détient les données demandées. Aux termes de [l'article 40 du décret n°2005-1755 du 30 décembre 2005](#) « Lorsque la réutilisation n'est possible qu'après anonymisation des données à caractère personnel, l'autorité détentrice y procède sous réserve que cette opération n'entraîne pas des efforts disproportionnés. ».

Outre la déclaration auprès de la CNIL, il **est obligatoire d'informer les personnes concernées de la finalité de la récolte de leurs données à caractère personnel et de leur droit d'accès, de rectification, d'opposition et de suppression des données qui les concernent.**

De manière générale, il est également indispensable de :

- prendre toutes précautions utiles pour préserver la sécurité des données et, notamment, empêcher qu'elles soient déformées, endommagées, ou que des tiers non autorisés y aient accès ;
- mettre à jour annuellement les données collectées.

Concernant la réutilisation des données autorisée par un texte législatif ou réglementaire :

Dans le cadre de ce portail, des données relatives au découpage parcellaire seront collectées.

L'article 127-10 du code de l'environnement dispose que l'Etat, les collectivités territoriales ainsi que leurs établissements publics respectifs qui ont vocation à en établir peuvent constituer des bases de données numériques comprenant des informations relatives au **découpage parcellaire ainsi qu'aux adresses des parcelles** (en procédant à des interconnexions avec des fichiers détenus par d'autres personnes publiques ou privées et comprenant des données à caractère personnel).

Ils peuvent procéder à la diffusion, y compris par voie électronique, auprès de l'ensemble des personnes publiques et privées, des informations contenues dans ces bases de données géographiques nationales ou locales de référence.

Ces bases de données géographiques nationales ou locales de référence ne peuvent inclure aucune information à caractère personnel autre que le découpage parcellaire et les adresses des parcelles.

Le décret n°2011-223 du 1^{er} mars 2011 précise que les informations susceptibles d'être diffusées comprennent (art. R127-10 du code de l'environnement) :

- Les données de localisation géographique relatives au découpage parcellaire cadastral : référence des parcelles cadastrales, localisation de celles-ci, localisation de leurs contours ;
- Les données de localisation géographique relatives aux adresses des parcelles : localisation et, le cas échéant, voie de situation, numéro dans la voie et compléments éventuels.

A ce titre, l'Etat, les collectivités territoriales ainsi que leurs établissements publics peuvent diffuser librement les données de localisation géographique relatives au découpage parcellaire cadastral et les données de localisation géographique relatives aux adresses des parcelles sans procéder à une quelconque anonymisation ni sans l'accord des personnes concernées.

Cet article du code de l'environnement ne mentionne pas les organismes privés chargés d'une mission de service public. Il est donc possible de s'interroger sur l'application de cette dérogation au portail si le portail est gouverné par une SAS.

Concernant l'anonymisation et big data :

La CNIL doit être à même de vérifier l'efficacité des procédés d'anonymisation utilisés. Il est recommandé de saisir la CNIL au préalable afin de faire valider un tel processus lorsqu'il concerne des traitements importants.

Dorénavant, avec la loi pour « une république numérique », la **CNIL peut certifier ou homologuer** et publier des référentiels ou des méthodologies générales aux fins de certification **de la conformité des processus d'anonymisation des données à caractère personnel**, notamment en vue de la réutilisation d'informations publiques mises en ligne (article 37 bis de la loi « Informatique et libertés » ajouté par la loi pour une « république numérique »).

Le G29, qui regroupe l'ensemble des CNILS européennes, propose 3 critères pour évaluer une bonne anonymisation :

- « **l'individualisation** : *est-il toujours possible d'isoler un individu ?*
- La **corrélation** : *est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?*
- **L'inférence** : *peut-on déduire de l'information sur un individu ?* »

Selon la CNIL, « un ensemble de données pour lequel au moins un des trois critères n'est pas respecté ne pourra être considéré comme anonyme qu'à la suite d'une analyse détaillée des risques de ré-identification ».

Pour qu'une donnée soit considérée comme anonyme, le **traitement doit être irréversible**.

La CNIL répertorie 3 méthodes d'anonymisation :

- la **substitution** parfois appelée aussi « **pseudonymisation** » (procédé consistant à remplacer l'identifiant d'une personne par un autre identifiant choisi arbitrairement)
Selon le rapport d'information du Sénat du 16 avril 2014 « il est nécessaire que, pour chaque personne, ce pseudonyme soit unique ».
- la **suppression** ou le « **masquage** » (procédé consistant « à dégrader l'information initiale, en supprimant certaines données, ou, au contraire, en ajoutant des informations qui noient la donnée initiale identifiante »)
- l'**agrégation** (procédé consistant « à rassembler plusieurs données de même type, afin de produire une donnée agrégée qui conserve l'information dont on a besoin, mais rend impossible l'identification de la part d'information agrégée qui correspond à un individu donné ». Selon le rapport d'information du Sénat du 16 juin 2016, la CNIL recommande de ne pas faire figurer de statistiques conçues à partir d'une agrégation inférieure à dix éléments.

A l'heure actuelle, l'anonymisation totale ne peut pas être garantie techniquement en raison de la masse de données de géolocalisation qui permettent à partir de quelques « points » de ré-identifier les personnes, exemple dans un article scientifique publié récemment : Arvind Narayanan and Vitaly Shmatikov « Robust de-anonymization of large datasets »

https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf.

Le rapport d'information du Sénat du 16 avril 2014 confirme que les procédés d'anonymisation mis en place ne présentent pas tous la même efficacité et que le risque de ré-identification n'est pas nul selon la nature de la donnée (voir également l'avis du G29 en ce sens²⁰).

Les rapporteurs nuancent toutefois ce propos en estimant « qu'à ce jour, les dispositifs d'anonymisation utilisés ont été suffisamment robustes pour éviter toute diffusion accidentelle de données personnelles ».

Néanmoins, quel que soit le procédé d'anonymisation utilisé, il est indispensable que le responsable de traitement contrôle régulièrement le caractère anonyme des données produites.

Etant donné que l'anonymisation parfaite n'existe pas, **il est recommandé d'informer l'utilisateur du modèle d'anonymisation utilisé** afin qu'il puisse consentir au traitement effectué en connaissance de cause. La technique utilisée pourra être précisée dans les Conditions Générales d'Utilisation (CGU).

2. Les données sensibles

Au sens de l'article 8 de la loi du 6 janvier 1978 modifiée, sont considérées comme sensibles les « données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci ».

Il est interdit de collecter ou de traiter de telles données. Le non respect de cette interdiction est passible de sanctions pénales.

Néanmoins, il est possible de déroger à cette interdiction dans les cas prévus par l'article 8 de la loi précitée :

- Les traitements pour lesquels la personne concernée a donné son consentement exprès ;
- Les traitements nécessaires à la sauvegarde de la vie humaine, mais auxquels la personne concernée ne peut donner son consentement ;
- Les traitements nécessaires à la constatation, à l'exercice ou à la défense d'un droit en justice ;
- les traitements nécessaires aux fins de médecine préventive et les traitements nécessaires à la recherche dans le domaine de la santé ;
- Les traitements mis en oeuvre par une association ou tout autre organisme à but non lucratif et à caractère religieux, philosophique, politique ou syndical pour la gestion de ses membres ou des personnes qui entretiennent avec l'organisme des contacts réguliers dans le cadre de son activité;
- Les traitements réalisés par l'Insee ou les services statistiques ministériels ;
- Les traitements faisant l'objet d'un procédé d'anonymisation « à bref délai » reconnu conforme par la CNIL ;
- Les traitements de données ayant été rendues publiques par la personne concernée ;
- Les traitements justifiés par l'intérêt public.

²⁰ Le big data pose la question de la possibilité de ré-identification des personnes. L'appréciation de l'anonymisation a évolué au fil du temps : voir l' Avis G29 du 10 avril 2014 : conditions et critères pour aboutir l'anonymisation total (Article 29 « Data protection working Party » http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf et sur le site de la CNIL <https://www.cnil.fr/fr/le-g29-publie-un-avis-sur-les-techniques-danonymisation-0>).

A noter que dans certains cas, la loi prévoit que l'interdiction de collecte de données sensibles ne peut être levée par le consentement de la personne concernée. Il est alors impossible de déroger à cette interdiction.

C. Les conditions d'accès, d'utilisation et de réutilisation des données

Il est indispensable de régir les conditions d'accès, d'utilisation et de réutilisation des données mises en ligne sur le site, ainsi que les modalités d'édition des données. Pour cela, il est nécessaire de connaître le titulaire de la donnée et d'obtenir son autorisation afin de ne pas contrevenir au droit de propriété littéraire et artistique qui pourraient être grevés à la donnée (1) puis de préciser les conditions de réutilisation possibles (2).

A ce titre, il convient d'insérer sur le portail des conditions générales d'utilisation (CGU) des données accessibles applicables aux utilisateurs des données et des conditions générales (CG) applicables aux fournisseurs des données (3).

Les CGU permettront également de faire figurer le numéro de déclaration effectué auprès de la CNIL en cas de collecte de données à caractère personnel et d'encadrer juridiquement les rapports et les conflits pouvant naître entre l'éditeur du site et le visiteur.

1. La propriété et la titularité des données

La propriété :

L'article L. 111-1 du code de la propriété intellectuelle dispose que « l'auteur d'une œuvre de l'esprit jouit sur cette œuvre, du seul fait de sa création, d'un droit de propriété incorporelle exclusif et opposable à tous ».

Cependant, par principe, les données créées par un salarié dans le cadre de ses missions appartiennent à l'employeur. L'auteur de la donnée devra néanmoins être cité comme auteur, à moins qu'il ne s'y oppose.

Si les données ont été créées par un salarié hors de ses missions, deux cas de figure apparaissent :

- La donnée a été créée hors missions du salarié mais se rattache à son domaine d'activité ou a été faite dans l'exécution de ses fonctions ou grâce aux moyens et connaissances mis à sa disposition : dans ce cas, le salarié est propriétaire de la donnée mais l'employeur peut se faire attribuer sa propriété ou obtenir sa jouissance sous réserve de lui verser une contrepartie financière,
- La donnée a été créée hors mission du salarié et n'est pas rattachable à son organisme : la propriété de la donnée lui appartient.

La titularité :

Pour disposer librement de la donnée grevée de droit de propriété littéraire et artistique, il ne suffit pas d'en être propriétaire, il est également nécessaire de détenir les droits qui y sont associés, d'en être titulaire.

Ainsi, si le propriétaire a concédé des droits d'utilisation et/ou d'exploitation à un tiers sur ses données, il convient de vérifier qu'il puisse encore en disposer librement pour pouvoir les ajouter sur la plateforme. Pour cela, il sera nécessaire que le propriétaire de la donnée contrôle le contrat de licence qu'il a conclu avec le tiers.

Ex : si le tiers bénéficie d'une licence exclusive sur la donnée, le propriétaire n'en sera pas titulaire et ne pourra a priori pas la rendre accessible sur le portail.

Les données peuvent également avoir été créées dans le cadre d'une collaboration de recherche. Dans cette hypothèse, les données seront certainement soumises à la copropriété. Il faudra alors se référer au contrat qui lie ces partenaires afin de savoir si le propriétaire peut en disposer librement ou s'il doit préalablement obtenir l'accord de son partenaire pour les diffuser.

En tout état de cause, le fournisseur doit avoir l'autorisation du(des) titulaire(s) des données grevées de droits d'auteur sur la diffusion, l'utilisation et la réutilisation et l'exploitation des données.

Spécificité des bases de données :

La protection des bases de données est régie par le code de la propriété intellectuelle (CPI) en son livre III, titre IV (article 342-1 et suivants).

Elle peut se faire par deux moyens, indépendant l'un de l'autre (art L.341-1 du CPI) :

- Par le droit sui generis : dans ce cas, il faut que la constitution, la vérification ou la présentation du producteur de la base atteste d'un investissement financier, matériel ou humain substantiel. C'est le risque économique et la prise d'initiative qui vont engager une éventuelle protection de la base de données.
- Par le droit d'auteur : dans ce cas, le choix et/ou la disposition des matières utilisées par l'auteur pour constituer la base de données doit être original. Il faut pouvoir déceler une création intellectuelle de la part de son auteur.

En somme, c'est la « forme » de la base de données qui va être protégée et non le contenu.

Le producteur est celui qui réalise l'investissement financier, matériel ou humain nécessaire à la protection de la base. Le titulaire de la base est en principe l'employeur puisque c'est lui qui réalise l'investissement financier (paiement du salaire, des outils du travailleur...). Si la base de données est fabriquée pour le compte d'un tiers, c'est à lui que revient la titularité de la base puisqu'il en a pris l'initiative et paye le prix de la commande.

L'auteur est le créateur de la base de données (de sa « forme »). Il détient les droits patrimoniaux et moraux sur la base. Sauf cession expresse, il en est le titulaire.

Par conséquent, lorsque le fournisseur souhaite rendre accessible ses données sur le portail, il est impératif qu'il certifie en être titulaire afin de ne pas contrevenir aux droits de propriété littéraire et artistique détenus par un tiers. Cet engagement de titularité s'obtiendra via les Conditions Générales (CG).

2. Conditions de réutilisation

Distinction entre droit d'accès et droit de réutilisation :

Il faut distinguer le droit d'accès et le droit de réutilisation d'une donnée qui **devront être intégrés conjointement dans le cadre du portail**. Cette distinction est importante car la liberté d'accès **n'implique pas un droit automatique à la réutilisation** :

L'accès aux données publiques :

L'accès est le fait de permettre à toute personne le droit d'obtenir la communication de documents. Par principe, **l'accès aux données publiques doit être gratuit** (article R324-1 du CRPA). Les données doivent être mises à disposition dans un format électronique ouvert et aisément réutilisable c'est-à-dire lisible par une machine » (art. L311-1 du code des relations entre le public et l'administration).

Le législateur impose aux administrations qui produisent des informations publiques de **mettre à la disposition des usagers un « répertoire des principaux documents dans lesquels ces informations figurent »** (art. L322-6). Ce répertoire doit recenser pour chacun des documents : leur titre exact, leur objet, leur date de création, les conditions de leur réutilisation et, le cas échéant, leur date et objet des mises à jour. Ce répertoire doit être accessible en ligne (art. R322-7) et mis à jour chaque année.

La réutilisation des données publiques :

La réutilisation des données publiques est le fait de donner la possibilité de réutiliser les informations publiques à d'autres fins que celles pour lesquelles elles ont été élaborées ou reçues (L321-1 du CRPA). La réutilisation ne peut pas être limitée dans son étendue, c'est-à-dire qu'une réutilisation à but commercial est possible (sauf exception).

Conformément au II.A.2 de la note et rappelé à l'article L.321-2 du CRPA, **ne constitue pas une information publique les documents sur lesquels des tiers détiennent des droits de propriété intellectuelle**.

Par conséquent, l'opérateur du portail ne peut permettre la réutilisation de données couvertes par un droit de propriété intellectuelle sans l'accord de son titulaire (aux termes de l'article L122-4 du CPI toute représentation ou reproduction faite sans le consentement de l'auteur est illicite).

En revanche, l'article L. 321-3, ajouté par la loi « pour une république numérique » dans le code précité, dispose que les droits **des administrations** (hors administration ayant une mission de service publique à caractère commerciale soumise à la concurrence) **ne peuvent faire obstacle à la réutilisation des bases de données qu'elles publient au titre du droit sui generis**, c'est-à-dire du droit des producteurs de base de données à interdire l'extraction et la réutilisation de la totalité ou d'une partie qualitativement ou quantitativement substantielle du contenu de la base.

- Modalités de réutilisation

La réutilisation des informations publiques s'entend « **sans altération, sans dénaturation et en mentionnant la source et la date de dernière mise à jour** » (art. L322-1 du CRPA).

L'article L321-1 du CRPA, modifié par la loi « pour une république numérique », dispose que **les informations publiques figurant dans des documents communiqués ou publiés par les administrations peuvent être utilisées par toute personne qui le souhaite à d'autres fins que celles de la mission de service public** pour les besoins de laquelle ces documents ont été produits ou reçus.

Il n'est pas possible d'apporter d'autres restrictions à la réutilisation des données sauf pour des motifs d'intérêt général et de façon proportionnée. Ces restrictions ne peuvent avoir pour objet ou pour effet de restreindre la concurrence (art. L323-2 du même code).

Toutefois, le code de l'environnement offre la possibilité aux « services de consultation permettant au moins d'afficher des données, de naviguer, de changer d'échelle, d'opter pour une vue panoramique, ou de superposer plusieurs séries de données consultables et d'afficher les légendes ainsi que tout contenu pertinent de métadonnées » de restreindre cet accès et d'empêcher le téléchargement de ces données afin d'éviter toute réutilisation commerciale (art. L.127-7 du code de l'environnement).

En tout état de cause, il est indispensable de mettre en ligne des conditions générales d'utilisation de ces données afin d'encadrer leurs conditions de réutilisation.

- Principe de gratuité et exception à l'accès et la réutilisation des données publiques

Par principe, l'accès de ces données publiques doit être gratuit et complet. Les données doivent également être utilisables et réutilisables gratuitement (art. L324-1 du CRPA).

La loi « pour une république numérique » dispose que lorsque la réutilisation à titre gratuit donne lieu à l'établissement d'une licence, celle-ci doit être choisie parmi une liste fixée par décret. Lorsqu'une administration souhaite recourir à une licence ne figurant pas sur cette liste, cette licence doit être préalablement homologuée par l'État, dans des conditions fixées par décret.

Nb : les décrets n'ont pas été publiés.

Cependant, il est possible de soumettre la réutilisation des données au versement d'une contrepartie financière lorsqu'elles sont tenues de couvrir par des recettes propres une part substantielle des coûts liés à l'accomplissement de leurs missions de service public (art. L324-1 du CRPA).

Pour cela, il faut proposer un service de commerce électronique ad hoc, et soumettre l'utilisation de ces données à une licence d'utilisation et/ou d'exploitation.

La redevance perçue ne doit pas dépasser le montant total des coûts liés à la collecte, à la production et à la mise à disposition au public. Le mode de calcul doit être public et transparent.

Ainsi, pour établir le montant des redevances, il faut tenir compte :

- Du coût de mise à disposition, voire d'anonymisation
- Du coût de collecte
- Du coût de production

- Voire d'une rémunération raisonnable de ses investissements comprenant, le cas échéant, une part au titre des droits de propriété intellectuelle

Les redevances doivent être nécessaire au maintien des données et des services correspondants (notamment s'il s'agit d'un volume important de données comprenant des mises à jour mensuelles).

Aussi, le produit total de la redevance ne doit pas être supérieur à la somme des coûts majorée de la rémunération des investissements sur la période d'amortissement, il ne doit pas permettre d'engendrer des bénéfices.

Ces conditions de réutilisations doivent être les mêmes pour des catégories comparables de réutilisation, elles ne doivent pas être discriminatoires.

Le montant pour la réutilisation de ces données doit être fixé à l'avance et publié sous forme électronique, via un contrat de licence (art. L.323-1 du CRPA). Dans le cas contraire, la réutilisation est considérée comme gratuite.

Le montant doit être révisé au moins tous les 5 ans (art. L324-3 code précité).

Des contrats d'exclusivité avec des structures privées ou publiques pour la réutilisation des données publiques sont possibles sous certaines conditions mais ils ne peuvent pas dépasser 10 ans et sont examinés tous les 3 ans.

La réutilisation des données privées

Ces conditions de réutilisation ne sont obligatoires qu'en ce qui concerne les données publiques. A contrario, les données privées, notamment les données grevées par un droit de propriété littéraire et artistique appartenant à un tiers privé, ne sont pas régies par ces conditions. Le fournisseur proposera l'accès, la diffusion et la réutilisation de la donnée selon les conditions fixées par le titulaire de ladite donnée.

3. Le contenu des CG

Compte tenu de ce qui précède, **il est nécessaire de prévoir dans le cahier des charges que le site Internet présente un lien facilement accessible sur les CG/CGU (CG du fournisseur et CGU de l'utilisateur)**

Pour le fournisseur, les CG devront notamment préciser :

- la nature des données mises à disposition ;
- Ses droits et obligations (notamment un engagement de titularité du jeu de donnée) ;
- Les droits de propriété intellectuelle afférant aux données ;
- Les droits qu'il confère à l'utilisateur en fonction des différents types de licence proposée (accès, utilisation, réutilisation...)
- Le degré de responsabilité de l'éditeur ;
- Les obligations relatives à la « loi Informatiques et libertés » (sur les données à caractère personnel)
- Le droit applicable.

Tel qu'énoncé, les CG devront préciser les différents types de licence pouvant être proposés selon le jeu de donnée mis à disposition :

Lorsqu'un fournisseur de données souhaite rendre accessible ses données, il devra préalablement déterminer les droits qu'il souhaite accorder sur ses données. Pour simplifier la procédure et obtenir une meilleure visibilité des droits accordés, il est préférable de proposer plusieurs types de licence :

- les licences fixées par décret conformément à la loi « pour une république numérique ».
- les licences spécifiques demandées par les fournisseurs eux-mêmes lorsqu'ils détiennent des droits d'auteur sur leurs données,
- les licences commerciales précisant les modalités financières.

Pour l'utilisateur, les CGU devront notamment préciser :

- la nature des données mises à disposition ;
- Les droits de propriété intellectuelle afférant aux données ;
- Les modalités d'accès aux données ;
- Les droits conférés à l'utilisateur en fonction des différents types de licence proposée (accès, utilisation, réutilisation...) et ses obligations, notamment l'obligation de citer les sources lorsqu'une donnée est mise en ligne sur le site ;
- Les droits et obligations des fournisseurs de données (notamment un engagement de titularité du jeu de donnée) ;
- Le degré de responsabilité de l'éditeur ;
- Les obligations relatives à la « loi Informatiques et libertés » (sur les données à caractère personnel)
- Le droit applicable.

III- Points essentiels devant figurer dans le cahier des charges

- Un lien sur les mentions légales (en page d'accueil),
- Un lien sur les CG (prévoir des CG pour le fournisseur et des CGU pour l'utilisateur),
- Un lien sur les contrats de licence (et la possibilité de modifier ces accords a posteriori),
- Un système de paiement si les données sont soumises à redevance,
- Un bandeau d'information si le site prévoit l'utilisation de cookies.

Par ailleurs, le contrat de prestation devra prévoir une cession des droits patrimoniaux du prestataire sur le site au profit du bénéficiaire. Le nom de domaine devra être déposé au nom et pour le compte de ce dernier.



www.irstea.fr

Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture

Direction générale
1, rue Pierre-Gilles de Gennes
CS 10030
92761 Antony Cedex
tél. +33 (0)1 40 96 61 21
fax +33 (0)1 40 96 62 25



SUIVEZ-NOUS SUR :

