



Direction des soutiens directs agricoles

POUVOIR ADJUDICATEUR :

Agence de services et de paiement (ASP)

2, rue du Maupas

87040 LIMOGES cedex 1

Annexe n°2 « Cadre organisationnel et description des applicatifs »

Table des matières

TABLE DES MATIERES	2
LISTE DES FIGURES	3
1 OBJECTIFS DE L'ANNEXE	4
2. DESCRIPTION DE L'ASP	4
LA DIRECTION DES SOUTIENS DIRECTS AGRICOLES (DSDA)	4
LA DIRECTION DU DEVELOPPEMENT RURAL ET DE LA PECHE (DDRP)	5
LA MISSION DE COORDINATION DES ORGANISMES PAYEURS DES FONDS EUROPEENS (MCFA)	7
3. CONTEXTE DE LA PAC	8
LA PAC EN FRANCE	8
L'ASP, ORGANISME PAYEUR ET ORGANISME DE COORDINATION DES ORGANISMES PAYEURS FRANÇAIS	9
ARCHITECTURE DES SI DECOULANT DE CETTE ORGANISATION	9
2 LA DESCRIPTION DES APPLICATIFS PAR GRAND CHANTIER	10
CHANTIER N°1. SYNAPSE : UN APPLICATIF DEDIE A LA PERFORMANCE	10
CHANTIER N°2. VALORISATION DES DONNEES D'ISIS	13
CHANTIER N°3. VALO RDR4 HSIGC FEADER	19
CHANTIER N°4. PORTAIL UTILISATEURS DU LAC DE DONNEES AGRICOLES	21
SYNTHESE DES APPLICATIFS DU LDA	22
3 DATES-CLES DU LDA	23
RECAPITULATIF	23
PROCESSUS DE TRAITEMENT DES EVOLUTIONS FONCTIONNELLES	23
ORGANISATION DE L'ASSISTANCE AUX UTILISATEURS	23
4 LES UTILISATEURS DU LAC	24
DESCRIPTION DES UTILISATEURS	24
GROUPES DE SECURITE	25
4. FONCTIONNEMENT DU LAC DE DONNEES DANS LE CADRE DU MARCHE SUBSEQUENT ISIS N°2 (2021-2024)	26
LA COMITOLOGIE ACTUELLE DU LDA (2021-2024)	26
LE DIMENSIONNEMENT DES EQUIPES MOE ACTUELLES	26
5. GOUVERNANCE DE LA DONNEE DU LDA	27
ACTEURS ET INSTANCES	27
LES DROITS D'ACCES AUX DONNEES ET LE PORTAIL	30

Liste des figures

Figure 1. Description des aides de la PAC par Fonds	8
Figure 2. Schéma d'organisation interne à l'ASP	10
Figure 3. Système IT pour la gestion de la PAC & de la performance	11
Figure 4. vue d'ensemble technique de SYNAPSE.....	12
Figure 5. Schéma fonctionnel du déversement de BasePerf DSDA	14
Figure 6. Articulation du Réplica avec la valorisation de données ISIS	15
Figure 7. Schéma d'architecture de BasePerf DSDA.....	16
Figure 8. Architecture de Dédoublx	17
Figure 9. Articulation Base Perf DSDA / Synapse / Dédoublx / Synapse.....	18
Figure 10. Traitement de Suivi Campagne PAC dans Dataiku	19
Figure 11. Présentation schématique de Valo FEADER HSIGC.....	20
Figure 12. Vue d'ensemble technique de Valo FEADER HSIGC.....	21
Figure 13. Diagramme organisationnel des acteurs et instances de la gouvernance des données.....	28
Figure 14. Architecture du LDA et catalogue de données.....	30
Figure 15. Page de connexion au Portail Utilisateurs du LDA.....	31
Figure 16. Mécanisme d'authentification au portail	31
Figure 17. Schéma de la gestion centralisée des droits.....	32
Figure 18. Mode d'accès RBAC	32
Figure 19. Schéma des données référentielles dispersées.....	33
Figure 20. Unique source de vérité	33
Figure 21. Arbre d'exposition	34

1 Objectifs de l'annexe

La présente annexe a pour objectif de décrire les éléments d'ordre organisationnel suivants :

- La description des autorités fonctionnelles du Lac de données agricoles (LDA ou Lac) ;
- La structure et les rôles de la direction de programme (DP) LDA ;
- La description des applicatifs du Lac ;
- Les organisations actuelles ;
- La gouvernance de la donnée mise en place.

2. Description de l'ASP

La direction des soutiens directs agricoles (DSDA)

Au sein de l'Agence de Services et de Paiement (**ASP**), la direction des soutiens directs agricoles (**DSDA**) pilote la mise en œuvre des aides directes de la politique agricole commune (**PAC**) et représente la maîtrise d'ouvrage du Système d'Information Intégré et de Contrôle (**SIGC**).

En tant que maîtrise d'ouvrage du SIGC, la DSDA assure :

- le pilotage de l'intégration du cadre réglementaire de la PAC dans le Système Intégré de Gestion et de Contrôle (SIGC) ;
- la traduction des règles communautaires dans la définition des processus métier (de la télé-déclaration jusqu'au paiement en passant par l'instruction, les contrôles, la liquidation et l'ordonnancement) et des expressions de besoins ;
- le pilotage de l'activité des services déconcentrés de l'Etat (DDTM/DAAF) en charge de l'instruction des aides ;
- le pilotage de l'activité des directions régionales de l'ASP (DR ASP) en charge des contrôles sur place ;
- la liquidation et l'ordonnancement des aides (le paiement des aides étant assuré par l'Agence Comptable de l'ASP) ;
- la prise en charge des audits et contrôles internes ;
- la valorisation de données métier (statistiques, contrôles, audits).

Les entités composant la DSDA

La DSDA est constituée d'une direction de pilotage, composée d'une directrice assistée de deux adjoints, à laquelle sont rattachés :

- le service d'instruction des aides ;
- le service des contrôles ;
- le service d'ordonnancement ;
- le service audits contrôle interne et lutte anti-fraude ;
- une direction des opérations ;
- une direction de programme Lac de données agricoles ;
- une direction de programme nouveau SIGC ;
- une direction de projet Suivi des surfaces en temps réel (3STR) ;
- une mission Statistiques.

La direction de programme Lac de données agricoles est l'interlocutrice principale du Titulaire actuel en charge du Lac de données agricoles et le sera de même avec le futur Titulaire de l'accord-cadre en charge de la Tierce Maintenance Applicative (TMA) du Lac de données agricoles. Elle interviendra aussi dans la relation avec le Titulaire de l'accord-cadre 24-21-04 en charge de l'exploitation et de l'hébergement du lac de données agricoles.

La direction de projet 3STR est l'interlocutrice principale du Titulaire actuel en charge du 3STR et le sera de même avec le futur Titulaire de l'accord-cadre en charge du 3STR.

La direction de programme du nouveau SIGC sera l'interlocutrice principale du futur Titulaire de l'accord-cadre 24-21-03 en charge de la construction du nouveau SIGC. Dès lors où le nouveau SIGC sera construit et où sa gestion basculera en TMA, la direction de programme du futur SIGC passera le relais à la direction des opérations.

La direction des opérations sera l'interlocutrice principale du futur Titulaire de l'accord-cadre 24-21-01 en charge de l'assistance au pilotage et à la coordination des MOE et de la conception fonctionnelle, du futur Titulaire de l'accord-cadre 24-21-02 en charge de la TMA du SI actuel, du futur Titulaire de l'accord-cadre 24-21-04 en charge de l'exploitation et de l'hébergement des SI, et dès lors où le nouveau SIGC sera construit et où sa gestion basculera en TMA, du futur Titulaire de l'accord-cadre 24-21-03 en charge de la TMA du nouveau SIGC.

La direction de programme du lac de données agricoles

La gouvernance repose sur une direction de programme (DP) dédiée, garante du bon déroulement du projet, du respect des plannings, de l'établissement des priorités et de la coordination de l'ensemble des intervenants. Il s'agit également d'un moyen fort pour garantir une forte coordination des ressources et des équipes provenant des deux directions métier agricoles : la DSDA pour le SIGC et la Direction du développement rural et de la pêche (DDRP) pour le hors-SIGC, pilotant conjointement le lac de données. Cette coordination reposera ainsi sur un directeur de projet unique qui répondra et rendra compte aux deux directions métier.

A titre informatif, au 1^{er} janvier 2025, les ressources humaines au sein de la direction de programme du lac de données agricoles se déclinera de la manière suivante :

- 7 ETP internes :
 - o 1 directeur de programme ;
 - o 1 directeur de programme adjoint (profil technique) ;
 - o 1 chef de projet informatique ;
 - o 1 datascientist ;
 - o 1 ingénieur donnée ;
 - o 1 référent fonctionnel Performance PAC ;
 - o 1 référente fonctionnelle Données ISIS ;
- 2 ETP en assistances externes :
 - o 2 architectes ;
 - o 1 recetteur / référent Synapse.

La DP s'appuie sur d'autres ressources de la DSDA (Correspondant RSSI, DAF) et sur la DDRP (DAF) pour piloter le lot et les aspects sécurité, et ponctuellement sur la DNSI. La mobilisation des experts métier est la suivante :

- pour la DSDA : 3 datascientists (2 agents de la mission Statistiques, 1 datascientist au sein du service audits contrôle interne et lutte anti-fraude) ;
- pour la DDRP : la DDRP mobilise une équipe dédiée au pilotage, et au suivi, et à la recette du projet Valo RDR4 (environ 8 personnes).

La Direction du développement rural et de la pêche (DDRP)

La direction du développement rural et de la pêche (DDRP) est l'organisme payeur du Fonds européen agricole pour le développement rural hors aides surfaces (Feader HSI^{GC}, hors Corse). Elle assure la maîtrise d'ouvrage des différents Systèmes d'Information 2023-2027 de la sphère HSI^{GC} hors Corse liés à la mise en œuvre du Plan Stratégique National de la programmation 2023-2027 (**PSN**).

En tant qu'organisme payeur du Feader HSI^{GC} et maîtrise d'ouvrage des SI 2023-2027 Feader HSI^{GC}, la DDRP assure :

- la conformité des aides Feader HSI^{GC} avec les règles communautaires
- la mise en œuvre globale des aides Feader non régionalisées dans le respect des aspects réglementaires et financiers du PSN et en ayant défini le cadre de délégation destiné aux Autorités de Gestion (AG) des aides Feader régionalisées
- le pilotage du paiement des aides nationales et confiées par le MASAF
- le pilotage de l'activité des directions régionales de l'ASP (DR ASP) en charge :
 - o des contrôles d'ordonnancement et de paiement (COP)
 - o des contrôles sur place (CSP) et contrôles ex-post des aides Feader HSI^{GC} non régionalisées (Protection des troupeaux contre la prédation)
 - o du contrôle au sens large de la délégation de l'OP envers les autorités de gestion (AG) :
 - audit national du Descriptif du Système de Gestion et de Contrôle (DSGC)
 - audit de Sécurité des Systèmes d'information des AG (SSI)
 - audits annuels et thématiques

- contrôles dossiers (contrôles administratifs)
- la liquidation et l'ordonnancement des aides (le paiement des aides étant assuré par l'Agence Comptable de l'ASP)
- la comptabilisation des paiements et la déclaration des dépenses Feader
- la prise en charge des audits et contrôles internes
- la valorisation de données métier (statistiques, contrôles, audits)

Les entités composant la DDRP

La DDRP est constituée d'une direction de pilotage, dirigée par une directrice assistée de deux adjoints pilotant les deux périmètres suivants :

- Direction des opérations ;
- Direction métier.

A la directrice sont rattachées une chargée de mission contrôle interne-externe et une Mission administrative et financière, assurant l'appui transverse à la direction, la gestion des ressources humaines, du budget et de la comptabilité ainsi que la communication.

La direction des opérations est composée :

- D'une ligne de produit RDR « Osiris – instruction » en charge :
 - o de la gestion du Système de gestion des aides HSIGC des programmations Feader antérieures à 2023-2027 et des aides FEP ; FEAMP et nationales ;
 - o de l'instrumentation de la solution MGDIS SAFRAN retenue dans le cadre des aides Feader non régionalisées, de l'assistance technique forfaitaire, des aides Feader dont le pilotage est national pour Mayotte et Saint-Martin et des aides nationales confiées à l'ASP, pour 2023-2027 ;
 - o de nouveaux outils tels que CERTIBASE dédié à la collecte des données des exploitations certifiées Haute Valeur Environnementale (HVE).
- D'une ligne de produit RDR « Référentiels performance – Cœur de métier OP » propre à la programmation 2023-2027 en charge :
 - o d'une partie référentiels et performance qui comprend le pilotage et la maîtrise d'ouvrage des produits et systèmes d'information suivants :
 - Dataplan destiné à la gestion partagée de la planification du PSN
 - RCIPAC destiné à l'identification des individus bénéficiaires de la PAC et des aides entrant dans la sphère HSIGC d'ordonnancement et de paiement
 - RefPAC destiné à la gestion des référentiels communs de la PAC et partagés entre les différents partenaires
 - Sigma : réceptacle et base de données relative aux données Feader HSIGC hors Corse nécessaires à l'établissement du rapport annuel de performance (RAP) alimentant Synapse qui est le Système d'information de la performance de la PAC
 - o d'une partie Cœur de métier de l'OP performance qui comprend le pilotage et la maîtrise d'ouvrage des produits et systèmes d'information suivants :
 - Logic destiné à la déclaration des dispositifs par les autorités de gestion, ainsi qu'à la gestion et le suivi des conventions de paiement des différents financeurs intervenant dans le financement des aides Feader HSIGC
 - Alpha destiné à la création et gestion des enveloppes budgétaires des différentes sources de financements Etat et Feader, et suivre ainsi la consommation budgétaire des crédits
 - Delta destiné à l'ordonnancement et au paiement des aides Feader HSIGC et nationales
- D'une ligne produit RDR en charge de l'intégration de la production et du testing pour la programmation 2023-2027
- D'un secteur outillage, sécurité et habilitation

La direction métier est composée :

- d'un Service métier en charge des Aides au Développement Rural et à la Pêche (SADRP), organisé en trois secteurs :
 - o le Secteur des Aides Nationales et FEADER qui met en œuvre des aides nationales et Feader non régionalisées et assure le suivi et l'accompagnement de la mise en œuvre des aides FEADER régionalisées

- le Secteur Contrôles Dossiers qui pilote les COP RDR3 et RDR4, les CCF et CSP RDR3, les contrôles dossiers dans le cadre des contrôles de la délégation RDR4, les CSP RDR4 (protection des troupeaux contre prédation et autres), produit un bilan des contrôles et apporte son expertise dans le cadre des audits externes
- le Secteur Activités Supports et Transverses qui assure une expertise juridique, le conventionnement (rédaction de modèles, avis métier, supervision), le suivi de l'activité des DR, l'accompagnement des DR (réponses aux questions, suivi des tickets incidents, formation), est chargé du contrôle interne et de la fraude.
- d'un Service Valorisation, Restitution et Performance (SVRP) organisé en trois secteurs :
 - le Secteur Financement de l'Agriculture et de la Pêche qui assure le suivi financier des crédits du Ministère de l'Agriculture, de la Souveraineté Alimentaire et de la Forêt et du FEADER toutes programmations confondues, la gestion des pré-débiteurs, ...
 - le Secteur Pilotage de la Performance qui supervise les données liées à la performance pour le Feader HSIHC hors Corse, garantit l'absence de doublons dans l'identification des individus via le RCIPAC précité (cf § DO) et administre le catalogue RefPAC (solution de partage des référentiels)
 - le Secteur de la Valorisation et des Restitutions Statistiques Contrôles qui réalise la valorisation de données sur les différentes programmations PAC, assure la MOA du portail valo et d'Osiris Contrôles et produit des restitutions statistiques notamment dans le cadre du bilan annuel des contrôles pour la Commission Européenne. C'est ce secteur qui est en charge des restitutions institutionnelles destinées au pilotage du Feader HSIHC hors Corse et de celles destinées au pilotage opérationnel.

La Mission de coordination des organismes payeurs des fonds européens (MCFA)

La Mission de coordination des organismes payeurs des fonds européens (MCFA) est l'organisme de coordination des 4 organismes payeurs français (ASP, FranceAgriMer, ODEADOM, ODARC), répondant à une exigence réglementaire européenne. La MCFA est attachée à l'ASP. Elle assure le reporting auprès de la Commission européenne des données financières de la PAC, pour le compte des 4 organismes payeurs. A ce titre, la MCFA est également chargée de produire et transmettre à la Commission le Rapport Annuel de Performance (RAP, cf. infra). Pour ce faire, la MCFA s'appuie sur la DP pour développer l'applicatif dédié à cette mission.

3. Contexte de la PAC

La PAC en France

Le Plan stratégique national français (PSN) de la Politique agricole commune (PAC 2023-2027), approuvé par la Commission européenne le 31 août 2022, est entré en vigueur au 1er janvier 2023. Le PSN de la France correspond à un document de planification stratégique, qui cherche à favoriser le développement d'un secteur agricole et sylvicole plus diversifié et résilient, et à accompagner la transition agroécologique engagée par les agriculteurs.

La PAC 2023-2027 repose sur un nouveau modèle de mise en œuvre qui consiste en une planification stratégique décrite dans un Plan stratégique national (PSN) établi pour 5 ans. Ce PSN est instauré par le Règlement (UE) n° 2021/2115 du Parlement européen et du Conseil du 2 décembre 2021 établissant des règles régissant l'aide aux plans stratégiques devant être établis par les États membres dans le cadre de la politique agricole commune et financés par le Fonds européen agricole de garantie (FEAGA) et par le Fonds européen agricole pour le développement rural (Feader).

Le ministère de l'Agriculture, de la Souveraineté alimentaire et de la Forêt est l'autorité de gestion nationale du Plan stratégique national (PSN). La mise en œuvre du Fonds européen agricole de garantie (FEAGA) est de la compétence de l'État. La mise en œuvre du Fonds européen agricole pour le développement rural (Feader) pour la programmation 2023-2027 est partagée entre l'État et les Régions.

Ce partage des compétences entre l'État et les Régions est défini par l'article 78 de la loi n°2014-58 du 27 janvier 2014 de modernisation de l'action publique territoriale et d'affirmation des métropoles modifié par l'ordonnance n° 2022-68 du 26 janvier 2022 relative à la gestion du Feader au titre de la programmation débutant en 2023. Les Régions, en tant qu'autorités de gestion régionales, sont responsables des interventions du Feader non liées à la surface tandis que l'État est responsable des interventions de nature surfacique et assimilées, tels que résumé ci-après :



Figure 1. Description des aides de la PAC par Fonds

Pour plus d'information : <https://agriculture.gouv.fr/pac-2023-2027-le-plan-strategique-national>

L'ASP, organisme payeur et organisme de coordination des organismes payeurs français

Premier payeur européen d'aides agricoles, l'ASP verse chaque année environ 8,5 milliards d'euros de la PAC.

L'ASP est le principal organisme payeur français désigné pour gérer, en majorité, les deux fonds agricoles européens de la Politique agricole commune (PAC) : le Fonds européen agricole de garantie (Feaga) et le Fonds européen agricole pour le développement rural (Feader). L'ASP en assure sa mise en œuvre en France, avec 3 autres organismes payeurs.

L'ASP assure, en tant qu'organisme payeur agréé par la Commission européenne, la gestion et le déploiement des deux fonds agricoles européens de la Politique agricole commune (PAC) :

- le premier pilier concerne le soutien des marchés et des revenus agricoles. Il est financé par le Feaga (Fonds européen agricole de garantie) ;
- le second pilier correspond à la politique de développement rural. Il est financé par le Feader (Fonds européen agricole pour le développement rural).

Dans le cadre de la PAC 2023-2027, l'ASP a décidé de bâtir un lac de données unique pour recueillir les données de tout type des quatre organismes payeurs (OP) français : ASP/DSDA, ASP/DDRP, FranceAgriMer (FAM) et Office du Développement agricole et rural de Corse (ODARC). L'objectif est d'améliorer et moderniser la valorisation des données de la PAC à des fins de pilotage, d'audibilité et de suivi-évaluation. **Ainsi, la plateforme du Lac de Données Agricoles (LDA) n'est pas une plateforme transactionnelle, et ne sert donc pas à instruire, à payer et à contrôler des aides à des bénéficiaires, comme le font habituellement les SI de gestion de l'ASP.**

Architecture des SI découlant de cette organisation

Pour la période 2023-2027, le passage de la PAC vers un pilotage axé sur la performance et les résultats a conduit la France à repenser l'ensemble des systèmes de gestion des aides des 2 piliers et à construire un outil de suivi de la performance à l'échelle du PSN, intitulé Synapse.

En effet, la mise en œuvre du cadre de performance de la PAC 2023-2027 implique la collecte, le stockage et la transmission de données de suivi et d'évaluation via un système d'information qui enregistre et conserve les informations essentielles sur la mise en œuvre du PSN. Il y a donc plusieurs sources de données à articuler, provenant de 3 périmètres différents :

- le Système Intégré de Gestion et de Contrôle (SIGC) opéré par la DSDA, qui s'appelle ISIS : FEAGA et FEADER surfacique et/ou animal, gérés par l'ASP/DSDA et l'ODARC ;
- le système de gestion des aides sectorielles (FEAGA, géré par FAM) ;
- les systèmes de gestion pour les interventions FEADER hors-SIGC opéré par l'ODARC sur son périmètre géographique
- les systèmes de gestion pour les interventions FEADER hors-SIGC opéré par la DDRP (pour les aides régionalisées et aides nationales de l'Etat – hors Corse).

L'ensemble de ces sources sont connectées au Lac de données agricoles, où est installé Synapse.

La répartition des responsabilités entre acteurs est synthétisée dans le schéma ci-après :

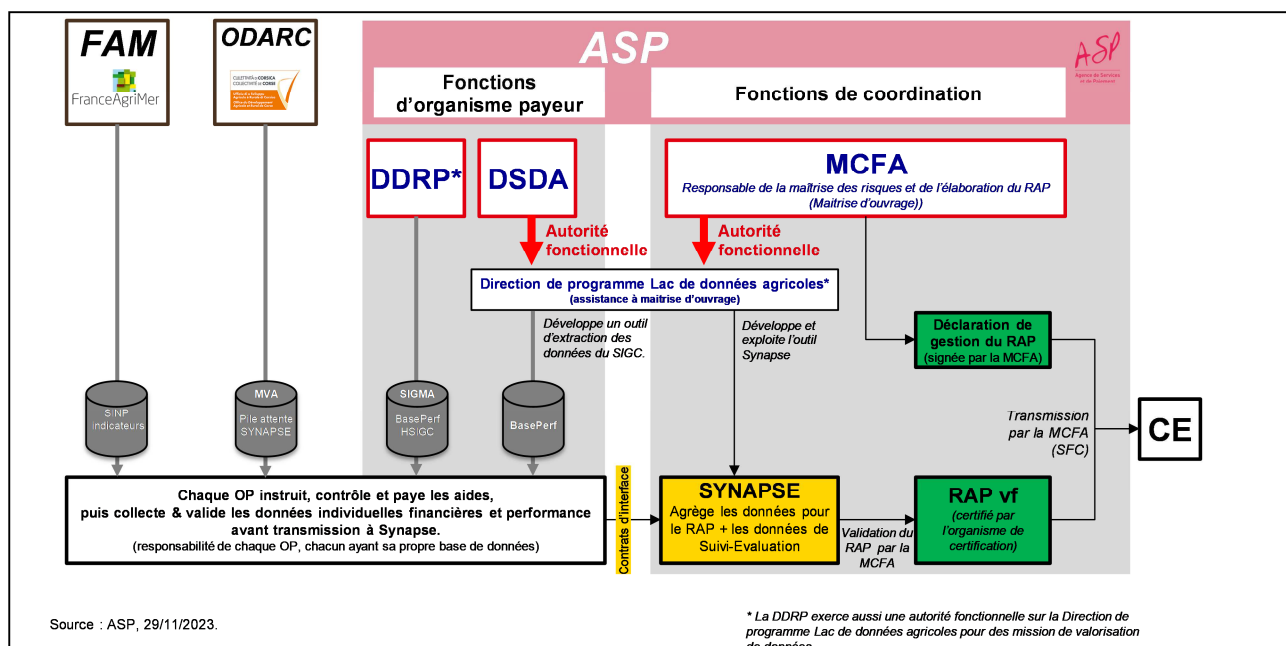


Figure 2. Schéma d'organisation interne à l'ASP

2 La description des applicatifs par grand chantier

Chantier n°1. Synapse : un applicatif dédié à la performance

Synapse permet de calculer automatiquement les indicateurs de la PAC nécessaires à la production du Rapport Annuel de Performance (RAP) – cf. ci-après.

Les enjeux de Synapse sont triples :

- **produire le Rapport Annuel de Performance dans le respect des délais réglementaires** (4 mois de délais, entre la fin de l'exercice financier au 16/10 n-1 et le 15/02 n), en coordination inter-OP et inter-AG ;
- **disposer d'un système unique regroupant l'ensemble des indicateurs du PSN (1er et 2nd piliers)**, pour suivre dans le temps leur évolution ainsi que leurs écarts aux ratios et jalons fixés dans le PSN ;
- **piloter la performance** pour in fine éviter tous risques de suspension et de réduction de paiements :
 - au niveau de chaque autorité de gestion ayant défini l'intervention : en donnant un accès continu aux données essentielles pour connaître (i) l'avancement des réalisations (engagements et paiements par intervention), (ii) l'efficacité (respect des montants unitaires planifiés) ;
 - au niveau du PSN l'efficacité de ses interventions : atteintes des jalons/cibles des résultats).

Le rapport annuel de performance contient les données financières, les indicateurs de réalisation et les indicateurs de résultat, qui sont agrégés à l'échelle nationale sans double compte, ainsi que les données relatives aux dépenses payées et certifiées. Il est transmis à la DG AGRI au plus tard le 15 février suivant l'exercice financier agricole. Sur la base de ce rapport, la DG AGRI procède annuellement au suivi de la performance du Plan stratégique de la PAC pour chaque exercice financier agricole.

Ce suivi de la performance est établi au travers de 2 mécanismes distincts :

- via **l'apurement annuel de la performance fondé sur les réalisations**, au moyen de 29 indicateurs de réalisation¹ quantifiant les actions mises en œuvre (respect d'un montant unitaire uniforme ou moyen) ; chaque intervention est reliée à un seul indicateur de réalisation ; le PSN approuvé par la DGAGRI contient environ 605 montants unitaires ;
- via **l'examen pluriannuel de la performance fondé sur les résultats**, au travers d'indicateurs de résultat mesurant les effets directs des réalisations. Chaque indicateur de résultat est doté d'une valeur cible à atteindre en fin de programmation ; les progrès vers ces cibles sont suivis annuellement (jalons annuels). La France a retenu 36 indicateurs de résultat (IR) dédiés au pilotage et à l'évaluation du PSN (rapportés chaque année). 19 Ri seront soumis à l'examen de la performance en 2025 (Exercice financier 2024) et en 2027 (EF 2026). Un écart important à la valeur du jalon en 2025 et en 2027 nécessitera un plan d'action, voire conduira à des suspensions de paiement en cas de sous-performance ou des réductions de paiement si les progrès au regard des jalons n'étaient pas suffisants. Chaque intervention est reliée à un ou à plusieurs indicateur(s) de résultat .

Par ailleurs, les indicateurs communs entre OP imposent une remontée des données brutes pour assurer un calcul sans double compte des indicateurs de réalisation et de résultat (pour lesquels l'absence de double-compte est requise).

Chaque Organisme payeur se dote d'un Système d'information permettant de stocker et d'envoyer les données financières et les données de la performance à Synapse. Ces échanges de données sont cadrés par des contrats d'interface qui en fixent les règles (déclenchement de l'émission, fréquence, description des données, structure, format, cardinalités, règles d'intégrations techniques et fonctionnelles, rapport d'erreurs, etc.).

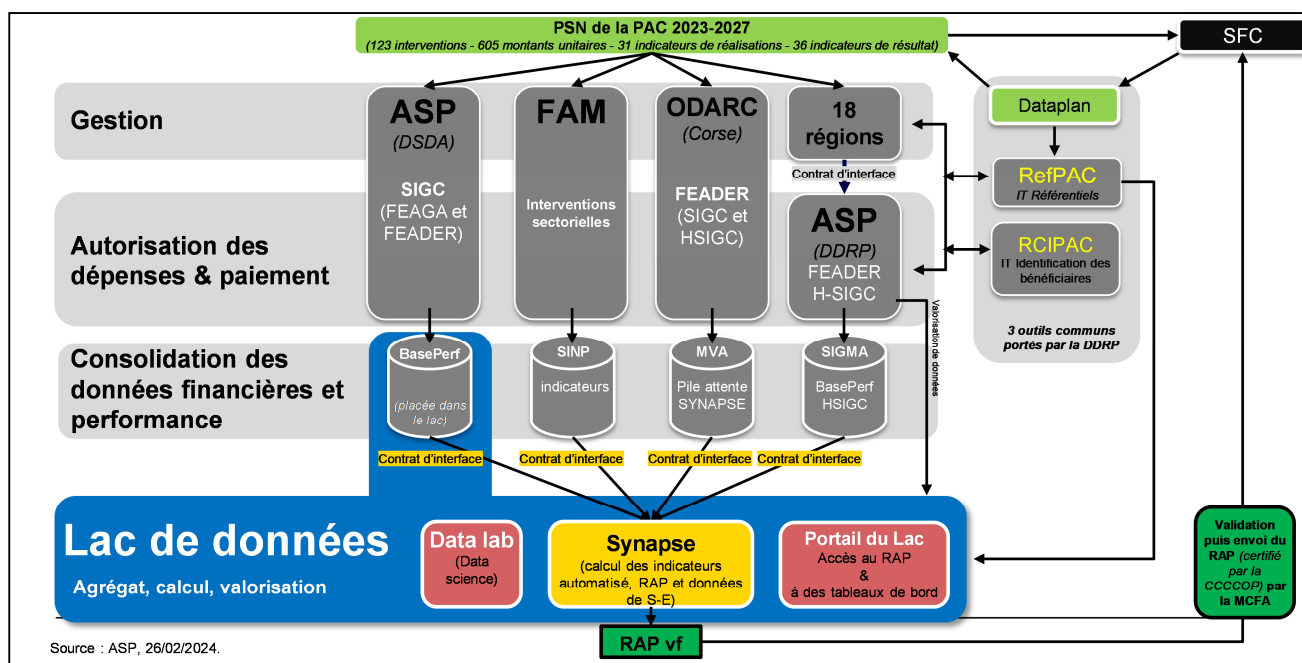


Figure 3. Système IT pour la gestion de la PAC & de la performance

Les premiers contrôles de la qualité des données sont effectués par les OP dans le cadre de leur mission de payeurs de fonds européens. La réconciliation des dépenses entre déclarations de dépense et données

¹ 2 autres indicateurs de réalisation sont utiles pour le suivi de la PAC (0.3 Bénéficiaires de la PAC et 0.34 Pratiques environnementales) et ne servent pas à l'apurement de la performance.

constitutives du RAP permet de veiller au bon fonctionnement du processus en identifiant et en corrigeant en amont de la finalisation du RAP les éventuelles anomalies.

Synapse procède également à des contrôles techniques et fonctionnels sur les données transmises par les OP. Il vérifie la présence et la cohérence des données de performance transmises à l'appui des paiements lors de l'intégration des données dans les bases de données. Il s'assure par cette vérification de la conformité des données reçues aux exigences réglementaires européennes d'une part et à la stratégie d'intervention définie par la France dans le PSN, d'autre part.

Synapse a été mis en production en décembre 2023, après avoir été homologué RGS obtenue en novembre 2023 et des tests d'arrimage avec l'ODARC, la DDRP et FAM. La MCFA a ainsi été en mesure de produire le premier RAP en s'appuyant sur Synapse. Le développement informatique de Synapse a continué en 2024, afin de mettre en production une nouvelle version de Synapse en octobre 2024 (pour établir le RAP de l'exercice financier 2024 en novembre 22024).

A la mise en production de Synapse, la DP Lac de données agricoles a validé un dossier d'exploitation regroupant l'ensemble des procédures permettant d'assurer le maintien en condition opérationnelle des services (c'est-à-dire l'ensemble des mesures prises pour garantir le bon fonctionnement du système d'information : correction des anomalies, optimisation des performances...), ainsi que l'ensemble de l'outillage d'administration de la plateforme du Lac de données agricoles. Une vue d'ensemble de Synapse est détaillée ci-après :

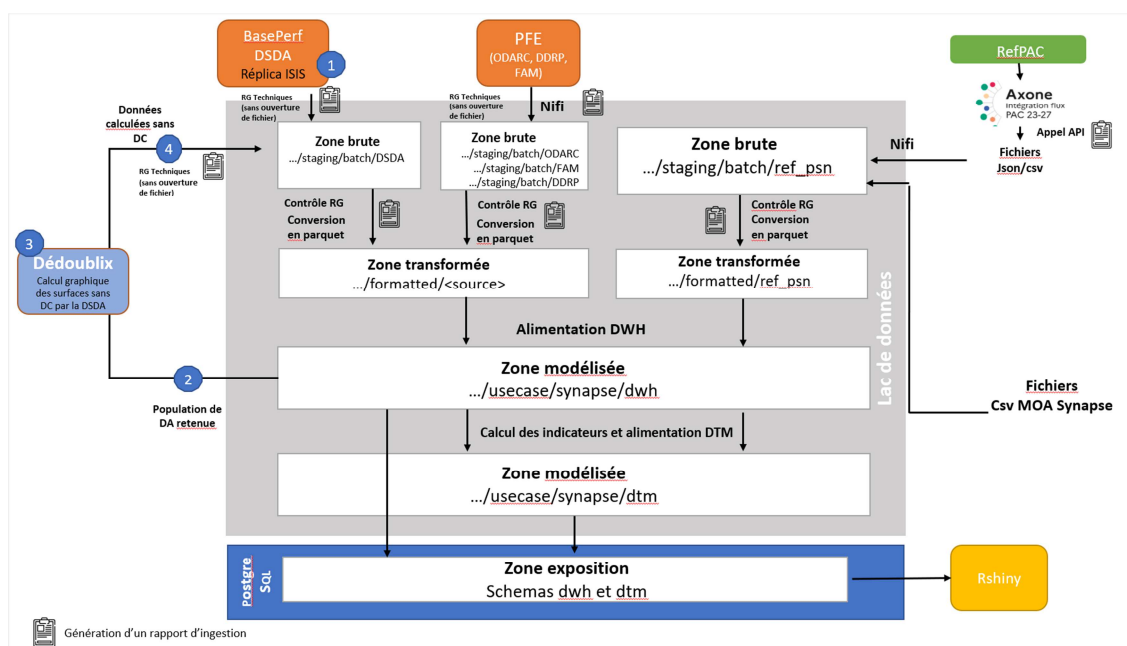


Figure 4. vue d'ensemble technique de SYNAPSE

Chantier n°2. Valorisation des données d'ISIS

Un déversement complet des données d'ISIS au travers du Réplica ISIS

Premier applicatif développé sur le LDA, le Réplica ISIS permet l'extraction des données du Replica ISIS Oracle vers l'espace de stockage brute du lac de données (HDFS). C'est le besoin premier de l'ASP, et qui a justifié la mise en place de cette plateforme Bigdata.

Toutes les tables et données du SI ISIS sont déversées dans le lac de données (HDFS).

Les règles de gestion technique de priorisation des tables à extraire, le filtre sur les schémas ainsi que la logique de déclenchement et d'ordonnancement sont formalisées et synthétisées ont été spécifiées. Le nombre d'exécutions en parallèle est paramétrable afin de lancer le plus de traitement en parallèle possible pour extraire toutes les données ISIS en temps optimisé. Certains schémas des bases sont exclus de l'ingestion.

Afin d'améliorer les performances et de gagner en temps global de déversement :

- un mécanisme full et delta est mis en place et est opérationnel depuis début 2023 (le Week end),
- un mode "delta" est exécuté à la demande du lundi au vendredi, sur un périmètre de tables plus restreint à déverser.

La chaîne de traitement permettant de déverser la totalité des tables de la base ISI2 s'effectue en plusieurs étapes :

- 1 Extraction des métadonnées : dans le cadre du mécanisme full, l'intégralité de la liste des tables du périmètre sont récupérés dans un fichier de métadonnées ;
- 2 Lancement de la chaîne d'extraction : les chaînes d'extractions sont exécutées au même moment. La plupart des traitements d'extraction sont réalisés sur le cluster Hadoop à l'exception des grandes tables contenant des champs géométriques qui dépassent en donnée les limites possibles des traitements sur Hadoop ;
- 3 Gestion des rejets / Traitement de fond de cuve : Le traitement des rejets permet de récupérer la liste de toutes les tables identifiées en erreur. Pour les tables géométriques, en cas d'erreur, ces tables sont considérées comme corrompues car incompatibles avec les librairies Arcgis, elles seront classifiées comme tables mixtes (géométrique et attributaire) lors du rejeu.
- 4 Tâche de fin de chaîne de traitement : permet de mettre à jour les informations temporelles pour pouvoir lancer la chaîne de traitements en delta.

Deux méthodes permettent de faire l'extraction :

- Mode Full : Sqoop pour toutes les tables sauf géométriques Arcgris, et Hadoop streaming et Python pour les tables géométriques ArcGIS ;
- Mode Delta il s'agit de déverser une partie des tables de la base et donc seulement les tables qui ont évolué par rapport à une date donnée, l'identification de ces tables est réalisée en se basant sur la table "dba_tab_modifications".

Le schéma fonctionnel du déversement est le suivant :

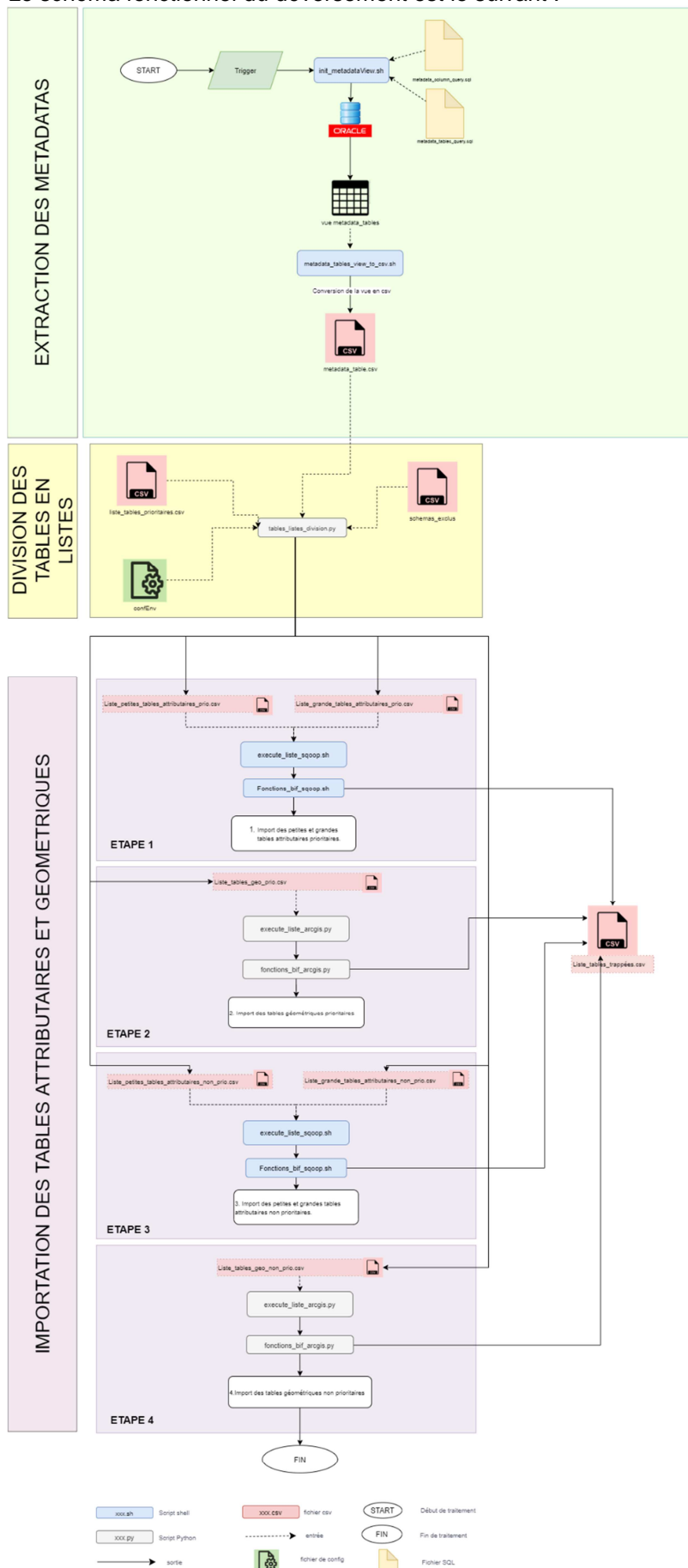


Figure 5. Schéma fonctionnel du déversement de BasePerf DSDA

En synthèse, les données déversées d'ISIS peuvent ensuite être valorisées au travers de 2 grandes approches :

- A. la BasePerf DSDA, accessible au travers de rapports Microstrategy,
- B. Dédoublis, pour le croisement de couches graphiques,
- C. des traitements de données développés avec Dataiku et accessible au travers de rapports Microstrategy.

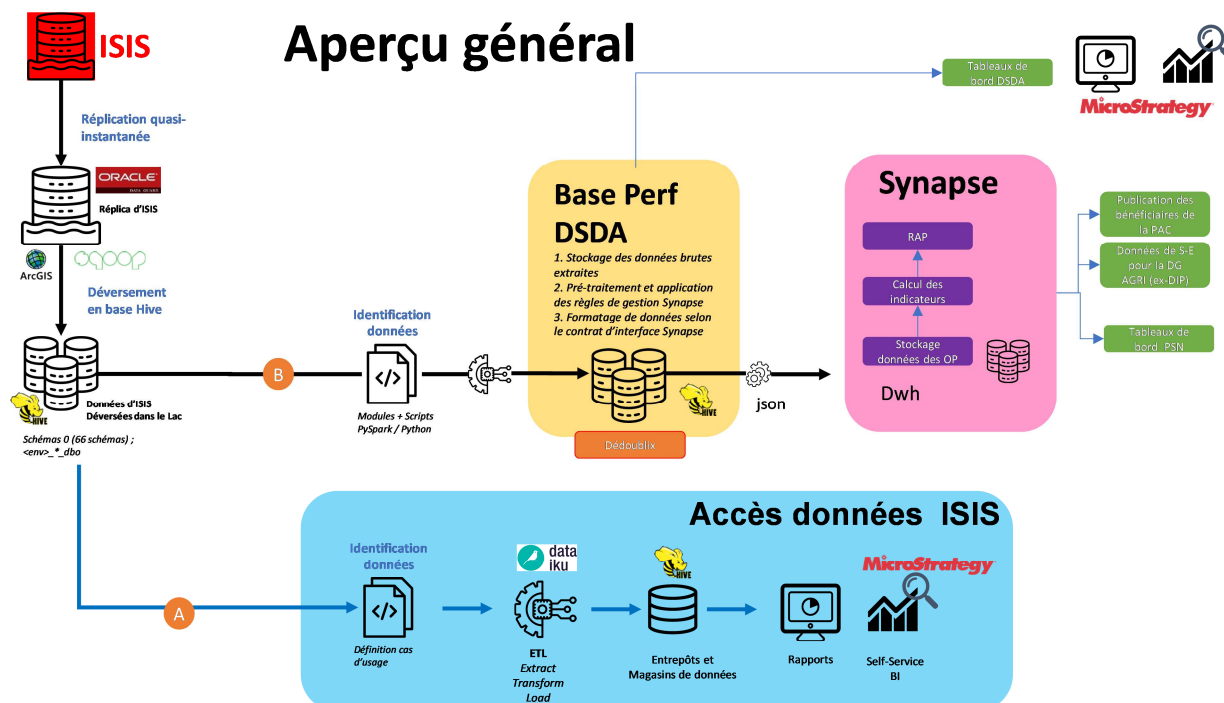


Figure 6. Articulation du Réplica avec la valorisation de données ISIS

B. BasePerf DSDA

La DSDA, comme Organisme payeur (OP) de la PAC 2023-2027, est tenue de restituer à la Commission européenne des données au sein du Rapport Annuel de Performance (RAP) de la PAC préparé par la MCFA et la DP Lac de données agricoles (via l'appliquatif Synapse), par exercice financier. Ces données permettront :

- le calcul des montants unitaires par intervention FEAGA et FEADER (sur la base des réalisations – apurement de la performance) ;
- le calcul des indicateurs de résultats surfaciques et non-surfaciques pour suivre les progrès de la PAC 2023-2027 (examen de la performance) ;
- le suivi évaluation de la PAC (soit intégrées dans le RAP ou soit transmises à la DG AGRI hors RAP).

La DSDA a ainsi identifié ces données dans les Bases ISIS et de facto déversée dans le LDA : la plupart seront extraites de son SI de gestion ISIS, d'autres seront déduites ou déterminées directement par la DSDA. Dès 2021, la DSDA a anticipé cet exercice de suivi de la performance de la PAC, en se dotant d'un Lac de données. Celui-ci favorise l'extraction de données du SI ISIS avec un système de réplique des bases Oracle d'ISIS fonctionnel. Les 11000 tables d'ISIS sont ainsi déversées tous les jours sur l'environnement de production du Lac de données agricoles.

A moyen terme, l'exploitation des données de Performance en autonomie par les agents de la DSDA permettra un meilleur suivi des aides de la PAC payées par la DSDA, en utilisant les données consolidées pour le RAP et en les agrémentant de données complémentaires à des fins de pilotage.

BasePerf est la solution de collecte des données ISIS pour les traiter et mettre à disposition les données permettant la construction du RAP et d'exploitation de données complémentaires.

Positionnée dans l'environnement de production du Lac de données agricoles, BasePerf DSDA a pour objectif de mettre à disposition des données "Performances PAC" en tant que "Source Référence". A ce titre, elle doit garantir la bonne collecte des données ISIS DSDA, la qualité des données qu'elle expose à la suite du traitement des données ISIS et la persistance de ces informations (historisation).

Elle est ainsi le point de passage obligé entre les exploitations des données Performances et la base de données ISIS, notamment en cas de demande de correction des données sources.

BasePerf met à disposition en priorité les données permettant de construire le RAP sous format JSON requis par Synapse (outil de consolidation, et de visualisation du RAP). Ce JSON présente les nouvelles DA traitées ou modifiées à la suite d'un rejet par Synapse. Dans un second temps, BasePerf devra prendre en compte l'utilisation de ses données pour d'autres exploitations (via MicroStrategy par exemple).

Le schéma d'architecture est le suivant :

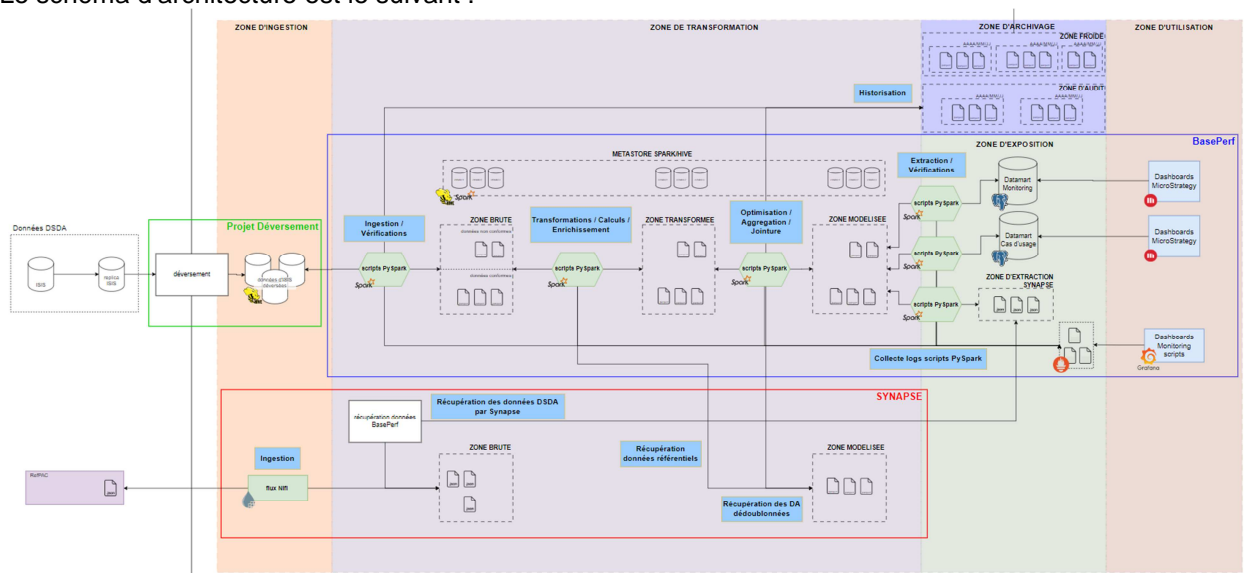


Figure 7. Schéma d'architecture de BasePerf DSDA

A. Dédoublx

Le rapport annuel de la performance de la PAC 2023 - 2027 porte sur 2 types d'indicateurs agrégés à l'échelle nationale, comprenant des données payées, certifiées et sans double compte :

- Indicateur de réalisation ;
- Indicateur de résultats.

Pour chaque type d'indicateur, deux méthodes de calcul peuvent être préconisées :

- Méthode alphanumérique : les données nécessaires au calcul d'indicateur sont des valeurs numériques,
- Méthode graphique ou surfaciques : les données nécessaires sont des couches graphiques présentes dans ISIS.

Un problème spécifique se pose lors du traitement des indicateurs dits "graphiques", qui exigent la superposition de couches graphiques. En effet, plusieurs interventions peuvent être associées à un indicateur donné. Ces indicateurs nécessitent des calculs sur la base de données géolocalisées, ce qui peut être complexifié par la juxtaposition d'interventions.

Etant donnée que plusieurs interventions sont cumulables à l'échelle de l'exploitation agricole et de la parcelle, il est actuellement impossible d'identifier les surfaces concernées par les interventions de l'indicateur sans double compte sur la base d'un calcul alphanumérique. Ainsi, pour ces indicateurs, des calculs sans double-compte de surface sont requis.

Au 1^{er} décembre 2024, les indicateurs concernés par un calcul graphique sont :

- Les indicateurs de résultat suivants : le R.31, R.33 (y compris ventilation SA/SF, étant entendu qu'à date le PSN v3.2 ne prévoit de rapporter que des surfaces agricoles) ;
- L'agrégat « surface » du O.14 pour le total et pour la ventilation par type d'intervention. Il n'y a en revanche pas de risque de doublons à l'échelle de l'intervention ou du code MU ou pour les agrégats non surfaciques (y compris le total et la ventilation par type d'intervention) ;
- L'indicateur de suivi O.34 pour le total et les ventilations (superficie agricole/superficie forestière).

Le calcul des indicateurs surfaciques « dits graphiques » nécessite l'utilisation d'un composant tiers permettant de superposer des couches graphiques afin de déterminer une surface sans double-compte : ce composant s'appelle Dédoublx. Il est composé entre autres d'une base PostgreSQL/PostGis.

Le process générique concernant le calcul des indicateurs graphiques se découpe en 10 étapes :

#	Etapes	Applicatifs
1	Envoi des DA dans le flux OP/Synapse	Baseperf DSDA
2	Contrôles des données à l'ingestion	Synapse
3	Stockage et modélisation des données dans l'entrepôt Synapse	Synapse
4	Sélection des demandes d'aide entrant dans le calcul des indicateurs	Synapse
5	Mise à disposition des DA sélectionnées	Synapse
6	Calcul graphique des surfaces sans double compte à l'échelle du bénéficiaire	Dédoublx
7	Envoi des valeurs calculées sans double compte	Dédoublx
8	Application de RG à l'ingestion puis alimentation d'une table d'audit spécifique	Synapse
9	Calcul de la valeur de l'indicateur toutes DA confondues	Synapse
10	Calcul de l'écart par rapport à la valeur prévisionnelle, le cas échéant	Synapse

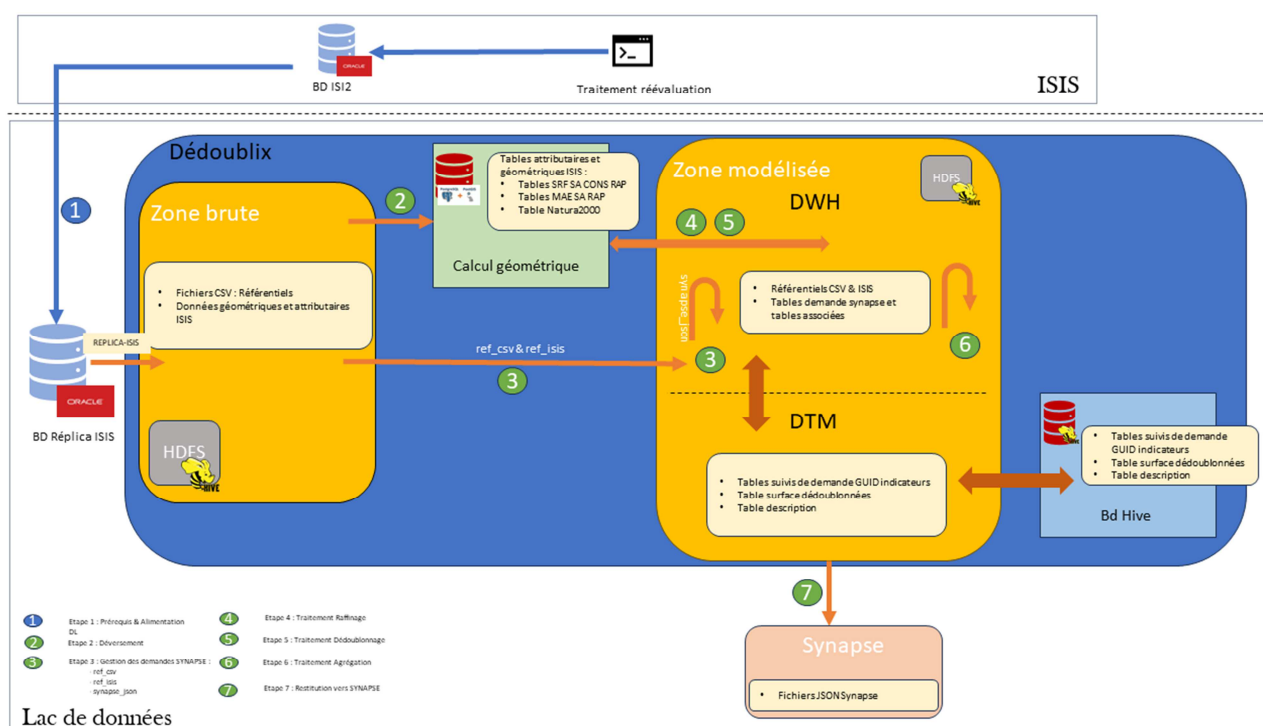


Figure 8. Architecture de Dédoublx

Le schéma de cheminement Base Perf DSDA / Synapse / Dédoublx / Synapse est le suivant :

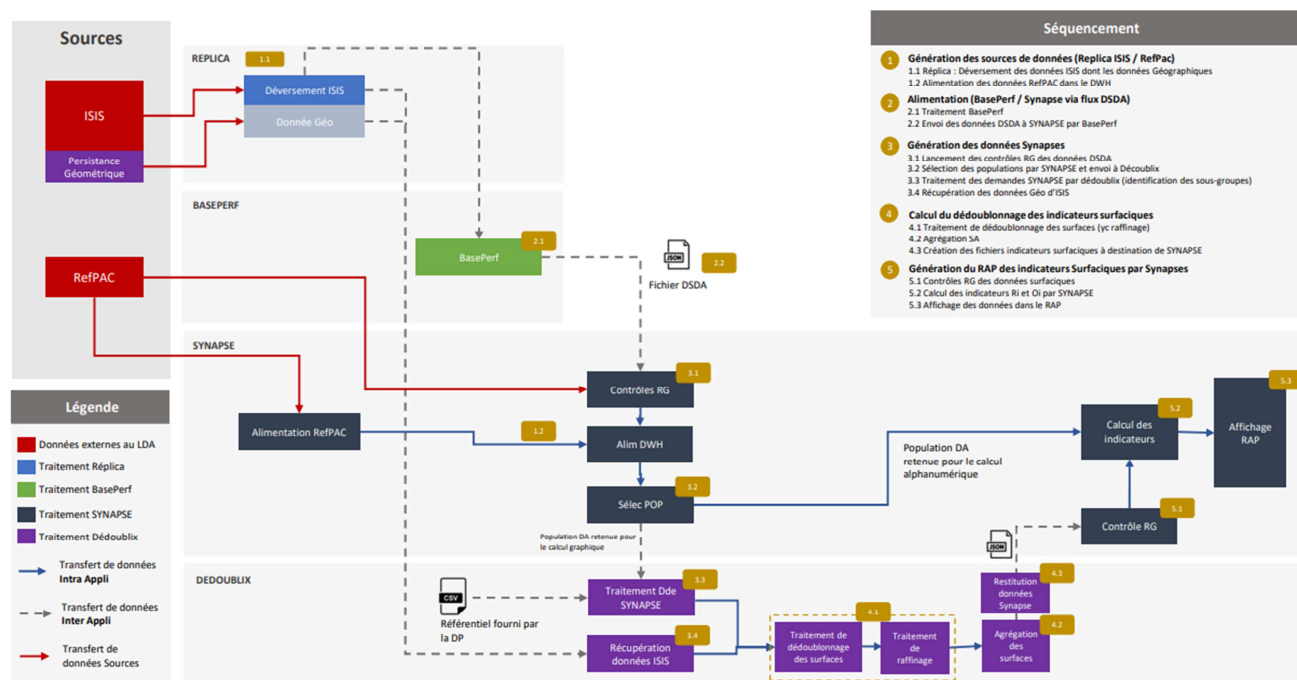


Figure 9. Articulation Base Perf DSDA / Synapse / Dédoublx / Synapse

B. Suivi Campagne PAC

Pour assurer un suivi de la campagne, la DSDA procède chaque année à la préparation de ces tableaux de bord au format Excel ; après validation de la Direction, ils sont diffusés auprès des DR et des DDT, principalement par transmission électronique (email). Actuellement, ils ne sont pas historisés (automatiquement), ni archivés dans l'application. Ces tableaux de bord ont plusieurs niveaux de complexité dans leur composition, et s'appuient principalement sur les tableaux de bord d'ISIS auxquels sont rajoutées certaines colonnes utiles pour les destinataires.

La DP LDA a mis en place l'ensemble de ces cas d'usage d'ici le démarrage de la Campagne 2024 (15 juin), sous forme de tableaux de bord accessibles via le Portail du Lac de données, en utilisant l'outil Microstrategy.

Les tableaux de bord standards ont été automatisés, à partir du Réplica ISIS, grâce à un applicatif développé grâce au logiciel Dataiku. Plusieurs tableaux de bord sont disponibles :

- ISIS_CU_1.1 - Etat des dossiers surfaces
- ISIS_CU_1.2 - Observations graphiques
- ISIS_CU_1.3 - SNA et ZDH
- ISIS_CU_1.4 - Ilots de référence
- ISIS_CU_1.5 - Droits de paiement de base
- ISIS_CU_1.6 - Instruction dispositif MAEC_BIO

Le schéma global du traitement est le suivant :

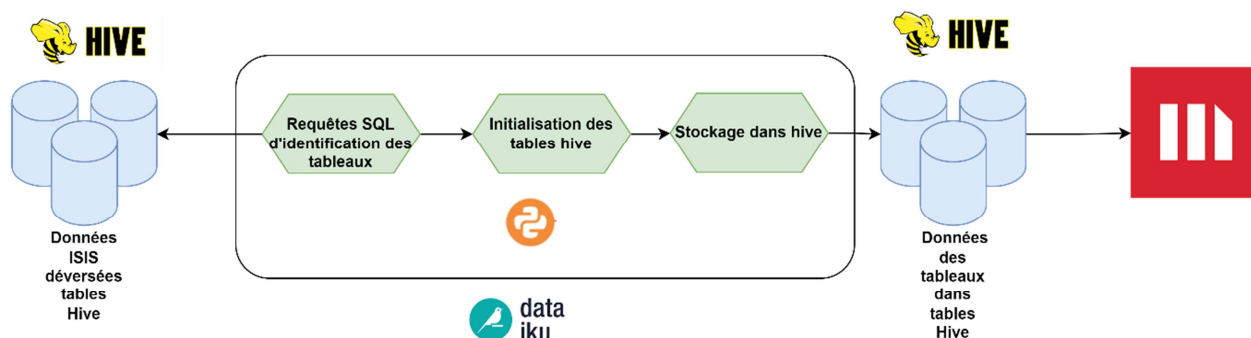


Figure 10. Traitement de Suivi Campagne PAC dans Dataiku

Chantier n°3. Valo RDR4 HSIGC FEADER

L'écosystème des SI du FEADER HSIGC est complètement modifié dans le cadre de la programmation de la PAC 2023-2027. L'organisation retenue par le Ministère de l'agriculture français est le suivant :

- des SI régionaux gérés par les Conseils régionaux qui, par délégation, instruisent et contrôlent les aides régionalisées du PSN ;
- des SI spécifiques gérés par la DDRP :
 - o sur le périmètre HSIGC : DELTA (ordonnancement), OMEGA (contrôles), ALPHA (enveloppes des aides nationales), SAFRAN (gestion des aides nationales et non régionalisées + Mayotte et Saint Martin), SIGMA (base performance HSIGC), LOGIC (lien conventions et dispositifs)
 - o sur les systèmes référentiels de la PAC : RefPAC, RCIPac (individus), Dataplan 2.0.

La DDRP a décidé de déverser une grande partie de ses données de gestion dans le Lac de données agricoles. La Valorisation des données FEADER HSIGC (RDR4) est ainsi utilisée pour

- aider au suivi budgétaire de la programmation,
- suivre l'avancement de l'état des dossiers au sein du processus d'engagement et de paiement
- favoriser l'identification des erreurs et des écarts,
- faciliter les contrôles croisés pour éviter le double financement.

Les enjeux sont :

- regrouper dans le LDA les données de pilotage des différents systèmes de gestion de la DDRP ;
- avoir un outil de valorisation des données de la PAC 23-27 pour les données gérées par la DDRP ;
- regrouper des données de pilotage de périmètres différents.

En synthèse, la VALO FEADER HSIGC RDR4 permet :

- de regrouper dans le Lac De Données l'intégralité des données financières et de pilotage des différents systèmes de gestion de l'ASP sur le périmètre FEADER (SIGC & HSIGC) et des aides nationales (hors PSN),
- d'avoir un outil de valorisation des données de pilotage de la PAC 2023-2027 pour les données gérées par l'ASP sur le périmètre FEADER (HSIGC) et des aides nationales (hors PSN).

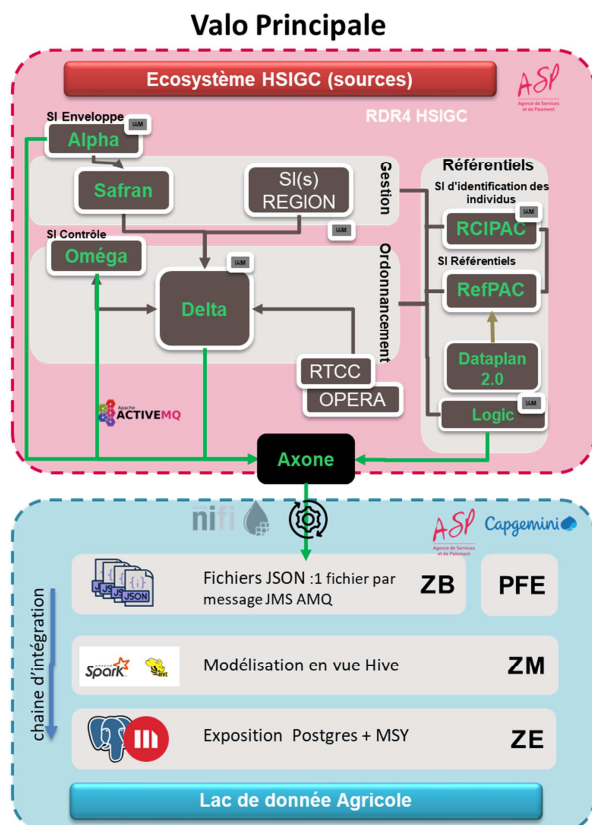


Figure 11. Présentation schématique de Valo FEADER HSIGC

Un document de Contrat d'Interface est établi entre VALO FEADER HSIGC et chaque SI amont.

Les 3 cas d'usages prioritaires qui couvrent le périmètre VALO sont :

- Le suivi des dossiers
- Le suivi financier de la programmation
- Le suivi des contrôles organisme payeur

D'autres cas d'usages ont été évoqués mais ne sont pas encore formalisés :

- Le suivi des demandes de paiements : du dépôt de la demande de paiement jusqu'au paiement,
- Le suivi des demandes d'aides : du dépôt de la demande d'aide jusqu'à la décision juridique,
- Le suivi des décisions juridiques,
- Le suivi des enveloppes spécifiques nationales : Fongibilités et gestion financières des enveloppes (Certaines restitutions seront portées par Alpha),
- le suivi des bénéficiaires,
- le suivi des financeurs,
- Le suivi des interventions / types d'interventions / dispositifs.

D'autre part, de nouveaux cas d'usages sont régulièrement découverts par l'ASP au fur et à mesure des demandes de valorisation des utilisateurs.

Ci-dessous la vue d'ensemble technique du projet VALO :

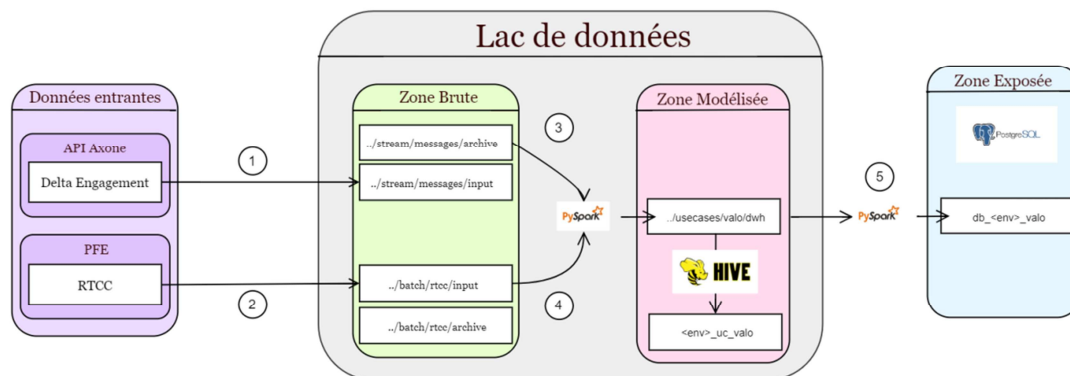


Figure 12. Vue d'ensemble technique de Valo FEADER HSGC

Chantier n°4. Portail utilisateurs du Lac de données agricoles

Le portail utilisateurs du LDA a été développé au travers d'une architecture Angular, en utilisant les principes et les composants du Système de Design de l'État (DSFR). La refonte en 2024 du portail développé en 2023 a permis d'améliorer l'expérience utilisateur, la performance et la maintenabilité de l'application.

Les fonctionnalités englobent les écrans suivants : page d'accueil (dashboard), page d'accessibilité, page des mentions légales, page des données personnelles, page de gestion du compte utilisateur, page de contact, page de plan du site et les pages d'erreur 404 et 500.

De plus, les éléments partagés tels que l'en-tête de page, le pied de page et le fil d'Ariane sont également pris en compte, ainsi que la configuration de Keycloak dans l'application Angular.

En outre, la mise à jour des applications et rapports RShiny (rap et ihm-monitoring) est incluse, ainsi que l'ajout d'un en-tête et d'un pied de page sur les tableaux de bord pour assurer une harmonisation avec le portail Angular.

Synthèse des applicatifs du LDA

Le LDA dispose, au 1^{er} décembre 2024, de plusieurs chaînes de traitement applicatives :

Traitement	Solution	Nombre de traitement VTOM	Périodicité	Nombre de flux en entrée	Nombre de tables	Nombre d'indicateurs / Flux en sortie
Déversement Replica ISIS	Développement spécifique	48 traitements élémentaires	Déversement FULL : Hebdomadaire Déversement DELTA : Quotidien	N/A (mécanisme de synchronisation par Oracle Dataguard- Hors Périmètre)	FULL : 8 600 tables Oracle DELTA : 107 tables	N/A
Synapse	Développement spécifique	24 traitements élémentaires	Ingestion quotidienne Calcul des indicateurs mensuels	4 json	dwh: 87 dtm: 101	70 indicateurs dans le RAP + données financières
Dédoublx	Développement spécifique	8 traitements élémentaires	Mensuel	3 json	dwh: 22 dtm: 8	3 flux JSON
BasePerf DSDA	Développement spécifique	20 traitements élémentaires	Mensuel	93 tables	baseperf_zb: 76 baseperf_zt: 31 baseperf_zm: 8	1 flux json contenant 1,3M d'enregistrements
Suivi Campagne PAC	Application Dataiku	1 traitement	Hebdomadaire	15 tables	15 tables	6 rapports
Valo Feader HSIGC (RDR4)	Développement spécifique	22 traitements élémentaires	Quotidien (Streaming)	8 jsons	dwh: 91	50 indicateurs
Portail Utilisateurs	Développement spécifique	0	Inscription des utilisateurs au fil de l'eau	200 utilisateurs à date, 400 à terme	N/A	N/A

Tous ces applicatifs sont en production ; des compléments applicatifs sont attendus au T1 2025 pour Synapse, BasePerf et Valo Feader HSIGC.

3 Dates-clés du LDA

Récapitulatif

Comme décrit précédemment, une année est rythmée par de nombreuses dates-clés, jalonnée par des échéances cycliques qu'il est indispensable d'anticiper et de planifier, et soumise à des aléas de toute nature nécessitant une forte réactivité de la part de toutes les équipes.

Traitements	Durant l'exercice financier agricole (16/10 n-1 au 15/10 n)	Echéance réglementaires
Déversement Replica ISIS	Hebdomadaire	<ul style="list-style-type: none">Cf BasePerf DSDA
Synapse	Traitement d'ingestion en continu Calcul des indicateurs RAP tous les 15 du mois	<ul style="list-style-type: none">Envoi d'une v0 vers le 15 novembrePartage d'une v1 à l'organisme de certification le 15 décembreEnvoi du RAP à la DG AGRI le 15 févrierTransmission des données individuelles de paiement (données du RAP désagrégées) le 30 avrilPublication des données individuelles sur internet avant le 30 mai
Dédoublis	Traitement lancé en mars, juin et septembre	<ul style="list-style-type: none">Transmission des indicateurs d'un exercice financier avant le 06 novembre
BasePerfDSDA	Traitement mensuel	<ul style="list-style-type: none">Transmission des données d'un exercice financier avant le 06 novembre
Suivi Campagne PAC	Traitement hebdomadaire	<ul style="list-style-type: none">Suivi de l'activité métier entre Avril et décembre
Valo Feader HSI GC (RDR4)	Traitement hebdomadaire	<ul style="list-style-type: none">Aucune (à date)Transmission des déclarations de dépenses tous les trimestres (à terme)

Processus de traitement des évolutions fonctionnelles

Pour chaque évolution réglementaire ou métier, la DP LDA ou ses autorités fonctionnelles rédigent une expression de besoins (EB) décrivant le contexte de l'évolution et la mise en œuvre souhaitée (création, modification ou suppression d'un écran ou d'une règle de gestion, d'un indicateur, ingestion d'une nouvelle source de données). Pour cela, elle dispose du règlement européen et de la déclinaison française à travers de notes et d'arbitrages établies par les Autorités fonctionnelles (DSDA, DDRP et MCFA).

Organisation de l'assistance aux utilisateurs

L'assistance aux utilisateurs est actuellement organisée en 3 niveaux :

- l'accompagnement des utilisateurs par la DP LDA (au travers d'un portail JIRA mis en place par la DNSI de l'ASP) ;
- l'analyse des anomalies et incidents « techniques » ou l'analyse réglementaire d'une demande par la DP LDA ;
- l'expertise technique et la correction des anomalies et incidents « techniques » par la MOE.

4 Les utilisateurs du Lac

Description des utilisateurs

Les utilisateurs du LDA proviennent de sphères différentes :

- différentes entités organisationnelles : directions de l'ASP (DSDA, DDRP etc..), MASA, prestataires techniques, OP, conseils régionaux et partenaires externes comme l'INRAE ;
- différents périmètres géographiques : national, un ou plusieurs départements ;
- différents rôles : management, experts techniques, consommateurs de tableaux de bord, auditeur.

Ainsi, une **matrice des profils d'utilisateurs** a été définie ainsi pour la période 2021-2025 :

Profils	Fonctions	Institutions	Niveaux géographiques	Environnement		
				MOE (DEV/INT/RECETTE)	DATALAB PREPROD	PROD
Administrateur Lac de données	Gère tout ou partie des outils et peut déployer en PROD Créer les groupes de sécurité	MOE	National	X	X	X
Développeur MOE	Développer, tester sur tous les environnements	MOE	National	X		
Administrateur des habilitations (Direction de la DP)	Octroyer et gérer les habilitations, au regard de la liste des profils définis et des groupes de sécurité	ASP	National	X	X	X
Architecte	Vérifie la bonne administration de la plateforme et pose les choix stratégiques avec la MOE	ASP	National	X	X	X
Développeurs ASP / Data scientist "expert" / Responsable applicatif	Développer avec des outils (Dataiku Jupyter, ...) Utiliser les outils de Machine Learning / deep learning (Dataiku +...)	ASP	National		X	
	Concevoir et consulter des tableaux de bord				X	X
Métiers requêteur / Chargé de mission Valo (requêteurs)	Lire et faire du requêtage sur Datamarts (peut être décliné par type d'environnement) Concevoir des tableaux de bord	ASP	National Régional Départemental		X	X
Métiers (Consommateurs de tableaux de bord)	Consulter des tableaux de bord	ASP MAA Conseils régionaux OP	National Régional Départemental			X
Auditeur	Avoir accès en particulier aux données de traçage des données liés aux événements métiers (ex: états de dossiers, cheminement de la donnée, calcul des indicateurs)	ASP C3OP	National			X

Cette matrice sera revue dans le cadre des nouveaux marchés Lot 4 et LDA.2026.

A noter que les utilisateurs ASP ont accès à tous les environnements (recette, pré-production et production - sauf développement et intégration) sans contrainte technique alors que les utilisateurs externes ne peuvent accéder qu'à l'environnement de production.

Lorsqu'une nouvelle personne rejoint la DP LDA, celle-ci lui configure les droits d'accès en fonction de ses missions, au travers de groupes de sécurité.

A terme, pour un individu externe, ce dernier devra remplir un formulaire Excel, disponible sur le portail du LDA (environnement de production). La DP réceptionne cette demande, valide ou non les accès et enfin ajoute l'utilisateur dans les groupes de sécurité. L'individu n'a plus qu'à créer son mot de passe.

Lors du départ d'une personne d'un projet lié au LDA, tous ses accès sont supprimés par la DP.

Groupes de sécurité

Les groupes de sécurité peuvent être déclinés par typologie :

- Des groupes permettant l'accès, outil par outil (Hue, Dataiku, Microstrategy, VTOM, ...) ;
- Des groupes permettant l'accès à des rapports exposés, dont le rapport annuel de performance (RAP) ;
- des groupes permettant de rattacher les utilisateurs à des organisations ;
- des groupes permettant de rattacher les utilisateurs à des périmètres géographiques ;
- des groupes permettant d'habilitier les utilisateurs à des données (PAC, FEAGA, FEADER) au travers de droits sur des objets techniques de stockages des données (fichiers parquet, bases Hive, bases PostgreSQL).

Pour chaque environnement, un portail d'administration permet de gérer la création des utilisateurs et leur affiliation à des groupes de sécurité ; ils sont gérés par la DP LDA. Ce portail est comme celui des utilisateurs, **indépendant d'ISIS**.

Une revue annuelle des habilitations au portail du LDA est réalisée en début de chaque année calendaire.

4. Fonctionnement du Lac de données dans le cadre du marché subséquent ISIS n°2 (2021-2024)

La comitologie actuelle du LDA (2021-2024)

La comitologie LDA se décompose comme suit :

- **un comité contractuel trimestriel** (1 heure 30) réunissant le pilotage MOE, la DP LDA et la DSDA/DDO pour dresser un état des lieux des commandes, des livrables et des facturations et pour entériner la valorisation des indicateurs de qualité et des pénalités qui en découlent ;
- **un comité de projet mensuel** (1 heure 30) réunissant le pilotage MOE et la DSDA pour faire le point sur les activités des différentes équipes MOE (conception, développement, production et exploitation) et sur la planification des évolutions et pour arbitrer si besoin sur les priorités ;
- **un comité technique mensuel** commun avec ISIS (1 heure) qui a pour but de faire un point de situation relatif à l'infrastructure, sur les architectures, les chantiers techniques (migration de logiciels de base, ...), l'exploitation applicative, la supervision, l'hébergement ;
- **des comités de suivi hebdomadaire** (1 heure) réunissant les fonctionnels MOE et la DP LDA pour assurer un suivi opérationnel sur un périmètre ciblé et sur une période spécifique.

Le dimensionnement des équipes MOE actuelles

Les équipes MOE dédiées au LDA sont les suivantes :

Prestations	Dimensionnement actuel	Marchés cibles concernés
Maintien en condition opérationnelle (hébergement, supervision, exploitation technique, conduite de production)	4 à 6 ETP	Lot 4 et marché LDA
Pilotage	2 à 3 ETP	Lot 4 et marché LDA
Activité développements	10 à 15 ETP	Marché LDA

5. Gouvernance de la donnée du LDA

Acteurs et instances

Les principaux acteurs et instances de la gouvernance des données du LDA sont :

Rôles Data	Missions
Directeur de la donnée	<ul style="list-style-type: none"> Pilote la stratégie de gestion des données de la PAC à des fins de suivi, évaluation et de valorisation, Met en place une stratégie Big Data dans le cadre du marché ISIS n°2, Organise la collecte des données de la PAC (internes et externes) de l'ASP, Supervise l'accès aux données par les différents services de l'ASP et des structures externes, Suit les évolutions technologiques des SI des Directions métiers impactant le lac.
Gouverneur de la donnée	<ul style="list-style-type: none"> Est le contact privilégié du DPD (délégué à la protection des données ou DPO en anglais) de l'ASP, Pilote la politique de gouvernance : <ul style="list-style-type: none"> Anime le comité des cas d'usage, Veille à la bonne compréhension des enjeux autour de l'utilisation des données (données sensibles, définitions, outils à utiliser), Assure l'intégrité et la qualité des données et fournit les résultats au directeur de la donnée, Dans le futur, partage les bonnes pratiques avec les instances globales de gouvernance de la donnée de l'ASP, Définit les processus liés à la gouvernance des données, à la fois d'un point de vue organisationnel et technique, et s'assure du bon respect de ceux-ci.
Architecte	<ul style="list-style-type: none"> Garantit le respect des normes et standards du cadre général d'architecture de l'ASP (quand le cadre s'applique) et du Dossier d'Architecture Technique du Lac de données, Conseille le gouverneur de la donnée sur la sécurisation technique des données, Valide la bonne application des processus de sécurisation technique auprès des référénts techniques, Informe le gouverneur de la donnée sur les évolutions techniques des projets impactant la gouvernance des données.
Référent métier	<ul style="list-style-type: none"> Connait parfaitement les données métier de son périmètre, Définit, sur son périmètre : <ul style="list-style-type: none"> Les termes métiers, Les règles de qualité de données, Identifie sur son périmètre les données sensibles ou personnelles exploitées et les cas d'usage associés, avec l'aide des Data Scientists, du référent fonctionnel, des référents métiers des projets amont et aval. En cas de doute, sollicite immédiatement le gouverneur de la donnée. Valide et archive les règles de correction des données et en fait un rapport au gouverneur de la donnée, Confirme la complétude métrologique mise en place par le référent fonctionnel.
Référent technique (ou Référent IT)	<ul style="list-style-type: none"> Connait parfaitement les traitements techniques de son périmètre, Définit de manière précise sur son périmètre, avec l'aide du référent métier et dans le respect du cadre général d'architecture : <ul style="list-style-type: none"> Les données à intégrer, d'un point de vue technique, Les flux techniques de données, La modélisation des données pour répondre aux cas d'usage ; Demande conseil à l'architecte en cas de nécessité, Applique ou fait appliquer les processus de sécurisation technique des données sur son périmètre, en assure la documentation. Assiste le référent fonctionnel dans ses tâches d'analyse en cas de besoin exprimé.
Référent fonctionnel (parfois appelé "Intendant de données" ou "Data Steward")	<ul style="list-style-type: none"> En coordination avec le référent métier, <ul style="list-style-type: none"> Fait le lien entre champs techniques et termes métiers, Met en place les rapports de mesure de qualité, Implémente ou fait implémenter des contrôles qualité de données, Analyse les remontées d'inconsistance de données faites par les Data Scientists à l'aide de la métrologie mise en place et en fait un rapport aux référents métiers/techniques des projets concernés.

Rôles Data	Missions
Data Scientist	<ul style="list-style-type: none"> Implémente des algorithmes de Machine Learning et d'analyse de données, Construit des tableaux de bord aidant à la décision et au monitoring des politiques publiques agricoles, Informe <u>spontanément</u> les référénts métier projet des données sensibles ou personnelles nécessaires pour ces algorithmes et le traitement associé <u>avant</u> leur exploitation, Remonte <u>spontanément</u> les inconsistances de données constatées dans les cas d'usages exploités aux référénts fonctionnels projet.
Data Analyst	<ul style="list-style-type: none"> Analyse les données existantes et les rend compréhensibles par les consommateurs de données, Construit des tableaux de bord aidant à la décision et au monitoring des politiques publiques agricoles, Présente les résultats de ses analyses aux équipes concernées pour les aider à la prise de décisions stratégiques, notamment avec l'utilisation d'outils de Data Visualisation, Remonte <u>spontanément</u> les inconsistances de données constatées dans les cas d'usages exploités aux référénts fonctionnels projet.
Comité des cas d'usage	<ul style="list-style-type: none"> Orienté vers l'offre de service la plus pertinente Centralise et qualifie les C.U <u>Exemples</u> : valider les cas d'usage, élaborer le processus de remédiation de la qualité des données... <u>Participants</u> : direction de programme, direction métier, référent métier, référent fonctionnel et tout autre partie prenante au sujet.

Le Diagramme organisationnel des acteurs et instances de la gouvernance des données résumé ci-après l'organisation en place :

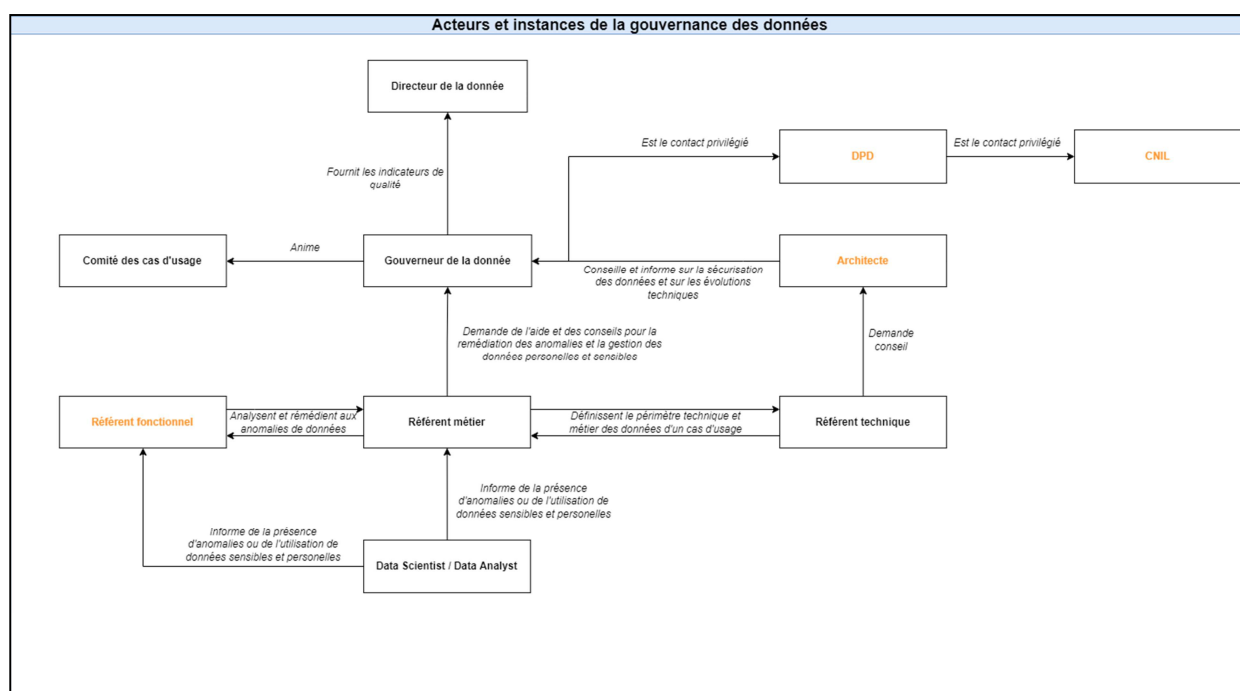


Figure 13. Diagramme organisationnel des acteurs et instances de la gouvernance des données

Les rôles en orange sont facultatifs ; le DPD et la CNIL sont consultés uniquement si des données sensibles et personnelles sont présentes. Le référent fonctionnel peut être aussi le référent métier ou technique, en fonction des spécificités du cas d'usage. Enfin, l'architecte aide et conseille sur demande uniquement du gouverneur de la donnée ou des référents.

Glossaire métier

Un **glossaire métier** est un outil clé de la gestion de la qualité des données. Il correspond à un dictionnaire de **termes métiers** utilisés dans un cas d'usage ou plus généralement au sein du LDA.

L'objectif est de s'assurer que tous les utilisateurs possèdent le même langage en définissant de manière concise et cohérente les concepts métiers.

Les bénéfices du glossaire métier sont :

- **un partage de vocabulaire / de sémantique commune**, qui permet une meilleure communication et compréhension ;
- **un accès facilité à la connaissance** via des écrans accessibles en navigateur Web voire directement dans l'outil de reporting (et non plus de manière hétérogène et dispersée dans des documents de spécifications) ;
- **une meilleure confiance dans les données utilisées**, en réduisant le risque de mauvaise compréhension des données ;
- **un lien entre glossaire métier et le catalogue de données** ;
- **un point d'entrée de base pour que les utilisateurs métiers prennent leur responsabilité dans la gouvernance des données.**

Périmètre des données à dictionnariser

Les termes métier qui doivent être inscrits dans le glossaire métier sont les **termes qui présentent un intérêt dans l'exécution d'un processus métier**. Les termes qui sont indispensables à la bonne exécution du processus métier sont des **données critiques**.

Il ne s'agit pas de documenter tout le savoir de l'ASP mais de procéder de manière méthodique pour supporter les usages métiers. De plus, ces informations sont valorisées manuellement, il faut donc trouver le bon équilibre.

Si un terme métier représente un indicateur, on renseigne alors dans le glossaire métier les définitions du terme métier et des termes métiers constitutifs, ainsi que la règle de calcul métier de l'indicateur.

Dans le cadre du LDA, le glossaire métier est déployé par cas d'usage.

Outillage

Pour le LDA, l'outillage sélectionné pour le glossaire métier est ATLAS. Les principales notions manipulées sont les suivantes :

Notion	Signification
Terme	Le terme qui fait l'objet d'une définition.
Glossaire	Regroupement de termes métiers. 1 seul niveau. Un glossaire correspond à un cas d'usage.
Catégorie	Sous-ensemble d'un glossaire. Un glossaire peut avoir n catégories. Une catégorie peut avoir m sous-catégories. Un terme appartient à un glossaire et peut appartenir à plusieurs catégories.
Classification	Permet de classer un terme métier à l'aide d'un "Tag" Exemple : "Donnée Personnelle". Utile pour les recherches et pour définir des politiques d'accès aux données basées sur ces tags.
Termes liés	Liens entre termes métiers. Exemples de types de relations standard : "SeeAlso", "Synonyms", "Antonyms", "PreferredToTerms", ...
Entité	Donnée instanciée dans le système d'information (ex : la table HIVE "Customer_Hive_table", le champ "Prénom" de cette même table).

La création d'un glossaire ou d'un terme intervient généralement lors de l'**instruction d'un nouveau cas d'usage**, en mode projet. C'est l'**équipe projet du cas d'usage**, incluant les référents métiers, qui pilote l'élaboration des nouveaux termes métiers. Elle doit s'assurer que les données du cas d'usage ne soient pas présentes dans le LDA. Si elles y sont, il faut qu'elle se rapproche du référent métier voire du gouverneur de la donnée afin de s'assurer que la concordance soit de 100%.

Si les **données sont différentes**, il sera nécessaire d'établir un plan d'action : modifier les contrats d'interface, écrire une définition légèrement différente...

Si la **concordance est de 100%**, la donnée peut être réutilisée et il faudra simplement accorder l'accès à cette donnée.

Périmètre des données à cataloguer

Les données à cataloguer (tables Hive, format parquet et tables Postgre) sont :

- données brutes,
- données transformées,
- données modélisées,
- données précalculées, quand elles ne sont pas présentes en zone modélisée.

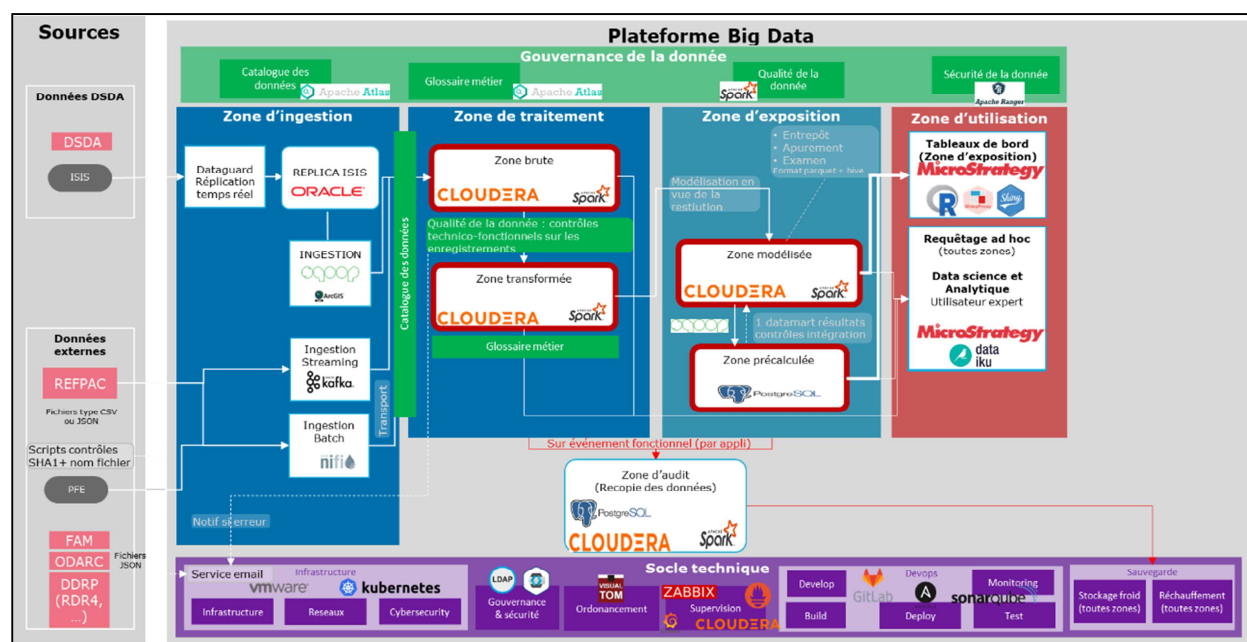


Figure 14. Architecture du LDA et catalogue de données

Les droits d'accès aux données et le portail

Définition et objectif

La **gestion de l'accès aux données** est un enjeu important pour le LDA, de par le fait de la nature sensible des données qu'il héberge. C'est pourquoi a été réfléchi une stratégie d'accès au LDA, dans le respect du cadre de sécurité de l'ASP (SMSI et PSSI) et « du droit à en connaître ».

Elle se base sur l'**authentification**, qui est le mécanisme qui permet à un utilisateur de faire reconnaître son identité par le système auquel il essaye d'accéder, plus simplement de se « logguer ». Une fois reconnu par le système, l'utilisateur est « connecté » au système.

Enfin, les **droits d'accès aux données** (habilitation) permettent à un utilisateur authentifié d'accéder aux données pour lesquelles il est autorisé.

L'objectif de ce chapitre est de décrire la gestion des droits d'accès aux données pour le LDA.

Authentification

L'authentification se fait par un portail des utilisateurs du LDA (un par environnement, soit 3 au total). Il est développé en RShiny et est **indépendant du portail TELEPAC/ISIS**.

L'accès à la plateforme sur l'environnement de production est accessible par le Réseau Internet de l'Etat (RIE) via l'adresse : xxx

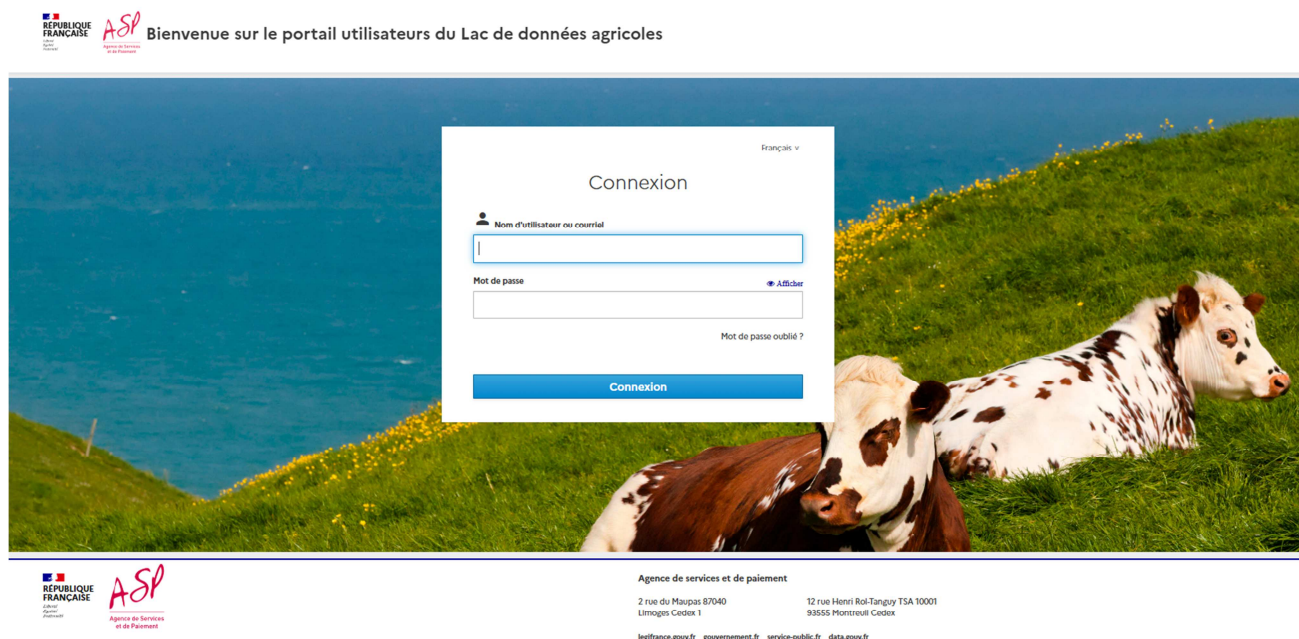


Figure 15. Page de connexion au Portail Utilisateurs du LDA

Le principe de la solution d'authentification est schématisé ci-dessous :

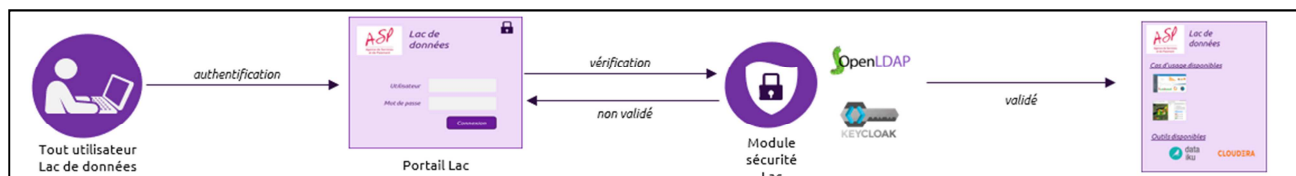


Figure 16. Mécanisme d'authentification au portail

Le LDA a son propre annuaire OpenLDAP combiné avec Keycloak, qui est un logiciel à code source ouvert permettant d'instaurer une méthode d'authentification unique au travers de la gestion par identité et par accès.

Cette solution permet une gestion centralisée des droits d'accès.

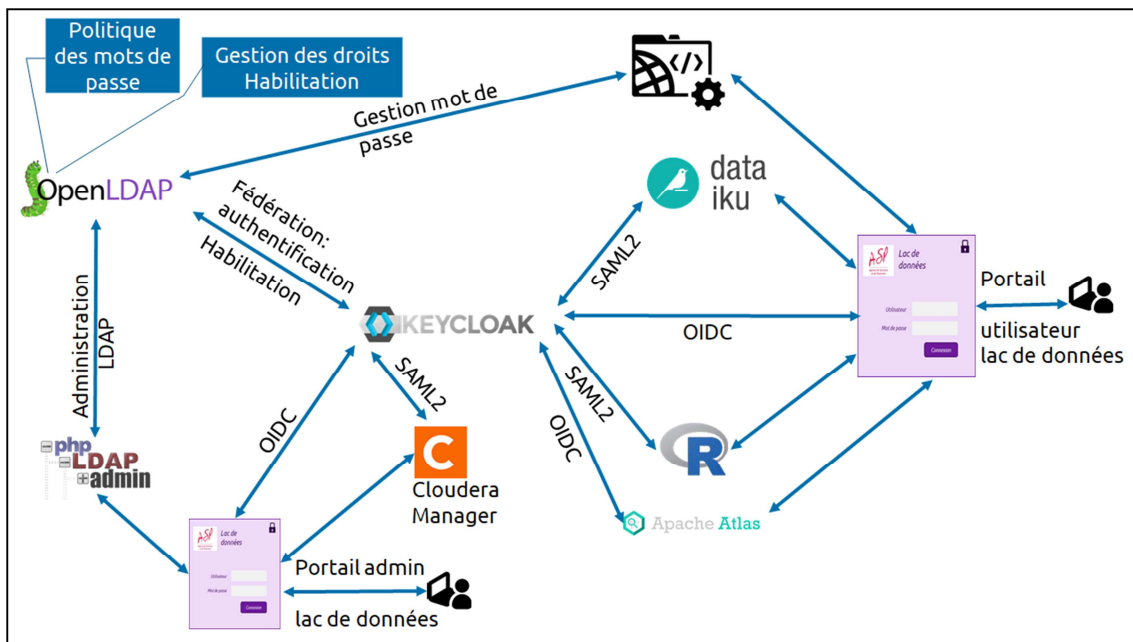


Figure 17. Schéma de la gestion centralisée des droits

Rôles et groupes

Il existe deux modes implémentables d'accès aux données :

- **RBAC** : Role Based Access Control
 - Accès basé sur le rôle de l'utilisateur. « Tel rôle est autorisé à accéder à telle ressource ».
- **ABAC** : Attribute Based Access Control
 - Accès basé en fonction des métadonnées. Exemple : « ne pas accéder aux données tagguées données personnelles »

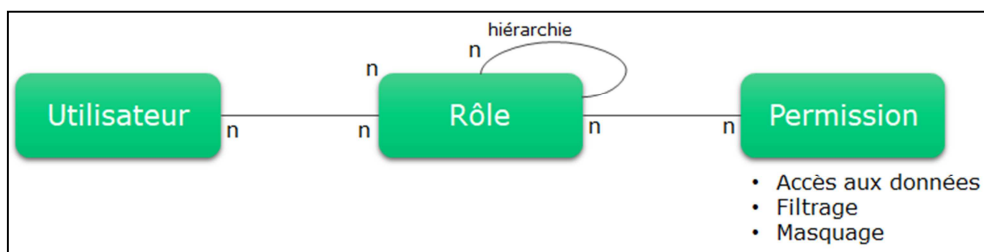


Figure 18. Mode d'accès RBAC

Il a été décidé au sein du LDA de mettre en premier lieu en place l'approche RBAC. Le mode ABAC sera utilisé pour la gestion des données personnelles et sensibles.

Gestion des données personnelles et données sensibles

Le LDA héberge tout type de données, dont des données personnelles et sensibles.

Une **donnée est dite « personnelle »** lorsque qu'elle se rapporte à une personne physique identifiée ou identifiable. Par exemple, le nom et prénom. Une **donnée sensible** peut être une donnée personnelle (ce n'est pas toujours le cas au sens du RGPD) et a un effet discriminant car portant sur (liste non exhaustive) :

- l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques
- l'appartenance syndicale
- les données génétiques, les données biométriques traitées uniquement pour identifier un être humain
- les données concernant la santé (numéro de sécurité sociale)

Le RGPD interdit de collecter ces données sauf dans certains cas (intérêt public, consentement de la personne...).

Certaines données sont considérées **sensibles** au sein du LDA **mais pourtant ne sont pas personnelles**. Par exemple, des informations nominatives concernant les agents de l'ASP validant les dossiers instruits sont archivées. Elles font partie de l'activité professionnelle du collaborateur et donc ne sont pas considérées comme des données personnelles par le RGPD. Elles restent malgré tout très sensibles, car la divulgation d'informations liées à cette personne pourrait lui causer préjudice (exemple : harcèlement suite au refus du paiement d'un dossier d'agriculteur instruit par le collaborateur).

Les données personnelles et sensibles récoltées dans le LDA sont listées dans un document Excel, et annexé à l'Analyse d'impact relative à la protection des données (AIPD) du Lac de données agricoles ; le glossaire-métier contiendra également ces informations.

Il est **impératif**, lors l'élaboration d'un nouveau cas d'usage, de réfléchir aux nouvelles données qui seront versées dans le LDA afin d'éviter d'avoir des données personnelles et sensibles non pertinentes qui sont stockées dans la zone brute ou archivées.

Gestion des référentiels

Les données référentielles (ou *master data* en anglais) sont des données stables, essentielles, produites et utilisées par de nombreuses personnes au sein d'une organisation. Par exemple, le genre de l'exploitant agricole ou la liste des indicateurs de réalisation de la PAC sont des données référentielles. Ce ne sont pas des données transactionnelles. Elles sont le plus souvent dispersées dans les différents SI métiers des organisations, comme c'est le cas à l'ASP. En effet, le catalogue RefPAC, administré par la DDRP, contient l'ensemble des référentiels de la PAC et est situé en dehors du LDA. Toutefois, l'applicatif Synapse consomme et héberge aussi ces référentiels et d'autres qui lui sont propres.

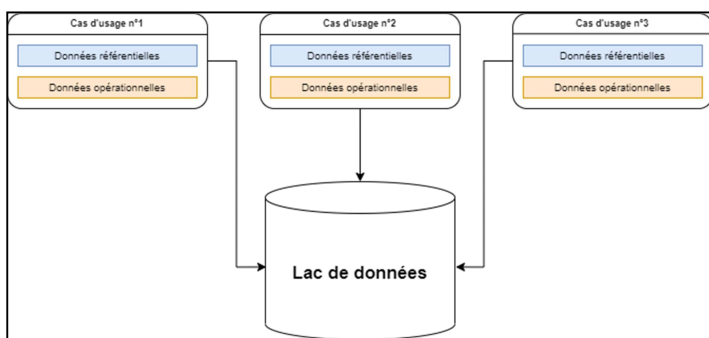


Figure 19. Schéma des données référentielles dispersées

Cette situation, où les données sont en silos, engendre des risques importants :

- Perte de qualité de données : comment déterminer l'unique source de vérité (*single source of truth* en anglais) lorsque deux SI utilisent des versions différentes d'une même donnée référentielle ?
- Augmentation des coûts : lorsqu'il y a plusieurs sources de vérité, il est nécessaire de procéder à des vérifications coûteuses pour définir laquelle est l'unique.

La gestion des données de référence (*master data management* en anglais) vise donc à **stocker les données référentielles à un seul endroit**, qui deviendra la source unique de vérité. Elle permet ainsi de contrôler leur qualité et de les rendre accessibles à tous les cas d'usage qui souhaiteraient les consommer.

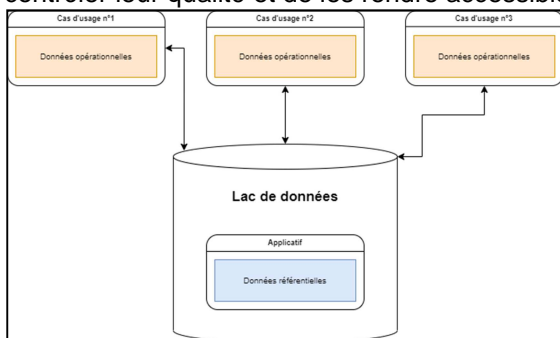


Figure 20. Unique source de vérité

Les données référentielles utilisées par Synapse seront hébergées au sein du LDA dans un espace applicatif dédié, et **chaque cas d'usage pourra les utiliser comme « bien commun »**. A noter que les données référentielles à usage d'une seule application sont exclues.

Choix des outils et valorisation des données

Afin d'aider les équipes projet des cas d'usage à définir quel outil utiliser pour la valorisation des données, un arbre de décision a été construit. En effet, via le portail utilisateurs du LDA, plusieurs logiciels sont disponibles pour de la valorisation des données :

- Dataiku : outil plus tourné Data Science avec un nombre limité de licences (10 au total),
- MicroStrategy : permet de créer des tableaux de bord et une couche sémantique pour un grand nombre d'utilisateurs (> 300),
- R-Shiny : n'est utilisé que pour le RAP.

Le schéma ci-dessous répond à la question : **quel outil pour quelles données ?**

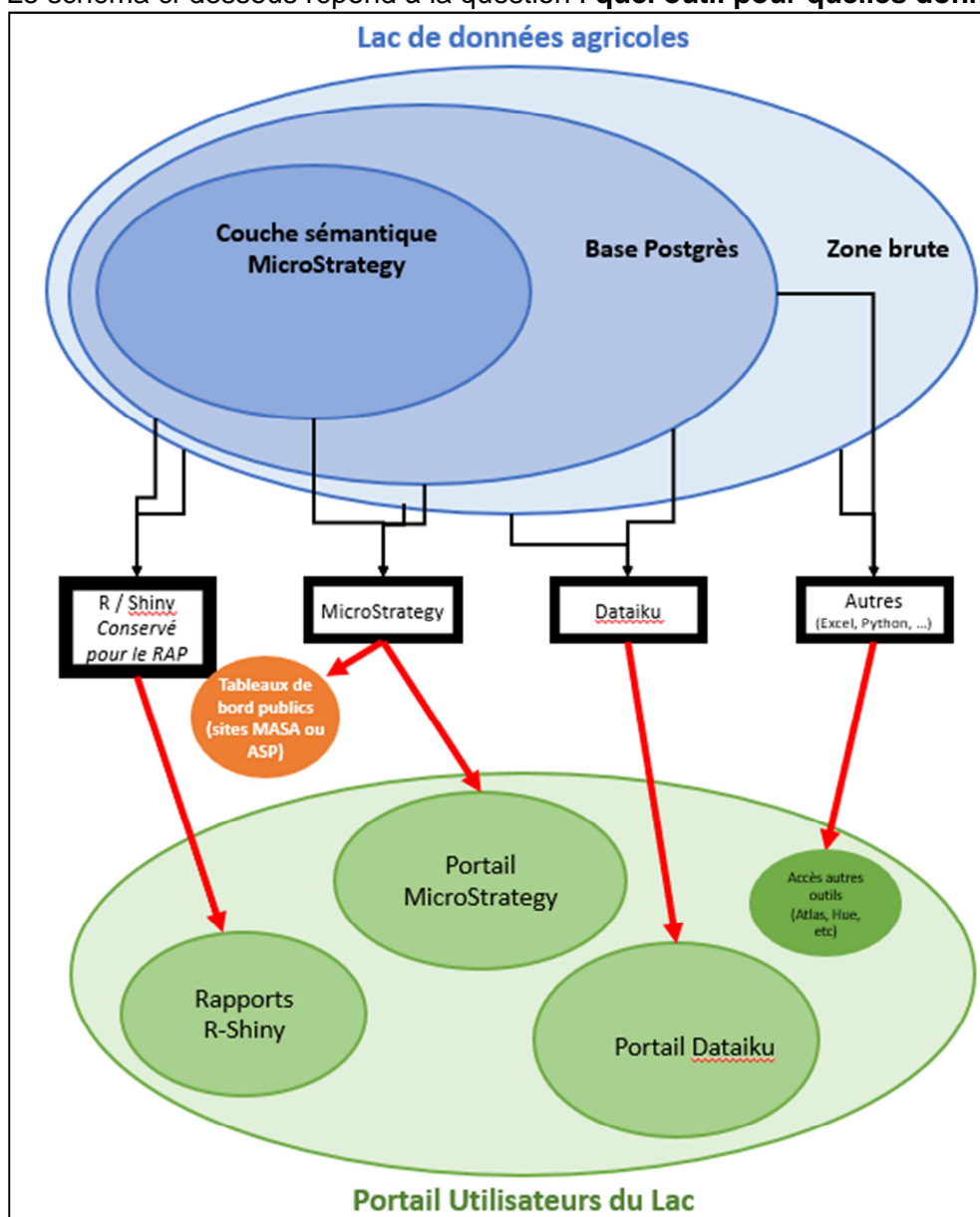


Figure 21. Arbre d'exposition