

Le directeur
Réf : 2021-CMR-019

Paris, le 09/06/2021

A Madame la secrétaire générale
du ministère de l'agriculture et de l'alimentation

Objet: Projet de mise en place d'un lac de données par l'Agence de Services et de Paiement – Avis conforme défavorable

Réf: Courrier et dossier de saisine du 10 mai 2021

En application de l'article 3 du décret n° 2019-1088 du 25 octobre 2019 relatif au système d'information et de communication de l'Etat et à la direction interministérielle du numérique, vous avez saisi la DINUM pour avis sur le projet de mise en place d'un lac de données par l'Agence de Services et de Paiement (ASP).

1. Présentation du projet

Ce projet s'inscrit dans le cadre de la prochaine programmation 2021-2027 de la Politique Agricole Commune (PAC) qui entrera en vigueur début 2023. La PAC représente un tiers du budget de l'Union Européenne et la France est le plus grand bénéficiaire (environ 18%). Pour répondre aux nouvelles orientations de l'Union Européenne, une refonte du SI PAC de l'ASP, le plus important organisme payeur des aides aux agriculteurs en Europe et le principal en France¹, est prévue. Ce programme de refonte comporte 3 chantiers principaux :

- Chantier Performance : il consiste à mettre en place un nouveau modèle de mise en œuvre de la PAC en donnant plus de marge de manœuvre aux Etats membres dans la définition de leur politique agricole (avec beaucoup moins de règles définies au niveau européen). Un rapport annuel de performance permettra d'apprécier la conformité des mesures prises au niveau national avec les objectifs de politique agricole fixés par la France dans son plan stratégique national et validés par l'UE. Pour mettre en œuvre ce nouveau modèle, un nouveau SI de pilotage de performance dialoguant avec l'ensemble des SI de la PAC via plusieurs interfaces de programmation d'application (API) et un lac de données sont prévus. La construction

¹ L'ASP est également l'organisme de coordination des 4 organismes payeurs français (ASP, FAM, ODARC, ODEADOM).

d'un Référentiel Commun des Individus de la PAC (RCI PAC) constitue le 3^e volet du chantier, prérequis des 2 volets précédents.

- Chantier Construction du SI de gestion et de contrôle: il consiste à remplacer le système existant « OSIRIS »² dédié aux aides non « surfaciques » mais qui ne répond pas à la première étape de décentralisation réalisée lors de la programmation précédente. De nombreux rapports (cour des comptes, inspection générale des finances...) appellent à son remplacement au plus tôt. Le chantier prévoit :
 - a) un portail à destination des demandeurs d'aide leur permettant de déposer les demandes pour les dispositifs gérés par l'Etat et d'en suivre l'avancement administratif;
 - b) une composante de gestion, assurant le traitement des dossiers d'aide depuis leur réception, leur instruction, leur programmation, et les contrôles jusqu'à la pré-liquidation sur le même périmètre³;
 - c) une composante propre à l'organisme payeur, permettant notamment d'assurer le paiement des dossiers, y compris ceux instruits par les régions.
- Chantier Monitoring: le contrôle des aides aux agriculteurs repose aujourd'hui sur un système complexe de contrôle sur place par échantillonnage, contrôles réalisés par photo-interprétation assistée par ordinateur, éventuellement complétée par des déplacements sur le terrain. Le chantier consiste à fiabiliser le Registre Parcellaire Graphique (RPG) en historisant son évolution dans le temps d'une part et à mettre en place une infrastructure de connexion permettant d'accéder aux images satellites Sentinel du programme Copernicus délivrées par le dispositif européen DIAS d'autre part. L'objectif est d'utiliser de nouvelles technologies (ex: utilisation d'algorithmes d'intelligence artificielle) pour vérifier l'éligibilité des bénéficiaires de la PAC aux aides qu'ils ont demandées, renforcer le contrôle et le suivi « en temps réel » et simplifier la relation avec les bénéficiaires.

Le budget de la refonte du SI PAC de l'ASP a été estimé à 44,4 M€ hors titre 2 sur la période 2020-2022. Le programme a obtenu en 2020 un financement du FTAP à hauteur de 28,8 M€.

Le projet de mise en place d'un lac de données fait partie du chantier performance. Il vise à mettre en place un lac de données (datalake), incluant un entrepôt de données (datawarehouse) et un laboratoire de données (datalab) qui doit permettre de stocker et de traiter entre autres des données graphiques, en cohérence avec la mise en place du monitoring imposé par la Commission Européenne.

Le budget du projet de mise en place du lac de données a été estimé dans le dossier FTAP à 5,55 M€ hors titre 2 sur la période 2020-2022 dont 3,61 M€ financés par le FTAP. Il vient d'être re-estimé à 7,5 M€ (et à 12,5M€ en incluant 2 ans de MCO) hors titre 2 entre 2020 et

² Le principal système de gestion existant de l'ASP, dédié aux aides « surfaciques » aux agriculteurs (ISIS) sera maintenu.

³ Le remplacement d'OSIRIS pour les dossiers gérés directement par les régions sera pris en charge par les régions sur leurs fonds propres, sous la supervision de l'ASP.

2022 dans la fiche MAREVA.

Pour mettre en place le lac de données, un marché subséquent à l'accord-cadre ISIS mono-attributaire notifié en 2019 à la société Capgemini a été approuvé par le Conseil d'Administration de l'ASP le 12 avril 2021 sous réserve de l'avis conforme de la DINUM.

2. Analyse et constats

Sur la base des documents transmis⁴ et des échanges organisés avec les équipes de l'ASP, je souhaite partager avec vous les constats suivants :

1. Le besoin de disposer d'un lac de données, incluant un entrepôt de données et un laboratoire de données pour produire les indicateurs de performance avec les données « historisées » d'ISIS et répondre aux impératifs de traçabilité de l'ensemble des opérations conduisant aux paiements successifs aux exploitants dans le cadre de la programmation 2021-2027, est tout à fait compréhensible. Néanmoins **la déclinaison opérationnelle de ce besoin en premiers cas d'usage et en plans projets décisionnel et analyse de données paraît insuffisante.**
 - La seule source de données identifiée du lac de données à ce jour est ISIS (réplica, photos et événements historisés), dont la moitié au moins relèverait du stockage à froid. L'urbanisation des éléments applicatifs d'ISIS est évoquée mais pas programmée. L'urbanisation permettrait de réduire son coût de fonctionnement et de fluidifier la circulation des données de référence externes et de gestion des autres organismes payeurs qui alimentent aujourd'hui directement ISIS.
 - La mutualisation des besoins avec d'autres SI de l'ASP, en premier lieu OSIRIS (ou ses futurs remplaçants) n'est pas prévue à ce jour.
 - Le besoin en termes de projets décisionnel et analyse de données n'est pas détaillé. Les pratiques actuelles de gestion pour ISIS (pas d'outil décisionnel, exploitation des données reposant en partie sur des extractions non industrialisées, requêtes et tableaux de bord disponibles limités) ne démontrent pas une maturité suffisante pour réussir rapidement un projet tel que celui envisagé.
 - La maturité du besoin auquel le lac de données doit apporter une réponse dépend du calendrier de la nouvelle programmation de la PAC et surtout de la stabilisation des nouvelles règles de gestion. Elle dépend aussi de l'avancement du chantier monitoring qui s'avère complexe à mettre en œuvre.
 - Le besoin en termes de sécurité de lac de données doit être développé, par exemple l'anonymisation ou le chiffrement des données.
2. Les compétences internes adéquates et suffisantes sont essentielles pour réussir un projet de cette nature. Ces compétences nécessaires vont bien au-delà des tâches classiques qui incombent à la maîtrise d'ouvrage comme le pilotage, les tests et la

⁴ L'ASP n'a pas répondu à la demande des documents complémentaires adressée le 14 mai 2021.

conduite de changement. Ainsi, **la capacité à conduire un tel projet qui nécessite des compétences spécifiques (datascientists, experts métiers, etc.) n'a pas été démontrée.**

- Aucune précision sur les ressources et compétences internes disponibles et leur dimensionnement n'a été fournie dans le dossier.
 - L'ASP indique qu'une équipe de 7 personnes est prévue en interne. Elle est composée de 2 interlocuteurs métier, 2 datascientists⁵, 2 experts techniques et la direction des opérations de la DSDA. Leur disponibilité pour le projet, au-delà de leurs missions actuelles, n'est pas précisée. Nous estimons donc qu'il existe un risque fort de dépendance prolongée aux compétences du titulaire du marché, qui n'est pas couvert par un plan d'internalisation des compétences. Cette situation peut avoir pour conséquence une difficulté de maîtrise de la trajectoire du projet et de la pérennité du produit développé par l'ASP.
 - La mobilisation des experts métier doit être précisée.
3. La solution technique proposée est très complète et repose sur des technologies à l'état de l'art. Les exigences de scalabilité et de résilience ont été prises en compte dans la réflexion sur la construction du socle technique. En tenant compte de ces éléments, **le coût du projet reste cependant très élevé avec notamment une part de coûts fixes trop importante.**
- En combinant les éléments fournis dans la fiche MAREVA et dans le rapport de présentation au comité technique des marchés, le coût total de l'opération s'élève à 27 M€ hors titre 2 sur 5 ans⁶, sans tenir compte des ressources internes.
 - Certaines dépenses, comme par exemple la réplique de la base ISIS (UO11.1-FONCT) ou la prise de photos historisées hebdomadaires d'ISIS (UO11.3-FONCT) s'élèvent respectivement à 1,7 M€ HT et 1,5 ETP du titulaire de marché et à 3,5 M€ HT et 2 ETP du titulaire de marché pendant 4,5 ans, alors que ces opérations devraient être automatisées dès la phase de construction.
 - Un tiers de prestations, notamment les projets décisionnel et analyse de données font bien l'objet des UO unitaires à la commande (UO14 et UO15) en fonction du besoin réel et de la capacité interne de l'ASP à les conduire, mais deux tiers des prestations⁷ sont forfaitaires et susceptibles d'être revues uniquement à la hausse (via UO11.x-COMPL et UO12). Or les coûts de fonctionnement de l'entrepôt de données (UO11.2-FONCT) ou du laboratoire de données (UO11.4-FONCT) par exemple dépendront bien de la volumétrie des données à stocker et du nombre de projet à mener effectivement.
 - Le nombre de projets décisionnel et analyse de données retenus pour fixer le forfait

⁵ Dont un externe via le marché UGAP.

⁶ Composés de 23,5 M€ pour le marché subséquent, 2,3 M€ pour l'utilisation des logiciels Cloudera/Dataïku porté également par le même titulaire du marché et 1,2 M€ pour le cadrage du projet en 2020.

⁷ Il s'agit du pilotage (UO10), du socle technique (UO11) sous forme de 5 types d'UO forfaitaires trimestrielles mais indissociables, ainsi que l'utilisation des logiciels Cloudera/Dataïku pour une durée de 3 ans à minima.

est très élevé. A titre d'exemple le forfait est basé sur le déploiement de 32 algorithmes comme la détection de la fraude et la mobilisation de 10 datascientists sur 5 ans pour les projets analyse de données.

- L'ASP a évoqué la clause de réexamen (article 1.5.4 du CCP du marché subséquent) pour réduire le cas échéant les coûts fixes d'infrastructures. Nous constatons que cette clause est soumise à de nombreuses conditions et dépend *in fine* de la bonne volonté du titulaire du marché.

4. Le socle technique est hébergé chez OVH dans un *cloud* privé mais contractualisé par le titulaire du marché.

- La durée de l'hébergement est limitée à 5 ans.
- En l'état, le marché ne permet pas la possibilité de préserver la solution d'hébergement et les logiciels de la solution technique indépendamment de l'intégrateur. La pérennisation du datalake serait donc conditionnée par la poursuite du contrat liant l'ASP et Capgemini, sauf nouvelle négociation que l'ASP devrait mener avec Capgemini dans des circonstances qui lui seraient largement défavorables.
- Le marché prévoit une réversibilité de 4 mois à 123 K€ (UO17), mais elle ne concerne que le transfert des connaissances et des documents de l'intégrateur, le coût de portabilité du socle technique sur d'autres hébergeurs n'est pas estimé.
- Le rôle et la responsabilité respectifs de l'intégrateur et de l'hébergeur sur les différents services, licences ou outils proposés par le socle technique ne sont pas clairement définis⁸. **En cas de reprise, les coûts cachés pourraient être potentiellement très importants.**

3. Conclusion

De manière générale, le projet paraît surdimensionné par rapport au besoin réel de l'ASP et à sa capacité interne à le conduire en 5 ans. La possibilité d'ajuster le coût de fonctionnement du socle technique au plus près de son utilisation réelle n'est pas prévue explicitement dans le marché subséquent. Le dossier fourni par l'ASP ne démontre nullement sa capacité à piloter les engagements du titulaire du marché⁹, et les difficultés rencontrées par les précédents projets ne permettent pas d'être confiant en la matière.

Ces constats structurants m'amènent à **émettre un avis conforme défavorable sur le projet de mise en place de lac de données de l'ASP** tel qu'il m'a été soumis.

⁸ Au-delà des VM, de l'utilisation de Cloudera, Dataïku et de licences Oracle (prises en charge par le marché ISIS).

⁹ Nous nous interrogeons par exemple sur les constats suivants :

- Le dossier de cadrage a été rédigé par le titulaire du marché avec une contribution de l'ASP limitée à « l'expression macroscopique des grandes orientations évoquées par la DSDA lors de la réunion de lancement du 21 novembre 2019 » (annexe 4 du CCP, page 5).
- « Les coûts sont issus des forfaits définis par le titulaire dans l'annexe financière, des quantités définies par le titulaire dans le DQE et de la planification de mobilisation des unités d'œuvre définie par le titulaire dans son mémoire technique pour répondre à la trajectoire définie au cahier des charges de l'ASP » (rapport de présentation au comité technique des marchés, page 12).

L'exploitation des données est le fruit d'un processus d'apprentissage long et le principe de « pensons grand et avançons petit » doit être recherché. J'invite l'ASP à redimensionner le projet sur une base notablement plus réduite, quitte à l'ajuster à la hausse dans un second temps, avec des UO complémentaires, y compris pour le socle technique si l'ampleur du besoin tel qu'il est prévu dans le dossier est confirmé.

Je souhaite aussi qu'elle étudie la possibilité d'utiliser le support de marché UGAP dédié aux services d'informatique en nuage de l'Etat qui permettrait de traiter avec les mêmes acteurs mais avec la possibilité d'assurer au mieux la pérennité et la portabilité du socle technique, en faisant de l'ASP le primo-contractant de l'hébergement *cloud*.

La DINUM se tient à la disposition de l'ASP pour l'accompagner dans le recadrage du projet et la contractualisation avec ses prestataires.

Le projet pourra être réexaminé afin de prendre en compte d'autres éléments (qualité et sécurité des données, interdépendance entre les différents chantiers, notamment avec le nouveau SI de performance et le nouveau système monitoring, gouvernance...) que l'ASP va communiquer à mes services dans le cadre des travaux relatifs à la sécurisation de la mise en production du nouveau SI PAC qui ont été demandés à la DINUM par le ministère de l'agriculture et de l'alimentation et le ministère des comptes publics.

Conformément au décret n° 2019-1088 du 25 octobre 2019 relatif au système d'information et de communication de l'Etat et à la direction interministérielle du numérique, la transmission du présent avis met fin à la procédure de saisine.

A handwritten signature in blue ink, consisting of a series of overlapping, vertical and horizontal strokes that form a complex, somewhat abstract shape. The signature is positioned above the name and title of the signatory.

Nadi BOU HANNA

Directeur interministériel du Numérique

Copie :

Monsieur le Premier ministre

A l'attention de :

- Monsieur le directeur de cabinet
- Madame la secrétaire générale du gouvernement

Monsieur le ministre de l'agriculture et de l'alimentation

A l'attention de :

- Monsieur le directeur de cabinet
- Madame la directrice générale de la performance économique et environnementale des entreprises
- Monsieur le secrétaire général adjoint
- Monsieur le sous-directeur du numérique
- Monsieur le président-directeur général de l'agence de services et de paiement

Monsieur le ministre de l'économie, des finances et de la relance

Monsieur le ministre délégué chargé des comptes publics

A l'attention de :

- Monsieur le directeur de cabinet
- Madame la directrice du budget

Madame la ministre de la transformation et de la fonction publiques

A l'attention de :

- Monsieur le directeur de cabinet
- Monsieur le directeur interministériel de la transformation publique