

Du capteur à l'indicateur : les entrepôts de données spatiales au service d'une meilleure maîtrise des consommations énergétiques des entreprises agricoles

Emilie Donnat, D. Boffety, S. Bimonte, André G, J Capdeville, J.M. Deumier, Sophie Gendre, Georges J, A. Cottais, A Tailleur, et al.

▶ **To cite this version:**

Emilie Donnat, D. Boffety, S. Bimonte, André G, J Capdeville, et al.. Du capteur à l'indicateur : les entrepôts de données spatiales au service d'une meilleure maîtrise des consommations énergétiques des entreprises agricoles. Innovations Agronomiques, 2017, 55, pp.201-214. 10.15454/1.51377788991795E12 . hal-02605913

HAL Id: hal-02605913
https://hal.inrae.fr/hal-02605913

Submitted on 31 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Du capteur à l'indicateur : les entrepôts de données spatiales au service d'une meilleure maîtrise des consommations énergétiques des entreprises agricoles

Donnat E.[1], Boffety D.[2], Bimonte S.[3], André G.[2], Capdeville J.[4], Deumier J-M.[5], Gendre S.[5], Georges J.[6], Cottais A.[7], Tailleur A.[8], Huneau T.[9], Le Frileux Y.[10], Dejean C.[11], Erignoux A.[12], Chanet J.-P.[3]

[1] ACTA, 149 rue de Bercy, 75595 Paris Cedex 12
[2] IRSTEA, centre de Clermont-Ferrand, UR TSCF, Technologies et systèmes d'information pour les agrosystèmes, Domaine des Palaquins, 03150 Montoldre
[3] IRSTEA, centre de Clermont-Ferrand, 24 avenue des Landais, BP 50085, 63172 Aubière Cedex
[4] Institut de l'Elevage, antenne de Toulouse, Service Bâtiments-Environnement, BP 42118, 31321 Castanet Tolosan Cedex
[5] ARVALIS- Institut du végétal, Service Agronomie-Economie-Environnement, Irrigation-Gestion de l'eau, 6 chemin de la côte vieille, 31450 Baziège
[6] Chambre d'Agriculture de Haute-Garonne, 28 route d'Eaunes, 31605 Muret
[7] Chambre d'Agriculture de Bretagne, avenue du Général Borgnis Desbordes, BP 398, 56009 Vannes Cedex
[8] ARVALIS- Institut du végétal, La Jaillière, 44370 La Chapelle St Sauveur
[9] Chambre d'Agriculture de Loire-Atlantique, ferme expérimentale, 44590 Derval
[10] Institut de l'Elevage, Station expérimentale caprine du Pradel, 07170 Mirabel
[11] IRSTEA, UMR GEAU, Domaine de Lavalette 361 rue JF Breton, BP 5095, 34196 Montpellier Cedex 5
[12] NPN Ingénierie, 64 avenue du Général Bizot, 75012 Paris

Correspondance : emilie.donnat@acta.asso.fr

Résumé

Les questions énergétiques liées aux exploitations agricoles constituent un sujet de recherche qui préoccupe la communauté internationale et ce, pour des questions environnementales et économiques. En matière de systèmes d'information, la technologie des entrepôts de données et les bases de données multidimensionnelles font émerger de nouvelles solutions pour le recueil et l'utilisation décisionnelle des données. Ces technologies permettent d'établir, sur des données collectées et archivées, toutes sortes d'indicateurs résultant d'opérations agrégées (somme, moyenne) en fonction de différents critères (période de temps, types de machines, types de travaux, entités de gestion, etc.). Dans ce projet, les chaînes d'intégration de l'information, du capteur jusqu'à l'indicateur, ont été développées pour les consommations de carburant liées aux équipements mobiles (cultures) et pour les consommations d'électricité et d'eau liées aux activités en bâtiments d'élevage (laitier et caprin) et à l'irrigation. L'exploitation des données collectées grâce aux solutions technologiques mises en place sur le terrain, puis leur intégration dans l'entrepôt de données en vue d'établir des indicateurs pertinents de performance énergétique permet le suivi en temps réel et la traçabilité des consommations d'énergie à différentes échelles de l'exploitation agricole, à travers tableaux de bord et diagnostics. Ce projet a ainsi débouché sur des solutions technologiques pour l'acquisition fiable de données, sur un entrepôt des données capable d'intégrer des données hétérogènes du point de vue de leur qualité et de leur granularité à une échelle fine afin de fournir des indicateurs de performance pertinents.

Mots-clés : entrepôt de données spatiales, capteur, indicateur, performance énergétique

Abstract: From sensor to indicator: spatial data warehouse serving for better management of energy consumption of agricultural farms

Agricultural energy consumption is an important issue that concerns the international community for environmental and economic reasons. Technology of data warehouse and multidimensional database opens the door to new solutions for collect and decisional use of data. These technologies enable to have access, on collected and stored data, to a whole range of indicators resulting from aggregated operations (sum, mean) in function of different criteria (period of time, type of machine, type of work or management entity). In this project, integration chains of information from the sensor to the indicator have been developed for gas consumptions in mobile farm equipment (for cultures) and for electricity and water consumption in breeding buildings (cow and goat breeding) and for irrigation. The use of data collected thanks to the technological solutions implemented on the field, then their integration in the data warehouse in order to establish relevant indicators of the energetic consumption enables monitoring in real time and traceability of energy consumption at different scales of the farm, through dashboard instrument panel and diagnosis. This project has led to technological solutions to obtain reliable data, on a data warehouse able to integrate heterogeneous data in terms of quality and granularity at an accurate scale, in order to provide reliable performance indicators.

Keywords: spatial data warehouse, sensor, indicator, energetic performance

Introduction

Dans le contexte européen et mondial actuel, les enjeux de développement durable, de préservation des ressources naturelles et de lutte contre le changement climatique affectent directement les entreprises agricoles. Concernées par chacun de ces enjeux, les dépenses énergétiques des entreprises agricoles sont au cœur des préoccupations socio-économiques : quantification et maîtrise des flux, réduction des coûts liés aux dépenses énergétiques directes (carburant, électricité), empreinte énergétique.

Par ailleurs, les technologies de l'information sont à même de fournir des chaînes d'intégration de l'information, de la donnée brute jusqu'à des indicateurs de différentes échelles spatiales et temporelles (Chaudhary *et al.*, 2004, Iftikhar, 2009). Grâce à l'intégration des données issues de capteurs placés sur les équipements des exploitations dans un système d'information de type entrepôt de données spatiales, on peut faire des analyses fines des données suivant différentes dimensions (spatiales, temporelles, type de cultures, type d'équipements, etc.) permettant de dégager des indicateurs assurant une meilleure compréhension des exploitations agricoles (Nilakanta *et al.*, 2008, Thornsbury *et al.*, 2003).

Lors d'un premier projet portant sur l'évaluation fine des performances énergétiques des entreprises agricoles par l'utilisation des NTIC et intitulé EnergéTIC (Donnat *et al., 2013*), les résultats ont montré la faisabilité et l'intérêt d'une collecte en routine de données pour le pilotage opérationnel (semaine, mois) de ces performances grâce à l'utilisation de capteurs bas-coûts, réseaux sans fil, entrepôts de données. En particulier, la technologie des entrepôts de données a permis l'élaboration de bilans énergétiques plus fins à l'échelle de la parcelle, de l'atelier de production ou de l'opération dans l'entreprise agricole (Pradel *et al.*, 2009). Cette première étude a cependant souligné des points restant encore à approfondir comme la fiabilisation des recueils automatiques de données, la valorisation des indicateurs vers les différents acteurs (agriculteurs, conseillers) en fonction des objectifs (outils de suivi, aide à l'amélioration du bilan énergétique des exploitations, etc.) ou encore la prise en compte de l'hétérogénéité des données dans les entrepôts de données spatiales. Ce dernier point est capital si l'on veut définir des indicateurs ayant du sens.

Associant l'ACTA, ARVALIS-Institut du végétal, les chambres d'agriculture de Bretagne, de Haute-Garonne et de Loire-Atlantique, l'Institut de l'Elevage et l'IRSTEA, ce projet (acronyme : EDEN) s'est ainsi attaché à promouvoir des solutions technologiques pour une meilleure maîtrise des consommations énergétiques, en utilisant un système d'information de type entrepôt de données à partir de l'intégration de données collectées à une échelle plus fine et hétérogènes par leur niveau de qualité, de précision et de granularité. Pour ce faire, le projet s'est structuré en trois phases :

1. poursuivre, par la fiabilisation des solutions technologiques (capteurs, systèmes d'acquisition embarqués), la collecte des données sur les sites expérimentaux des partenaires et enrichir la base de données issue d'EnergéTIC avec les nouvelles données ;
2. construire un entrepôt de données qui prenne en compte l'hétérogénéité des données et développer le processus d'intégration des données dans l'entrepôt ;
3. valoriser des indicateurs de performance énergétique obtenus grâce aux techniques citées précédemment pour le suivi en temps réel, la traçabilité des consommations et la réalisation de diagnostics à différentes échelles de l'exploitation agricole.

1. L'acquisition des données sur les sites expérimentaux : fiabilisation des solutions technologiques et collecte des données

L'acquisition des données s'est effectuée sur les sites expérimentaux des partenaires du projet (Figure 1).

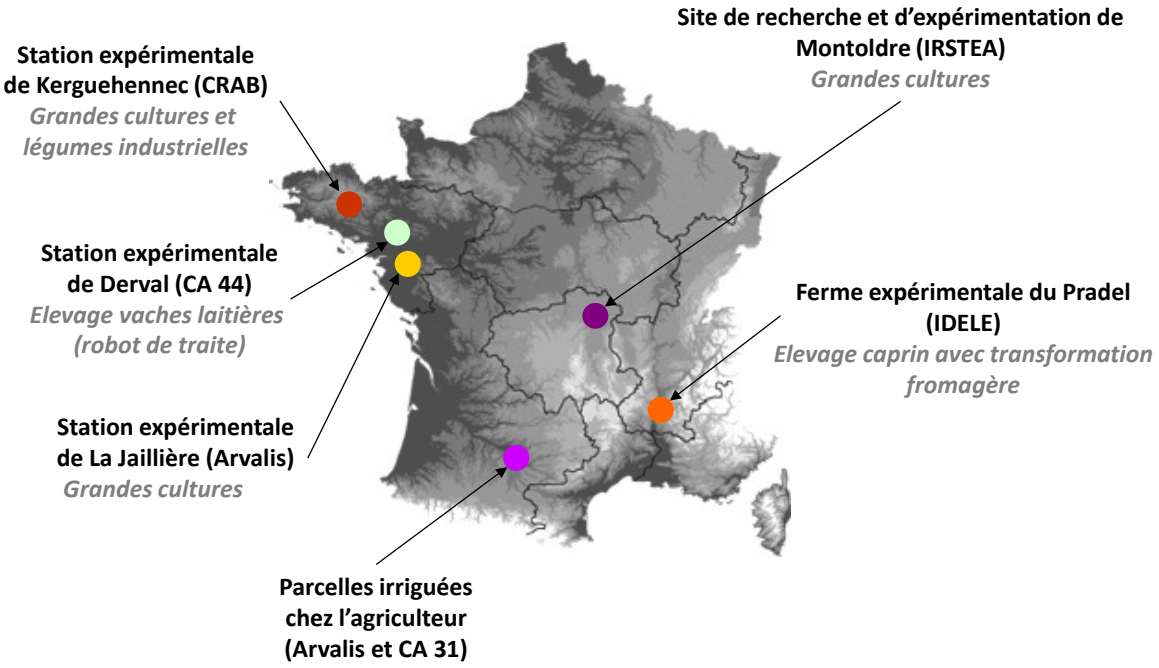


Figure 1 : Carte des sites expérimentaux des partenaires

1.1. Fiabilisation des solutions technologiques développées dans le projet EnergéTIC

Dans le cadre du projet EnergéTIC, un système d'acquisition automatique géoréférencé de la consommation énergétique des équipements mobiles a été développé. Celui-ci implique une unité d'acquisition embarquée intégrant un GPS, une jauge à carburant numérique, différents capteurs "tout ou rien" et un système d'identification des équipements utilisés à partir de badges RFID (Radio

Frequency Identification). Le développement industriel a ensuite été confié à ExoTIC Sytems basée à Clermont-Ferrand, dans l'objectif d'équiper certains matériels mobiles de sites expérimentaux du projet EDEN.

En fonction des travaux agricoles réalisés, le dispositif automatique proposé permet de suivre l'activité du tracteur et des différents matériels associés lors des chantiers avec une mesure de la quantité de carburant dans le réservoir à chaque changement d'état des capteurs installés sur le tracteur. Ces différents capteurs "tout ou rien" placés sur certains éléments du tracteur permettent d'identifier les tâches réalisées à partir des états de fonctionnement des outils (état de la commande de la prise de force, état de la commande d'un distributeur hydraulique auxiliaire, état du relevage arrière, état du bras du troisième point, état du piton d'attelage). Directement intégré au boîtier d'acquisition, un système GPS localise le tracteur et un système automatique à partir de badges RFID identifie les outils attelés au tracteur et les opérations techniques qui y sont associées. Cette identification porte avant tout sur l'outil principal, mais peut également être étendue à l'association d'outils ou d'équipements, aux opérateurs, aux installations fixes ou mobiles à la condition qu'ils soient munis d'un badge RFID. Toutes les informations d'identification RFID délivrées sont collectées sur le boîtier d'acquisition. La Figure 2 montre le principe de la solution d'acquisition automatique ExoTIC Systems.

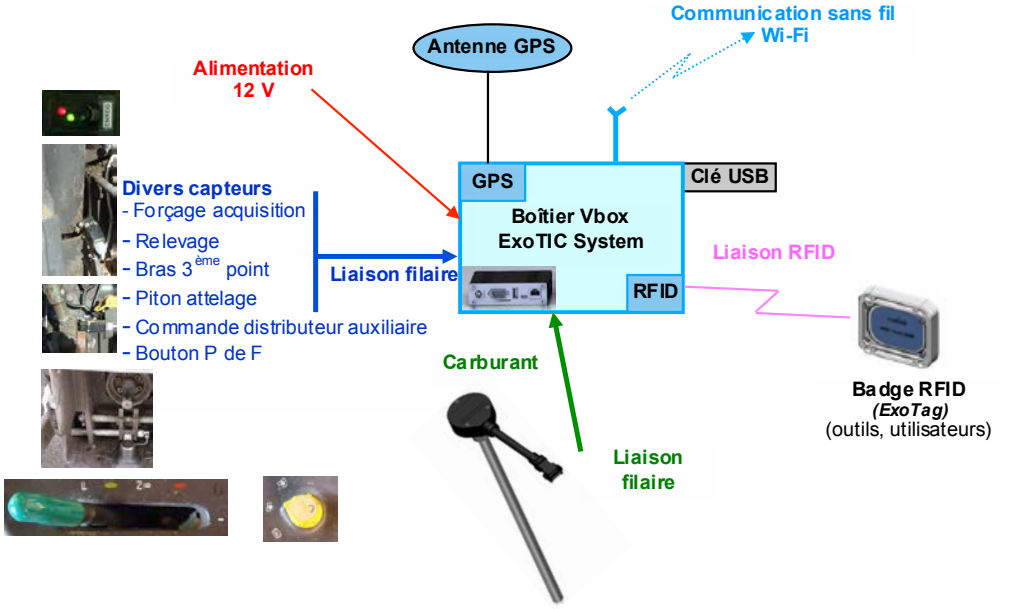


Figure 2 : Schéma de principe de la solution d'acquisition automatique

Afin de limiter la quantité de données enregistrées, l'acquisition se fait soit lorsqu'un changement d'état est détecté sur un des capteurs "tout ou rien", soit sur demande de l'opérateur sur pression du bouton poussoir du poste de conduite. Le transfert des fichiers de données pour exploitation ultérieure se fait à partir de la mémoire interne du boîtier d'acquisition par copie sur une clé USB. Les données brutes recueillies et assemblées constituent des fichiers textes organisés avec les informations des différents capteurs : date, temps, données GPS, état des capteurs utilisés, informations concernant le relevé de la quantité de carburant et l'identification RFID. Ces fichiers sauvegardés chronologiquement sur la clé USB peuvent donc ensuite être transférés vers la base de données pour être exploités.

1.2. Collecte des données sur matériel mobile

Les données collectées sont : l'heure, la latitude et la longitude données par le GPS, une valeur binaire représentant les états des capteurs tout ou rien (TOR), la valeur de la quantité de carburant du réservoir et l'identifiant des badges "Tag" associés aux outils ou équipements utilisés (Figure 3).

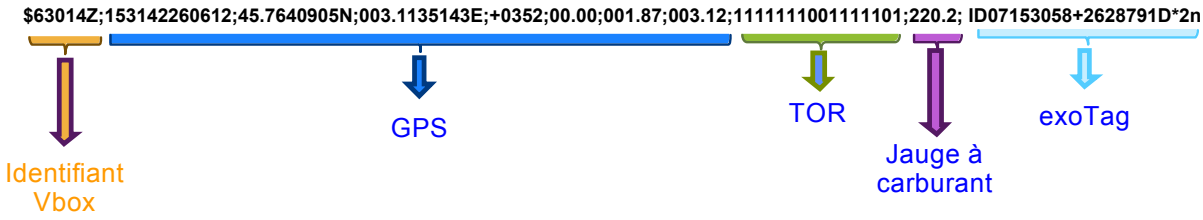


Figure 3 : Format de la trame des données enregistrées à chaque changement de valeur de capteur

Les fichiers obtenus (Figure 4) sont traités automatiquement et les données sont intégrées dans la base de données. Les opérations sont attribuées aux parcelles grâce au traitement des informations GPS. Le traitement des données spatiales est réalisé grâce à l'extension spatiale de l'outil ETL Talend.

$Ares610;3.905;13.784;150553040615;46.1362419N;003.2101500E;+0323;00.00;006.86;006.96;1111100011111001;183.4*D6
$Ares610;3.905;13.837;150558040615;46.1362457N;003.2101510E;+0324;00.00;006.35;006.44;1111100001110001;183.4*D3
$Ares610;3.915;13.802;150559040615;46.1362457N;003.2101512E;+0324;00.00;006.34;006.42;1111100011111001;183.4*D2
$Ares610;3.915;13.775;150614040615;46.1362534N;003.2101541E;+0324;00.00;007.31;007.38;1111100001110001;183.4*DF
$Ares610;3.918;13.465;150621040615;46.1362534N;003.2101548E;+0324;00.00;006.79;006.85;1111100011111001;183.4*D5
$Ares610;3.915;13.775;150633040615;46.1362495N;003.2101555E;+0324;00.00;006.11;006.17;1111100001110001;183.4*DC
$Ares610;3.900;13.767;150641040615;46.1362457N;003.2101560E;+0324;00.00;005.60;005.65;1111100011111001;183.4*D3
$Ares610;3.899;13.846;150647040615;46.1362457N;003.2101562E;+0324;00.00;005.43;005.48;1111100001110001;183.4*D6
$Ares610;3.897;13.740;150653040615;46.1362457N;003.2101567E;+0324;00.00;005.18;005.23;1111100011111001;183.4*D2
$Ares610;3.904;13.793;150656040615;46.1362419N;003.2101569E;+0324;00.00;005.11;005.17;1111100001110001;183.4*DC
$Ares610;3.905;13.846;150703040615;46.1362419N;003.2101572E;+0324;00.00;005.31;005.40;1111100011111001;183.4*D7
$Ares610;3.910;13.767;150715040615;46.1362381N;003.2101586E;+0324;00.00;005.45;005.55;1111100001110001;183.4*D4
$Ares610;3.928;13.625;150743040615;46.1362305N;003.2101693E;+0324;00.00;005.49;005.46;1111100001111001;183.4*D9
$Ares610;3.899;13.829;150753040615;46.1362267N;003.2101696E;+0324;00.00;005.61;005.56;1111100011110001;183.4*DC
$Ares610;3.918;13.740;150755040615;46.1362190N;003.2101727E;+0324;00.00;005.11;005.04;1111100011111001;183.4*D5

Figure 4 : Extrait du fichier "ARES610_10_06_2015.LOG" enregistré lors d'un épandage avec tonne à lisier

1.3. Collecte des données sur bâtiment d'élevage laitier

La collecte des données de consommations d'électricité a été initiée sur le site de la ferme expérimentale de Derval (44). Cinq compteurs d'électricité avec une sortie en impulsions ont été placés sur les quatre plus gros équipements consommateurs (tank, ballon d'eau chaude, robot de traite, surpresseur) ainsi que sur le bâtiment d'élevage des vaches laitières (bloc-traite, racleur, pompes de transfert de lisier, éclairage, …). La température et l'hygrométrie ont été mesurées dans le bâtiment. Les consommations d'eau du bloc-traite et d'abreuvement des vaches laitières ont été relevées par des compteurs d'eau avec une sortie en impulsions. L'ensemble de ces données a été collectée et centralisée dans un seul et même système d'acquisition. Le même dispositif de collecte des consommations d'électricité et d'eau a été mis en place sur la ferme du Pradel (09) afin d'étendre les références à l'espèce caprine et à la production fermière. Des enregistrements périphériques ont été réalisés sur les quantités de lait (produites, vendues et séparées pour la transformation), le nombre total d'animaux, le nombre en production, le nombre de traites, les quantités de concentrés distribués... Les suivis ont été réalisés sur plusieurs années afin d'évaluer les fluctuations de ces consommations en fonction des évolutions de la production laitière, de la conduite de l'élevage et de l'alimentation des vaches laitières et des chèvres.

Les documents techniques et la notice d'utilisation de la centrale d'acquisition de données sont disponibles pour chacun de ces deux sites d'élevage. Les données stockées dans les centrales d'acquisition sont accessibles par le web au moyen d'un accès sécurisé et protégé.

Au terme de ce projet, on bénéficie donc d'enregistrements pour plusieurs types d'élevage. Ils permettront de valoriser dans l'entrepôt de données des informations portant sur des pas de temps

assez longs (plusieurs années) ou sur des périodes zootechniques particulières (hivernage en bâtiment, versus période de pâturage par exemple).

1.4. Collecte des données d'irrigation

Différents types de matériel d'irrigation ont été pris en considération : pivot, enrouleur et couverture intégrale et différents équipements ont été installés pour réaliser les chaînes d'acquisition des mesures de pression, de position et de consommation (électrique et eau).

Sur l'ensemble des installations d'irrigation suivies au cours du projet, trois l'ont été à l'aide d'une chaîne d'acquisition qui a permis d'acquérir des données précises sur l'ensemble de la campagne d'irrigation. Pour les autres installations, seules des mesures ponctuelles ont été réalisées et permettent d'approcher la consommation. En effet, étant donnée la régularité des consommations pour une même position, un suivi continu n'était pas nécessaire. Sur les trois sites équipés (figures 5, 6 et 7), les données suivantes ont été suivies :

- La consommation d'énergie électrique a été enregistrée au moyen de compteurs d'électricité avec impulsions ;
- L'énergie hydraulique brute a été déduite des mesures de débit et de pression : les pressions ont été enregistrées en continu sur la saison en tête de parcelle et en sortie de station de pompage. Les débits ont été contrôlés périodiquement, de façon à pouvoir caractériser le couple débit-pression relatif à chacune des configurations hydrauliques rencontrées à l'aval de la station de pompage. Enfin, les compteurs d'eau ont été relevés à chaque passage sur l'exploitation. L'ensemble de ces données a fourni les temps de fonctionnement du matériel d'irrigation ainsi qu'indirectement les débits (connaissant les configurations matérielles) ;
- Pour obtenir de façon complète les calendriers d'irrigation au cours de la campagne, les durées et les doses d'eau appliquées ont été notées par parcelle, par culture et par type de matériel. Lorsqu'il s'agit de matériel mobile (enrouleur), la possibilité de repérer les positions irriguées au moyen de GPS a été étudiée, ce qui permettrait d'automatiser entièrement l'acquisition de cette donnée. Dans le cas contraire, la parcelle et la culture irriguée pour chaque poste doivent être notées manuellement.

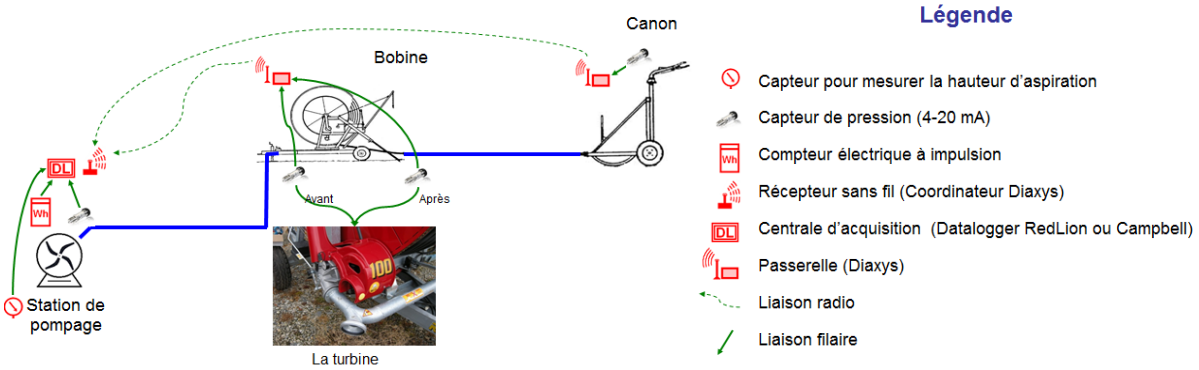


Figure 5 : Dispositif du canon enrouleur

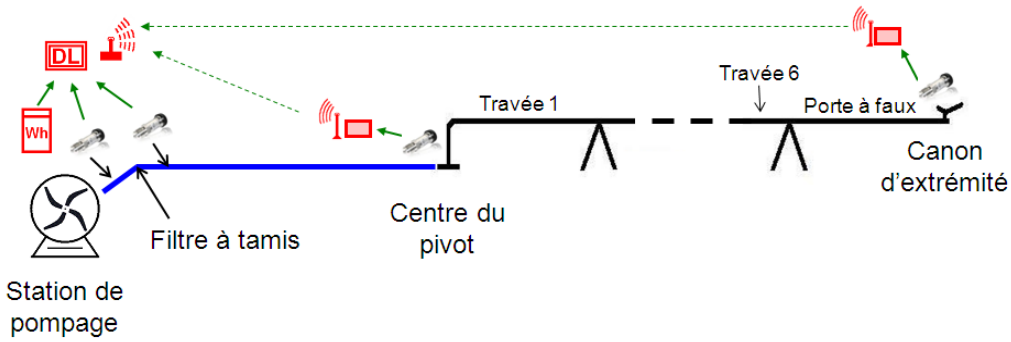


Figure 6 : Dispositif du pivot

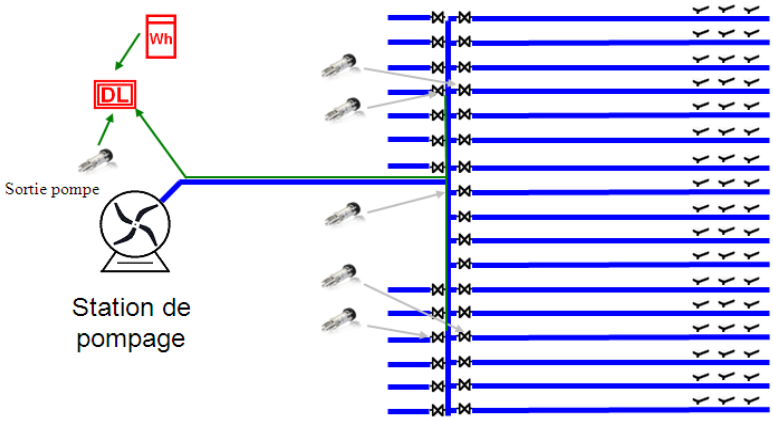


Figure 7 : Dispositif de couverture intégrale

2. Développement du système d'information : base de données et entrepôts

2.1. Format et qualité des données : un préalable à l'échange des données

Le choix a été fait de s'appuyer sur les standards afin de faciliter les échanges de données : utilisation des catégories AGRO-EDI, des listes et catégories de données mises en œuvre dans les outils Systerre® et Dia'terre®. Il sera ainsi plus facile d'être interopérable avec d'autres systèmes.

Le format des données brutes issues des capteurs a été formalisé pour les données de mesures de carburant des équipements mobiles et pour les données issues des bâtiments d'élevage. Pour les mesures de carburant, le fichier standard est celui fournie par la Vbox (Figure 4). Les données issues des capteurs d'irrigation ont été formalisées dans un fichier de type Excel pour permettre une intégration dans la base de données.

Un travail de formalisation des formats de données pour les sites d'élevage en vue d'un transfert vers la base de données a été réalisé. Les étapes de travail sur les données avant leur envoi ont été définies. Ces étapes visent à valider les données, à s'assurer de leur qualité et à les formater pour que les envois issus de sites différents puissent être intégrés à la base de la façon la plus semblable possible. Ce travail de formalisation est applicable à tous les sites d'élevage.

Un fichier standard pour les données d'élevage a été conçu et a donné lieu à la réalisation d'un logiciel nommé ANADATA[1] permettant la vérification de la qualité des données et leur mise en forme selon les procédures définies afin d'assurer leur importation dans la base de données. Une version 2 a été développée pour traiter de façon plus ergonomique et plus fiable les fichiers incomplets qui peuvent être

[1] Logiciel développé par NPN Ingénierie en la personne d'Alain Erignoux.

produits lorsque l'un des capteurs des unités d'acquisition de données est en dysfonctionnement sur un pas de temps assez long.

2.2. Intégration des données brutes dans la base de données

Le projet EnergéTIC avait permis de mettre en évidence des problèmes d'homogénéité des données au moment de calculer les indicateurs et de comparer les informations fournies par les différents sites concernés. Pour résoudre ce problème, les saisies libres dans la base de données ont été limitées. Pour cela, un travail de création et d'intégration de listes de données pérennes communes aux différents types d'exploitations et les plus exhaustives possibles a été indispensable.

La base de données a été restructurée (Figure 9) pour intégrer les données de tous les types d'exploitation : irrigation, culture, élevage et transformation fromagère. Le modèle de cette base et le dictionnaire de données ont fait l'objet d'un rapport technique (André, 2014).

Après validation des listes standardisées de produits, de types de produits, d'opérations techniques, ou encore de types d'équipements, etc., toutes les données pérennes des exploitations fournies par les différents partenaires ont pu être intégrées : celles des parcelles physiques et culturales, des bâtiments, équipements, des lots d'animaux… Cette intégration est à réaliser à chaque campagne et/ou pour chaque ajout de matériel, de lots d'animaux, etc.

Cette nouvelle base de données PostgresSQL permet à la fois une intégration directe des données issues des capteurs des différentes exploitations (par fichiers avec séparateur), ainsi qu'une saisie dans une interface web développée à l'aide du framework Django-Python.

L'intégration des données dans la base est automatisée par un programme développé en Python. Pour l'instant, il est nécessaire de lancer manuellement l'intégration afin de pouvoir vérifier à chaque fichier que les données sont correctes et correctement intégrées. Certaines données des fichiers ne proviennent pas directement des capteurs et sont donc saisies manuellement dans le fichier par l'exploitant : par exemple, le nombre d'animaux traits n'est pas obtenu par capteur.

Les données de la base sont ensuite traitées et intégrées dans l'entrepôt par un système automatisé utilisant un ETL (Extract-Transform-Load).

2.3. Automatisation et fiabilisation de l'outil ETL proposé dans le projet EnergéTIC

Le processus d'ETL de l'entrepôt de données spatiales a été mis en place à partir de la base de données développée dans le projet en utilisant un outil ETL distribué par la Société Talend. Cet outil possède une version libre open source : Talend Open Studio for Data Integration.

Grâce à cet outil, les routines ETL ont été fiabilisées et rendues maintenables.

2.4. Etude des différentes granularités de données et conséquences sur le modèle de l'entrepôt de données

La granularité des données est un facteur important et est liée à l'hétérogénéité des données enregistrées. Elles peuvent être acquises à différentes échelles spatiales (parcelle, exploitation…) ou temporelles (heure, journée...) ou à différents niveaux de représentation (machine, tâche, chantier...). Au sein d'un entrepôt, il est important de savoir gérer ces différentes granularités par rapport aux différents niveaux d'analyse de l'entrepôt afin de pouvoir agréger ou forer les données pour les analyses.

Une méthodologie de prise en compte des données à différentes granularités via l'estimation des données détaillées a été définie dans le projet (Koueya *et al.*, 2014).

Les opérateurs OLAP (OnLine Analytical Processing) intègrent des fonctions d'agrégation qui permettent la visualisation des données à différents niveaux de détail ou granularité. Au niveau des

granularités fines, on retrouve les données détaillées ou micros données, alors que les données agrégées sont retrouvées au niveau des granularités élevées. Les données ou mesures agrégées résultent des calculs (somme, moyenne, etc.) opérés sur les données détaillées. Elles sont souvent stockées pour faciliter la navigation dans les Bases de Données Multidimensionnelles (BDM). Cependant, les valeurs incomplètes sont systématiquement présentes au sein des BDM, et des analyses fondées sur des données incomplètes peuvent être inexactes. Le projet EDEN a donc conduit des recherches intensives sur des bases de données relationnelles, déductives et orientées objet, particulièrement concernant les valeurs manquantes - forme la plus connue d'information incomplète -. Le traitement des valeurs manquantes est couramment réalisé par le biais d'estimation pour des BDM. Les méthodes d'estimation se divisent en deux catégories : les méthodes horizontales et les méthodes descendantes.

Dans le cadre de ce projet, une adaptation des méthodes d'estimation existantes a été réalisée afin de prendre en compte les corrélations et la multigranularité des BDM. Celle-ci repose sur deux points : l'utilisation d'un prédicat de sélection qui aide les fonctions d'estimation à prendre en compte seulement les données sémantiquement utiles en utilisant l'organisation hiérarchique des données dans les BDM. Le second est l'utilisation d'une méthode d'ajustement des valeurs estimées aux mesures agrégées connues. Cette méthodologie se base sur la programmation linéaire pour l'ajustement des données. Une implémentation a été proposée dans une architecture R-OLAP basée sur Mondrian en la couplant avec Map4Decision ou JPivot (en accord avec la nouvelle architecture) et l'outil ETL Talend (aussi utilisé dans le projet).

2.5. Méthodologie de conception facilitée des entrepôts de données

Dans le cadre des entrepôts de données, les problèmes de qualité sont aussi liés à la structure multidimensionnelle qui ne correspond pas aux besoins des utilisateurs. Pour cela, nous nous sommes intéressés aux méthodologies de conception. Une distinction peut être faite sur les méthodologies de conception des entrepôts de données selon les besoins des utilisateurs : les approches axées sur les besoins ("requirement-driven approaches"), les approches axées sur les sources ("source-driven approaches") et les approches mixtes. Toutes ces approches sont basées sur l'hypothèse que les données sources pour alimenter l'entrepôt de données sont connues et disponibles à l'avance. Toutefois, dans certains cas, les données sources ne sont pas disponibles au début du projet, comme c'est le cas ici. Dans ce projet, les données sources sont identifiées et recueillies *a posteriori*, selon les besoins d'information représentés dans le schéma conceptuel.

Les décideurs impliqués dans ce projet non experts en TIC (Technologies de l'Information et de la Communication) ont jugé aussi très difficile d'exprimer leurs besoins d'analyse par des modèles conceptuels. Ceci a rendu le processus de validation des besoins des utilisateurs plus long et incertain. Ce problème a été résolu en leur montrant des exemples de résultats de requêtes OLAP possibles au lieu de travailler sur les schémas pour identifier leurs besoins d'analyses. Nous proposons ainsi une méthodologie assistée par un outil nommé ProtOLAP (Bimonte *et al.*, 2013, 2014) qui permet des tests rapides et fiables pour la validation des schémas des entrepôts de données dans des situations où les compétences en TIC des utilisateurs sont limitées et les données sources ne sont pas disponibles au début de projet.

3. Constitution, analyse et adaptation des indicateurs énergétiques pour leur utilisation sur des exploitations agricoles

Une fois les données collectées et intégrées dans l'entrepôt de données, l'étape suivante a visé à établir des indicateurs utilisables pour le pilotage des exploitations agricoles et définir la méthode à développer pour permettre leur interprétation et donc leur valorisation auprès des exploitations agricoles.

3.1. Choix et classification des indicateurs en fonction des objectifs et des acteurs destinataires

3.1.1 Cultures

Les indicateurs découlent du schéma UML de l'entrepôt de données qui fait apparaître les huit dimensions thématiques du modèle proposé :

- La dimension Productions correspond aux extrants de l'exploitation agricole (grains, légumes, etc.)
- La dimension Produits correspond aux intrants utilisés dans l'exploitation agricole (carburant, engrais, semences, etc.)
- La dimension Equipements correspond aux matériels et équipements utilisés lors des opérations techniques (tracteur, type d'attelage, etc.)
- La dimension Opération technique correspond aux activités culturales (déchaumage, épandage, semis, labour, etc.)
- La dimension Temps correspond aux dimensions temporelles usuelles (heure, jours, mois, année)
- La dimension Campagnes correspond à l'année où ont été relevées les mesures
- La dimension Localisations correspond à l'exploitation agricole et son parcellaire
- La dimension Opérateurs correspond à la personne réalisant l'opération technique.

3.1.2 Elevage

Pour parvenir à une définition aussi complète que possible des indicateurs de résultat pertinents pour les deux types d'élevage (bovins laitiers et caprins) issus de l'entrepôt de données multidimensionnel, il a été envisagé de travailler sur la situation la plus complète, à savoir l'élevage caprin avec transformation fromagère, et de transposer ensuite les acquis à l'élevage bovin lait.

Les types d'indicateurs techniques retenus sont les suivants :

- Les indicateurs zootechniques en entrée et sortie
- Les indicateurs énergétiques
- Les indicateurs de consommation d'eau
- Pour les calculer, le modèle présente huit dimensions thématiques (Bimonte et al. 2014) :
- La dimension Productions correspond aux extrants de l'exploitation agricole (lait, fromage, etc.)
- La dimension Produits correspond aux intrants animaux ou autres utilisés dans l'exploitation agricole (vache, fourrage, électricité, eau, etc.)
- La dimension Equipements correspond aux matériels et équipements utilisés lors des opérations techniques (tank à lait, prérefroidisseur, etc.)
- La dimension Opération technique correspond aux activités d'élevage laitier (traite, nettoyage, abreuvement, etc.)
- La dimension Temps correspond aux dimensions temporelles usuelles (jours, mois, année)
- La dimension Campagnes correspond à l'année d'activités
- La dimension Localisations correspond à l'exploitation agricole et aux lots d'animaux
- La dimension Période correspond à la période zootechnique (pâturage, stabulation, etc.).

3.1.3 Irrigation

Quatre familles d'indicateurs pour le volet irrigation ont été définies permettant de statuer sur la consommation en eau et en énergie et de poser un premier diagnostic technique sur le niveau des installations du matériel (Deumier *et al.,* 2014):

- Les indicateurs agronomiques : quantité d'eau d'irrigation/ha et quintaux/ha. La quantité d'eau d'irrigation est comparée aux besoins théoriques car une surconsommation d'eau entraîne une surconsommation d'énergie. Le niveau de production renseigne sur la performance de l'année, de la parcelle et la technicité de l'irrigant.

- Les indicateurs énergétiques : le ratio kWh par m3 consommé renseigne sur la performance énergétique de l'installation ;

L'entrepôt de données pour le volet irrigation est représenté selon un schéma UML. Cependant, dans le cadre du projet, le calcul des indicateurs n'a pas été valorisé dans l'entrepôt de données et n'a pas donné lieu à l'élaboration de tableaux de bord.

3.2. *Exploitation et valorisation des indicateurs : tableaux de bord et diagnostics*

Une réflexion a été menée pour identifier et sélectionner les formes pertinentes des requêtes préenregistrées qui permettront à un utilisateur de l'entrepôt de données de consulter ses propres résultats et d'ordonner les sorties de façon pertinente sans avoir besoin d'une compétence ou d'une connaissance particulière relative à la structure de l'entrepôt de données.

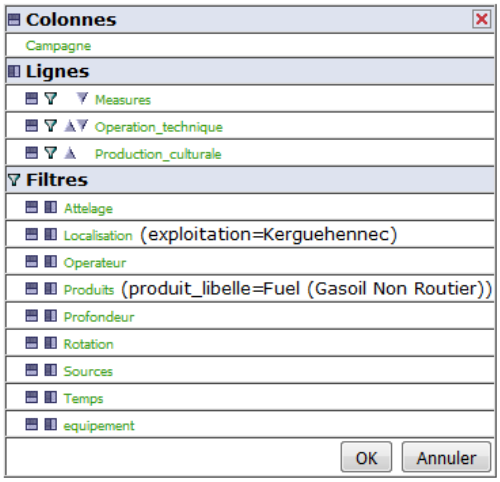
En partenariat avec IRSTEA, ceci a été mis en œuvre pour les volets culture et élevage sur un site web spécifique permettant un accès à partir d'un simple navigateur web aux tableaux croisés dynamiques (Figure 8).

3.2.1 Cultures

Huit types de requêtes ont été programmés pour le volet Cultures :

- L'analyse des consommations de carburant pour l'opération de déchaumage en litre par hectare par parcelle par exploitation et par profondeur et par attelage (tableau comparatif distinguant deux profondeurs différentes par type d'outil).
- L'analyse des consommations de carburant pour les épandages d'engrais organique (tout type de fumier) en litre par quantité d'engrais organique épandu par parcelle et par exploitation.
- L'analyse des consommations de carburant pour l'opération de déchaumage en litre par hectare de surface travaillée et surface culturale et par profondeur (tenant ainsi compte pour une année culturale des différents passages ce qui permet de connaître la consommation globale de carburant sur la parcelle).
- L'analyse des consommations de carburant par parcelle et par opération technique en litre par hectare (à l'échelle de la parcelle, les différentes opérations techniques de la campagne sont répertoriées avec la consommation par hectare, la surface cumulée travaillée (tous passages confondus) et la consommation globale).
- L'analyse globale des consommations de carburant à l'échelle de l'exploitation par type de culture en litre par hectare (évolution de la consommation annuelle par famille de cultures (céréales, légumes, protéagineux…), avec la possibilité de distinguer et de comparer chaque opération culturale).
- L'analyse globale des consommations de carburant à l'échelle de l'exploitation en litre par hectare (évolution annuelle des consommations avec la possibilité de visionner individuellement toutes les opérations culturales).
- La consommation de carburant par culture et par campagne en litre par hectare (comparaison pour une même culture de l'évolution des consommations au fil des ans et entre cultures).

EDEN culture



			Campagne		
Measures	**Operation_technique**	**Production_culturale**	**2011**	**2012**	**2013**
intrant_quantite_w	All Operation_technique	céréale	356.5	550.3	656.2
		culture légumière de plein champ	122.4	423.9	324
		oléagineux		55.9	
	Labour	céréale		85.4	104
		culture légumière de plein champ	41.3	122.8	75.1
cal-sur-campagne	All Operation_technique	céréale	102.74	117.1	145.14
		culture légumière de plein champ	58.95	108.3	61.9
		oléagineux		6.21	
	Labour	céréale	2.05	7.56	6.11
		culture légumière de plein champ	2.3	4.6	4.5
conso l/ha	All Operation_technique	céréale	3.4699	4.6996	4.5212
		culture légumière de plein champ	2.0763	3.9146	5.2342
		oléagineux		9.0016	
	Labour	céréale		11.2963	17.0213
		culture légumière de plein champ	17.9565	26.6957	16.6889
duree_w	All Operation_technique	céréale	23.3	36.8833	39.8667
		culture légumière de plein champ	7.6833	25.2667	23.7667
		oléagineux		3.6333	
	Labour	céréale		5.7	6.1833
		culture légumière de plein champ	2.7	4.9833	4.75

Figure 8 : Exemple d'un type de requête pour l'analyse globale des consommations de carburant à l'échelle de l'exploitation par type de culture exprimées en L/ha

3.2.2 Elevage

Quatre indicateurs précalculés ont été retenus dans l'entrepôt de données :

- Les consommations d'eau en litre et d'électricité en Kwh par litre de lait vendu par année
- Les consommations d'eau en litre et d'électricité en Kwh par litre de lait vendu par année et par période
- La consommation électrique en kWh par vache traite
- La consommation d'eau pour l'abreuvement par type d'animal

3.2.3 Irrigation

En raison du coût de la chaîne d'acquisition, le choix d'aller vers une méthode de diagnostic avec des mesures plus ponctuelles a été privilégié.

La méthode de diagnostic développée pour le volet irrigation est une méthode globale dans le sens où elle porte sur les aspects énergétiques, agronomiques et économiques. L'objectif est ici de pouvoir qualifier en termes de consommation d'eau et de consommation d'énergie une installation d'irrigation. La participation de l'agriculteur est indispensable au bon déroulement de la méthode. La démarche proposée pour le diagnostic comprend cinq phases (Figure 9).

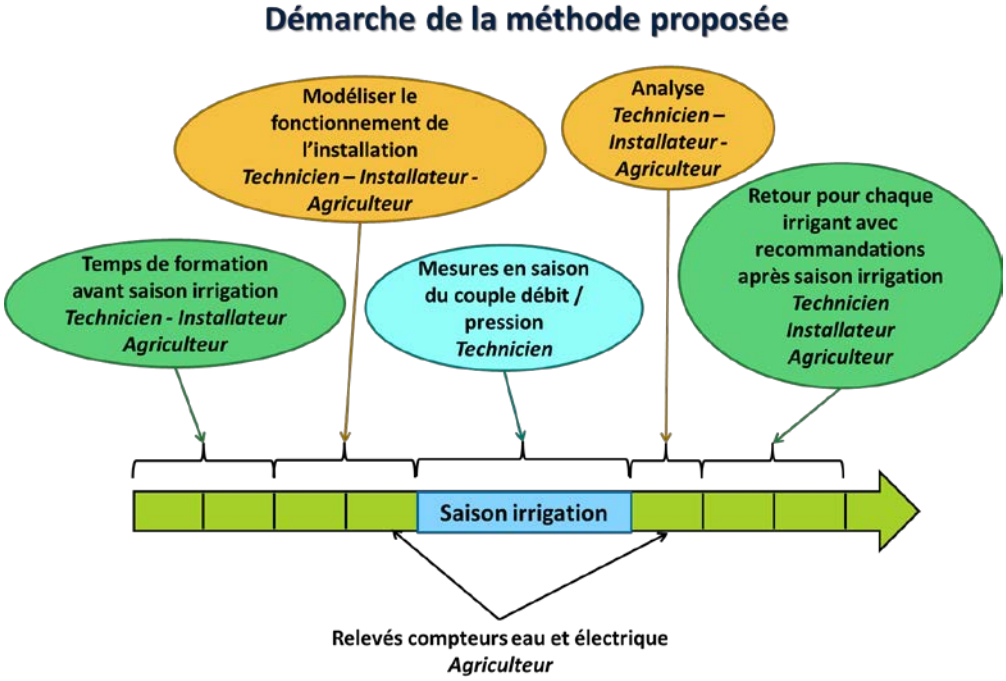


Figure 9 : Phases du diagnostic global

Le temps de formation avant la saison d'irrigation permet de compléter les fiches d'enquêtes développées pour collecter l'ensemble des données nécessaires au diagnostic. Un temps de modélisation du fonctionnement de l'installation est prévu en partenariat entre l'agriculteur, le technicien et l'installateur afin de noter s'il y a des soucis de compréhension du fonctionnement. La phase de récupération de données sur le terrain (mesures débit/pression, relevés des compteurs) se fait pendant la campagne d'irrigation. Les relevés de compteurs (eau et électricité) sont réalisés par l'agriculteur. Le technicien se charge des mesures de débit et de pression. A la fin de la période d'irrigation, le technicien réalise l'analyse complète de l'installation (agronomique, économique, énergétique). La dernière phase du diagnostic consiste en un retour personnalisé auprès de chaque agriculteur des résultats issus de l'étude technique de son installation.

Conclusion

L'intérêt des entrepôts de données spatiales a été démontré dans le cadre du projet pour calculer des indicateurs de pilotage et d'analyse plus stratégiques. Des méthodes de conception des entrepôts de données spatiales originales ont pu être développées également dans ce cadre. Dans la perspective de collecter davantage de données et références, une ouverture des entrepôts de données en lecture sous certaines conditions pourrait permettre à toute exploitation agricole de se situer, de façon anonyme, par rapport aux exploitations du même domaine d'activité, du même système de production ou en fonction des pratiques de production utilisées dans l'objectif d'optimiser ses performances énergétiques.

Le projet a permis d'identifier certains indicateurs importants pour l'irrigation, ce qui permettra à l'avenir d'identifier le matériel strictement nécessaire à l'analyse et ainsi de diminuer le coût de celui-ci. Une méthode de diagnostic simplifiée originale a été mise au point et devrait déboucher à terme sur un outil opérationnel. Son déploiement à l'échelle des régions Midi-Pyrénées et Aquitaine est prévu en 2016.

Références bibliographiques

André G., 2014. Dictionnaire de données BDD EDEN, version 2. Rapport technique IRSTEA, juillet 2014, 36 p.

Bimonte S., Chanet J.P., Capdeville J., Lefrileux Y., 2014. Energetic assessment of dairy activities using OLAP systems. Ageng 2014 International conference on agricultural engineering, Zurich, 6-10 juillet 2014, 10 p.

Bimonte S., Chanet J-P., Capdeville J., Tailleur A., Luciano M., 2014. Une étude sur l'efficacité des méthodes de conception et d'implémentation pour les Entrepôts de Données par une méthodologie « requirement-based » : cas d'étude de la consommation d'énergie en agriculture. 10èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne EDA, Vichy, 5-6 juin 2014, pp. 119-128.

Bimonte S., Nazih H., Kang M., Rizzi S., 2013. ProtOLAP: Rapid OLAP Prototyping with On-Demand Data Supply. ACM thirteenth international workshop on data warehousing and OLAP, San Francisco, 28 octobre 2013, ACM Press.

Chaudhary S., Sorathia V., Laliwala Z., 2004. Architecture of sensor based agricultural information system for effective planning of farm activities. IEEE International Conference on Services Computing, 15-18 septembre 2004, pp. 93-100.

Deumier J.M., Déjean C., Georges J., 2014. Matériel d'irrigation : de nouveaux critères de performance. Perspectives Agricoles 409, 3 p.

Donnat E., Boffety D., Pradel M., Tailleur A., Bimonte S., Chanet J.-P., 2013. Évaluation fine des performances énergétiques des exploitations agricoles par l'utilisation des NTIC. Innovations Agronomiques 30, 267-281.

Iftikhar N., 2009. Integration, aggregation and exchange of farming device data: A high level perspective. ICADIWT '09, Second International Conference on the Applications of Digital Information and Web Technologies, Londres, 4-6 août 2009. pp. 14-19.

Koueya N., Bimonte S., Mephu Nguifo E., 2014. Une nouvelle approche d'estimation pour les entrepôts de données multi-granulaires incomplètes. 10èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne EDA, Vichy, 5-6 juin 2014. pp. 129-144.

Nilakanta S., Scheibe K., Rai A., 2008. Dimensional issues in agricultural data warehouse designs. Computers and Electronics in Agriculture 60, 2, 263-278.

Pradel M., Boffety D., Abt, V., 2009. Vers une évaluation plus fine des performances énergétiques des exploitations agricoles : quels indicateurs et quelles solutions technologiques pour les renseigner ? ECOTECH'09 : Maîtrise de l'énergie à l'échelle de l'exploitation agricole : quelles perspectives technologiques?, Montoldre, pp. 133-141.

Thornsbury S., Davis K., Minton T., 2003. Adding Value to Agricultural Data: A Golden Opportunity. Review of Agricultural Economics 25, 2, 550-568.

Cet article est publié sous la licence Creative Commons (CC BY-NC-ND 3.0)



https://creativecommons.org/licenses/by-nc-nd/3.0/fr/

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « Innovations Agronomiques », la date de sa publication, et son URL)