

# Création d'un serveur de données l'Observatoire du développement rural

Cédric Gendre<sup>1</sup>

**Résumé :** Cet article présente l'organisation et les fonctionnalités de l'Observatoire du développement rural (ODR) au travers de son application web Carto-dynamique développée avec les outils et langages open source.

**Mots clés :** sciences sociales, développement rural, environnement, serveur de données, informatique, cartographie et traitements statistiques.

## Introduction

L'unité de service Observatoire du développement rural du département SAE<sup>2</sup> a développé une application web intitulée **Carto-dynamique** avec les outils issus de l'open source PHP-MySQL. Cette application a pour vocation de rapprocher des données de sources variées centralisées sur un serveur, autour des thématiques de développement rural et de faciliter la production de résultats statistiques cartographiés. Carto-dynamique est intégré aujourd'hui dans une plateforme logicielle plus large dénommée *Observatoire des programmes communautaires de développement rural*.

Ce document est divisé en trois parties ; nous présenterons l'observatoire et ses fonctionnalités, puis l'outil Carto-dynamique développé pour répondre aux objectifs de l'observatoire et, en dernière partie, les trois grands types de données traitées dans l'ODR.



Page d'accueil de l'Observatoire du développement rural

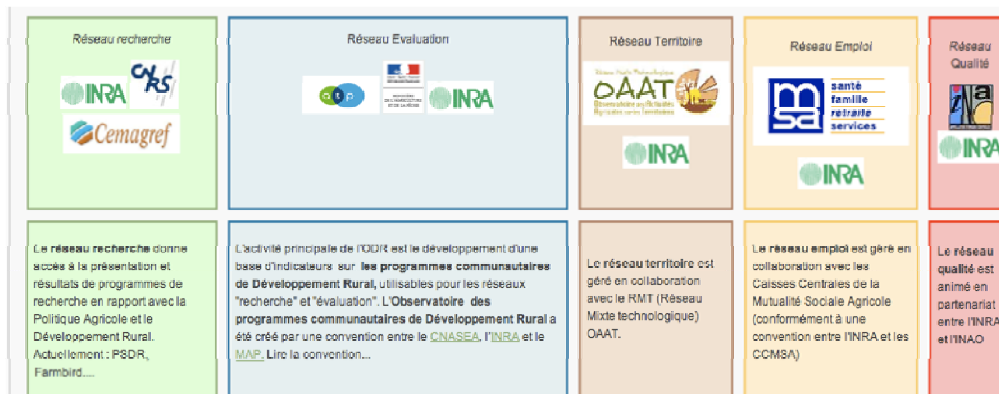
<sup>1</sup>, US0685 ODR -observatoire des programmes communautaires de développement rural – INRA -F-31326 Castanet-Tolosan ☎.05 61 28 53 48 ✉ [cedric.gendre@univ-tlse2.fr](mailto:cedric.gendre@univ-tlse2.fr)

<sup>2</sup> Département des sciences sociales, agriculture et alimentation, économie, environnement

## 1. Observatoire des programmes communautaires de développement rural

L'observatoire des programmes communautaires de développement rural (ODR) est un système d'information collaboratif géré en partenariat. Il a eu pour objectif initial de contribuer à la préparation et au suivi en France des politiques européennes de développement rural, plus particulièrement des mesures financées par les règlements de développement rural (RDR) de 1999 (programmation 2000-2006) et de 2005 (programmation 2007-2013). C'est aujourd'hui un outil coopératif pour la recherche en sciences sociales sur les politiques agricoles, le développement rural et l'agro-environnement.

L'observatoire est régi par une convention cadre et un comité de pilotage. Les partenaires fondateurs sont l'Agence de services et de paiement (ASP, fusion du CNASEA et de l'AUP), l'Institut national de la recherche agronomique (Inra) et le ministère de l'agriculture, alimentation et pêche (MAAP). Il peut accueillir de nouveaux partenaires comme récemment la Caisse centrale de la mutualité sociale agricole (MSA) et des « tiers agréés », le Centre national de la recherche scientifique (CNRS) et le Cemagref susceptibles de fournir ou d'utiliser des données. Il est administré et développé par un chef de projet, Gilles Allaire, et une équipe opérationnelle, l'unité de service ODR, situé au centre Inra de Toulouse-Auzeville<sup>3</sup>.



*Les partenaires officiels de l'ODR*

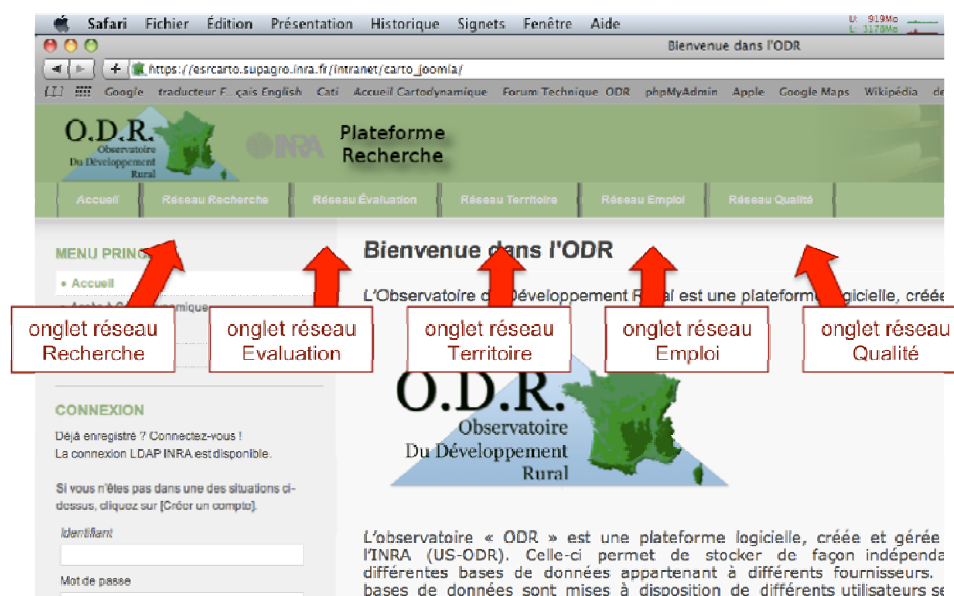
### 1.1 Un ensemble de réseaux gérés en partenariat

Outre les réseaux « évaluation et recherche » répondant aux objectifs initiaux de l'ODR, il héberge aussi un observatoire de l'emploi agricole et rural, géré en partenariat avec la MSA et un projet d'Observatoire des signes officiels de qualité en collaboration avec l'Institut national de l'origine et de la qualité (Inao). L'observatoire rassemble aujourd'hui un ensemble de réseaux sur son site web consultables en ligne :

- **Le réseau recherche** présente des projets menés par des équipes de recherches (Inra, Cemagref, CNRS) utilisant l'observatoire du développement rural et sa plateforme logicielle.

<sup>3</sup> lien web vers l'observatoire : <https://esrcarto.supagro.inra.fr/>

- **Le réseau évaluation** présente les résultats de l'évaluation du second pilier de la politique agricole commune (PAC) conduite avec l'assistance technique de l'équipe ODR.
- **Le réseau territoire** ouvre un portail territoire co-animé par le réseau des observatoires des activités agricoles dans les territoires (OAAT)
- **Le réseau emploi** en collaboration entre l'Inra et la MSA, via une convention cadre, ouvre sur la création de l'observatoire de l'emploi agricole et rural
- **Le réseau qualité** donne accès aux pages de l'Observatoire des impacts territoriaux et agro-environnementaux des signes officiels de qualité

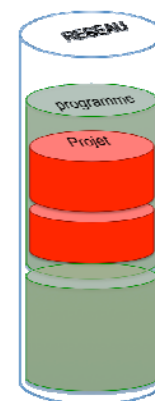


### *Les réseaux accessibles dans l'ODR*

Ces réseaux sont constitués de **programmes**, c'est-à-dire d'espaces de travail et de publications contractualisés avec un ou plusieurs partenaires. À l'intérieur de ces programmes, des utilisateurs particuliers, rédacteurs ou administrateurs de programmes, donnent accès à de la documentation consultable en ligne par les autres membres du réseau, ainsi qu'à des articles informatifs sur les thématiques de recherche développées<sup>4</sup>.

Les membres du réseau peuvent aussi, suivant leurs droits, utiliser la plateforme logicielle pour créer, stocker et rendre disponible à la consultation, des traitements statistiques et cartographiques.

L'utilisation de la plateforme logicielle pour la production de résultats statistiques cartographiés passe par la création de projets.

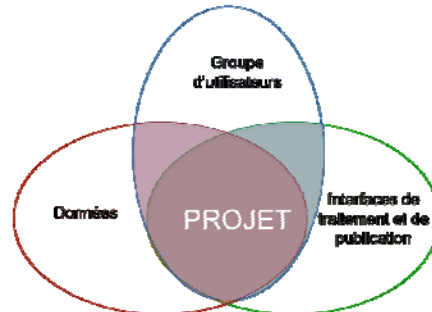


Un **projet** au sens de l'observatoire rassemble :

- un **groupe d'utilisateurs** dont nous verrons qu'il en existe de plusieurs types, autour d'une thématique de recherche ou d'étude, avec à la tête un responsable de projet ;
- des **données administratives ou statistiques, géographiques** déposées dans des zones de stockage de l'ODR et accessibles aux membres du projet ;

<sup>4</sup> Le gestionnaire de contenu utilisé pour gérer dynamiquement les pages des réseaux de l'observatoire est le CMS Joomla distribué sous licence GNU/GPL (gratuit).

– des **interfaces de traitement et de publication** sur internet utilisant des logiciels Open Source (Apache, PHP, MySQL, JOOMLA, MAPSERVEUR, POSTGIS/GRES, R) utilisables via l'ODR qui en garantissent la pérennité.



*Éléments constitutifs d'un projet*

Ces projets sont créés dans l'application **Carto-dynamique** outil central du fonctionnement de l'observatoire.

Les utilisateurs de l'observatoire ont la possibilité de publier certains résultats obtenus dans des **thèmes** du projet regroupés en **dossiers thématiques**.

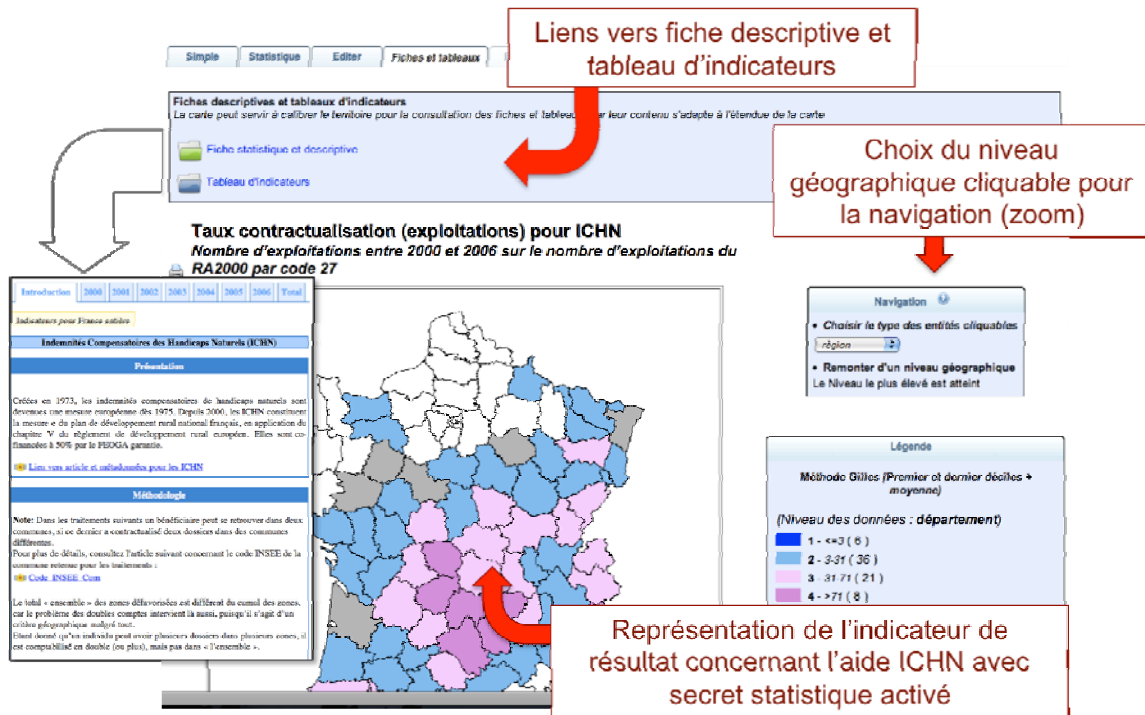
Un **thème** est un ensemble de traitements cartographiés préprogrammés présentant les résultats d'un projet. Il peut contenir, des cartes dynamiques (agrégables à différentes échelles spatiales et pour différents territoires), un ou plusieurs tableaux statistiques dynamiques, un document explicatif du thème.



*La page répertoire des dossiers thématiques*

Ces résultats sont visibles soit dans le projet concerné soit rassemblés dans un programme ce qui les rend accessibles en consultation en dehors du projet initial (cf. écran de consultation des dossiers thématiques ci-dessus).



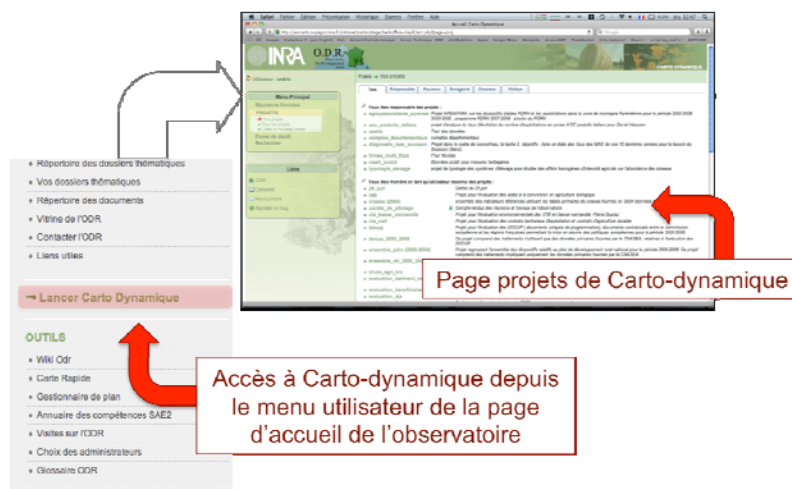


Exemple de thème publié par les membres de l'ODR

## 2. Principes de fonctionnement de Carto-dynamique

La plateforme logicielle de traitement et publication Carto-dynamique est principalement constituée d'un serveur de données utilisant des applications en PHP-MySQL, complété de logiciels open source permettant diverses actions sur les données : POSTGRES-GIS et MAPSERVER pour le traitement de données géolocalisées, R pour des traitements de statistiques descriptives ou de modélisation. Le serveur de données permet :

- le stockage de données repérées par un codage géographique ;
- la production de résultats (cartes et tableaux de synthèse) à partir des données du serveur ;
- la conservation de certains résultats en vue de leur consultation dynamique ultérieure.



Accès à Carto-dynamique depuis le menu utilisateur de la page d'accueil de l'observatoire

Lancer Carto-dynamique

## 2.1 Carto-dynamique : serveur de données pour un système d'information collaboratif

L'application permet une centralisation du stockage des données grâce à une gestion des utilisateurs en différents groupes avec des droits distincts d'accès au serveur et aux variables stockées.

### 2.1.a Droits d'accès des utilisateurs : visiteur ou titulaire

Tout membre de l'observatoire est utilisateur au sens qu'il a besoin d'être authentifié par un login/mot de passe pour entrer dans l'application. Les chercheurs Inra peuvent accéder à l'observatoire en utilisant leur login/mot de passe LDAP, tandis que les autres utilisateurs doivent être enregistrés préalablement dans le système.



*Page d'accueil de l'observatoire*

Les utilisateurs sont divisés en deux grands groupes, les utilisateurs visiteurs et les utilisateurs titulaires.

– **Les utilisateurs visiteurs** ont la possibilité de consulter et de télécharger, sous condition, des résultats publiés dans des programmes et des projets selon leur affiliation à des réseaux. Différentes interfaces sont dédiées à la consultation des résultats sous forme de dossiers thématiques.

Lors d'une première connexion à l'observatoire, l'utilisateur est automatiquement défini comme visiteur. Le statut de visiteur permet seulement la consultation de deux projets génériques (vitrine, zonages). Par la suite, il doit demander une habilitation au responsable d'un autre projet pour en consulter les publications.

– **Les utilisateurs titulaires** peuvent créer des projets ou y participer, déposer et gérer des données, produire, stocker et publier des traitements statistiques et cartographiques. Un

membre titulaire peut donc à la fois être dépositaire de données et/ou développer un projet soit comme responsable soit comme participant. Il existe un statut particulier et important de titulaires qui sont simplement fournisseurs de données dans un projet et ne produisent pas de traitement. Ils sont alors définis comme « donateurs » pour le projet concerné.

|                  |  |
|------------------|--|
| <i>Visiteur</i>  | simple consultation des résultats publiés  |
| <i>Titulaire</i> | utilisateur des données existantes, rapprochement avec ses propres données (importation sur serveur) |

The screenshot shows the INRA O.D.R. web interface. On the left, a navigation menu includes 'Menu Principal', 'Accueil', 'Projets', 'Zones de dépôt', and 'Rechercher'. Below this, a section for 'Evaluation\_finale\_rdr' contains options like 'Description', 'Gérer les membres', 'Créer un responsable', 'Ne retirer du projet', and 'Supprimer le projet'. A red arrow points from the 'Gérer les membres' option to a table of members. Another red arrow points from the 'Projets' menu to a 'Liste des projets' tab. The table lists members with columns for 'Nom', 'Statut', 'Mail', and 'Téléphone'. Red boxes with arrows identify specific roles: 'Membres titulaires responsable du projet' (cedric reconnu), 'Membres titulaires donateurs du projet' (cantalaube reconnu, bdechambre visiteur, jbeschet visiteur, lacomba visiteur), and 'Membres visiteurs du projet accédant aux résultats publiés' (ducroit visiteur, ehrtart visiteur, mitteault visiteur, urbano visiteur, longhi visiteur, leenhardt visiteur). A legend at the bottom right explains the colors: blue for 'Ajouter un nouveau Titulaire au projet' and red for 'Ajouter un nouveau Visiteur au projet'.

| Nom        | Statut               | Mail   | Téléphone |
|------------|----------------------|--------|-----------|
| cedric     | reconnu(groupe res.) | cedric |           |
| alilaire   | responsable          |        |           |
| onasea     | donateur             |        |           |
| soees      | donateur             |        |           |
| msa        | donateur             |        |           |
| cantalaube | reconnu              |        |           |
| bdechambre | visiteur             |        |           |
| jbeschet   | visiteur             |        |           |
| lacomba    | visiteur             | lacom  |           |
| ducroit    | visiteur             | alex   |           |
| ehrtart    | visiteur             | mich   |           |
| mitteault  | visiteur             | franc  |           |
| urbano     | visiteur             |        |           |
| longhi     | visiteur             | marc   |           |
| leenhardt  | visiteur             | sophi  |           |
|            |                      | danie  |           |
|            |                      |        |           |

*Gestion des membres et statuts dans un projet*

### 2.1.b Droits d'accès aux données : réservés ou publiques

Tout fournisseur de données a une gestion fine des droits d'utilisation de ses données dans l'observatoire. Les dépositaires peuvent ainsi définir principalement deux types de droit d'accès aux données : réservé ou publique.

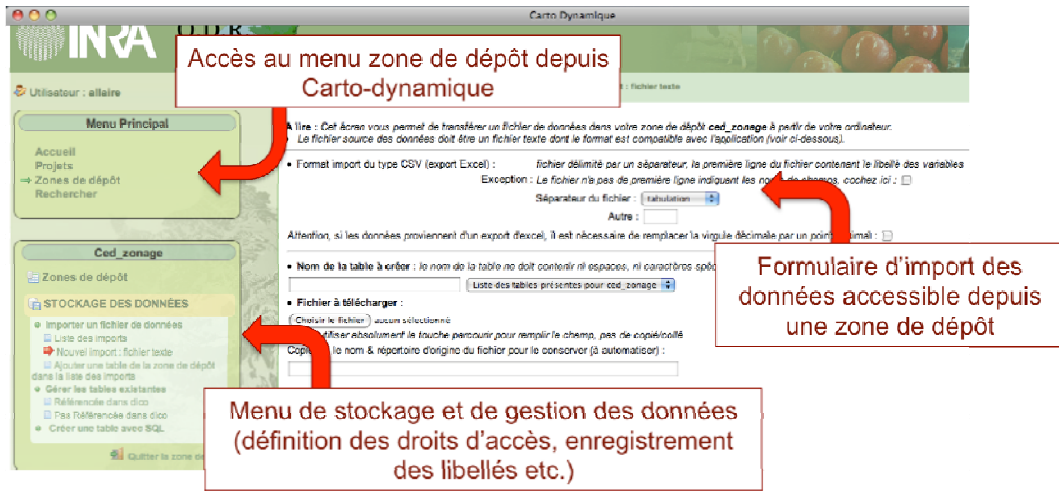
- Les données dites **publiques** sont déposées dans l'ODR et sont accessibles par tout membre enregistré et titulaire sans demande d'autorisation pour leur utilisation.
- Les données **réservées** sont accessibles aux utilisateurs titulaires selon des permissions attribuées par les dépositaires. Ces données peuvent être accessibles dans leur intégralité pour un projet précis ou soumises au secret statistique et à des restrictions selon les cas.

|                          |   |
|--------------------------|---|
| <i>Données publiques</i> | accessibles à tout utilisateur sans restriction.  |
| <i>Données Privées</i>   | possibilité d'ouverture de droits maîtrisés à d'autres utilisateurs par le dépositaire de la donnée |

La gestion des droits sur les données autorise la publication de résultats suffisamment agrégés à un niveau géographique, même si les données brutes traitées sont à des niveaux très fins (par exemple : données individuelles comme unité statistique).

Dans le même ordre d'idée, l'utilisateur titulaire dispose d'un ensemble d'outils pour vérifier les étapes intermédiaires du calcul nécessaire à l'obtention des résultats sur le serveur sans avoir un accès direct aux tables de données brutes.

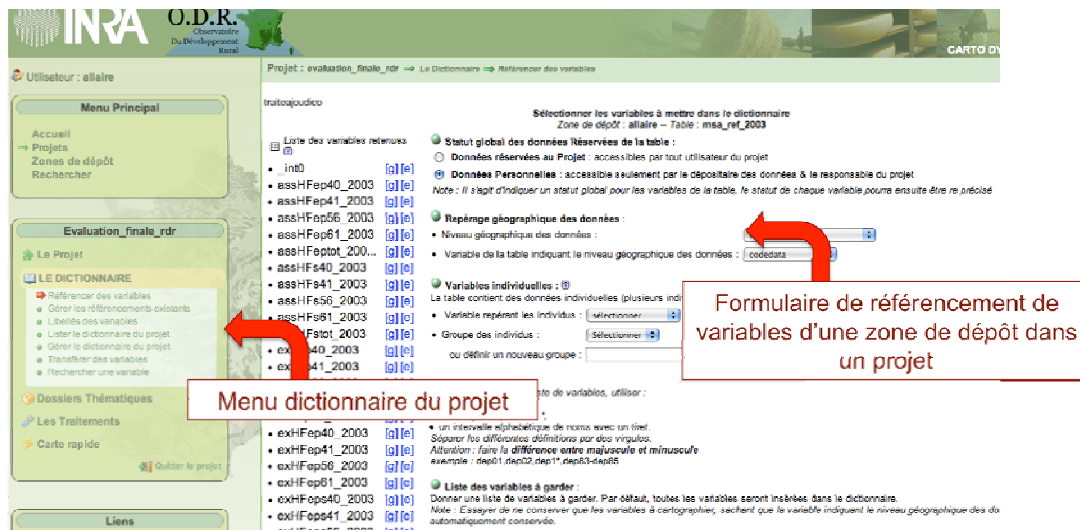
L'importation des données brutes est possible depuis le poste client dans une zone de stockage du serveur dite « **zone de dépôt** » appartenant au dépositaire.



### Zone de dépôt

Pour être disponible dans un projet, les variables réservées doivent être référencées par le propriétaire des données dans le **dictionnaire** du projet. Elles deviennent, sous condition, accessibles pour des traitements aux membres titulaires du projet concerné.

🔔 Rappel : Les données publiques n'ont pas besoin de se référencement.



### Formule de référencement du dictionnaire du projet

Les droits d'utilisation des données sont gérés par le dépositaire des données dans ces deux interfaces web de Carto-dynamique : le formulaire des zones de dépôt et le formulaire du dictionnaire de projet.

## 2.2 Connexion des données à différentes échelles spatiales

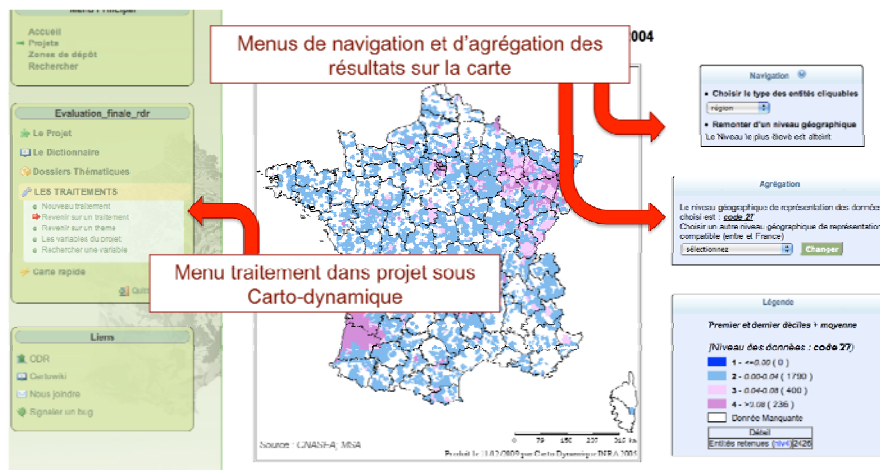
L'objectif de Carto-dynamique est de faciliter le traitement d'informations provenant de sources variées. Ceci est rendu possible par la mise à jour d'une table de correspondance entre tous les codages géographiques disponibles sur le serveur permettant leur rapprochement.

On désignera par « **géocodes** » ce codage géographique. Ce sont les attributs des polygones des fonds cartographiques enregistrés sur le serveur à un moment donné. Les géocodes de base sont ceux des communes présentes au recensement général de la population (RGP 1999).

Le serveur de données repose sur un système de gestion de base de données (SGBD). Le cœur de l'application est une métabase référençant à la fois les données brutes transférées sur le serveur, les utilisateurs et leurs droits, et les fonds de cartes disponibles. Via la métabase, le système peut alors rapprocher toute donnée stockée en tenant compte de la hiérarchie entre géocodes compatibles (exemple agrégation des données communales en données cantonales).

Il ne s'agit pas ici d'un système d'information géographique (SIG) bien que pouvant produire des résultats cartographiés dynamiques. On peut cependant enrichir la cartographie statistique de couches d'habillages Raster (image) pour en améliorer la lecture.

En pratique, cette spécificité permet, lors d'un traitement ou d'une consultation de résultat dans Carto-Dynamique, de changer facilement d'échelle géographique et aussi d'effectuer des traitements de manière automatique qui mélangent des données de différentes échelles géographiques par agrégation - par exemple produire une carte au niveau départemental en utilisant des données communales et cantonales.

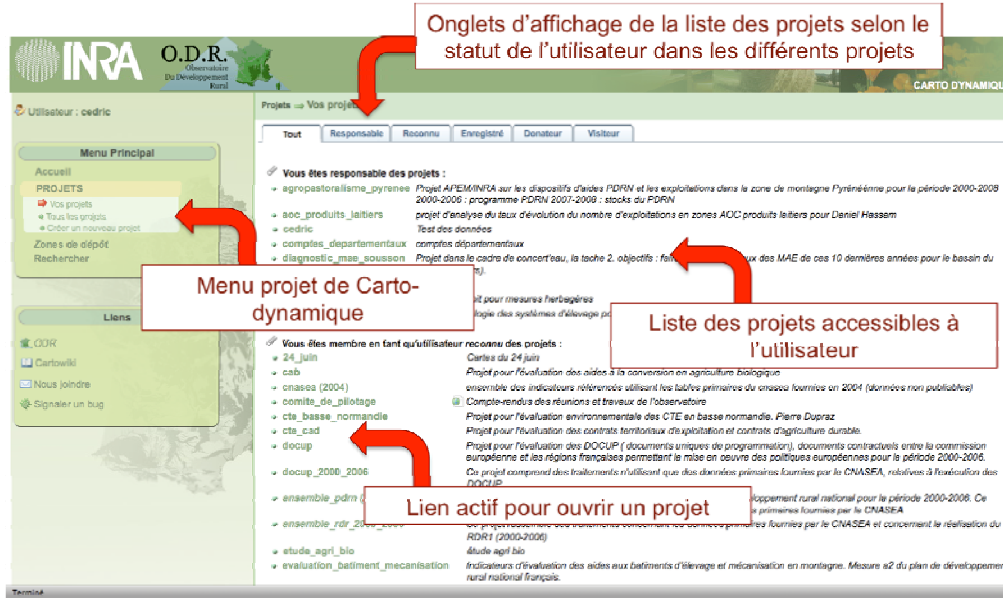


*Les menus de navigation dans l'interface de traitement*

## 2.3 Mise en œuvre pratique du serveur de données.

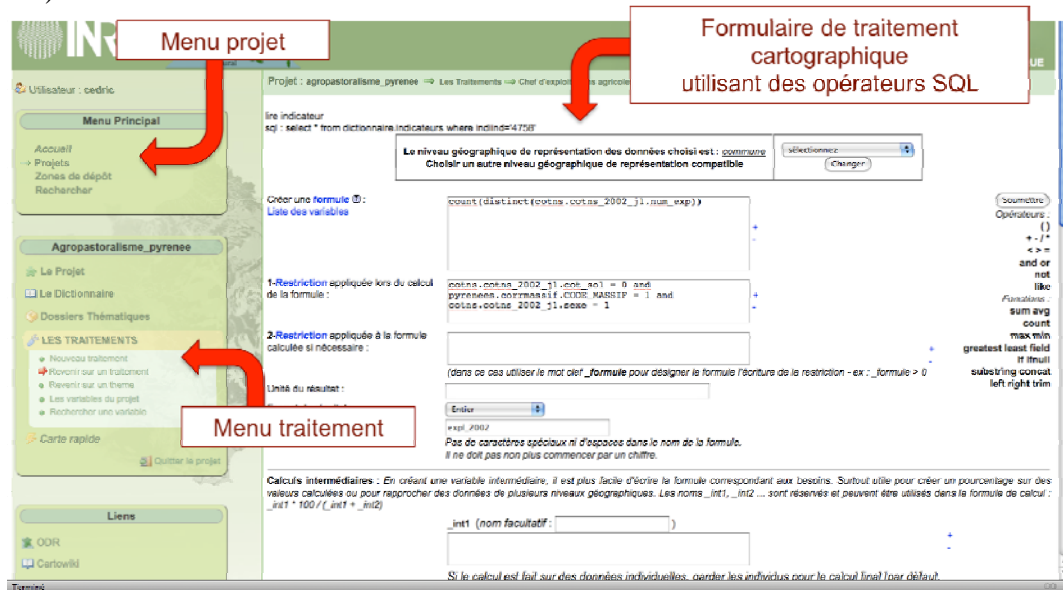
En plus de rapprocher toutes données actuelles ou à venir stockées sur le serveur, le serveur de données offre la possibilité à tout utilisateur titulaire de mémoriser son travail en le sauvegardant dans ses projets. Il peut ainsi revenir facilement et rapidement pour consulter à nouveau ou modifier un traitement existant.





Menu des projets mémorisés dans Carto-dynamique

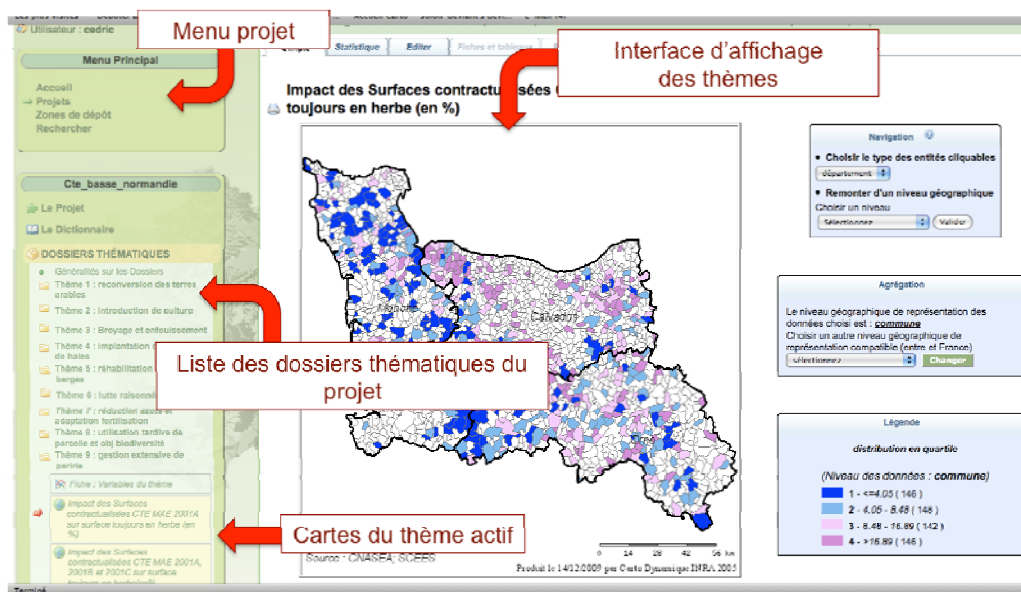
Les traitements sont réalisés dans une interface où il est possible de s'abstraire en partie des difficultés de mise en œuvre du langage SQL, langage de requête sur les tables du serveur et aussi, d'assurer un contrôle des données manquantes, incomplètes et des appariements (jointure) de données.



Menu « traitement » de Carto-dynamique accessible aux membres titulaires du projet

Une fois un traitement obtenu et validé, il peut être mémorisé dans le projet et même publié en ligne à travers l'interface du serveur sous forme de thème.

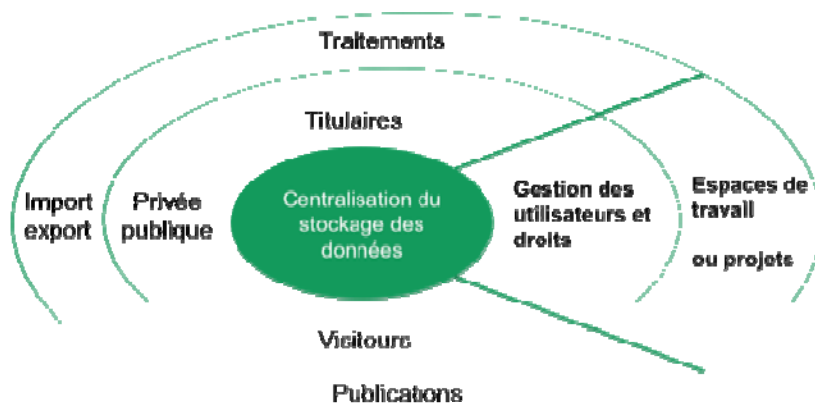




Menu « dossiers thématiques » accessible aux titulaires et visiteurs du projet

Selon les cas, et selon les droits de l'utilisateur, l'interface permet de faire l'export de résultat alliant les données agrégées (indicateurs) et leur géocode, en vue de traitements externes (Excel, logiciels statistiques).

En conclusion, le schéma suivant résume les fonctionnalités d'un serveur de données tel que Carto-dynamique développé pour l'observatoire.



Le cœur du système est la possibilité de centraliser et de rapprocher des données géocodées de différentes sources (par l'implémentation d'une métabase). Pour cela, des droits sont appliqués aux utilisateurs et sur les données et des services sont proposés aux utilisateurs pour traiter, publier et visualiser leurs résultats.

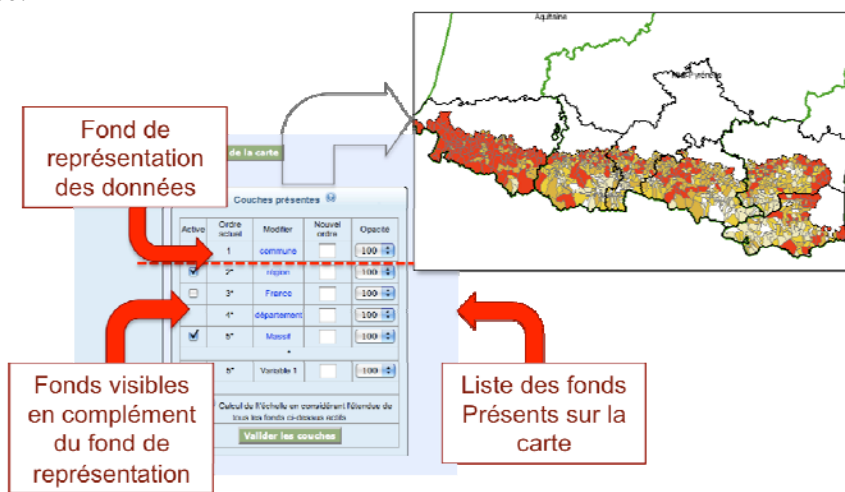
Pour répondre à ses objectifs, l'ODR doit aussi pouvoir proposer un ensemble d'informations lié aux thématiques de développement rural, c'est ce que nous proposons de décrire en dernière partie de cet article.

### 3. Les trois grands types de données disponibles dans l'ODR

L'ODR contient des données très diversifiées et des sources variées se rapportant aux thématiques du développement rural et à l'agro-environnement. Les informations gérées par l'ODR sont de trois types différents. : les fonds de carte, les données géocodées, les couches d'habillages.

#### 3.1 Les fonds de carte

Les données cartographiques ou **fonds de carte** fournissent les contours de différents découpages géographiques, communes et dérivés, bientôt bassins versants et zonages environnementaux ou agricoles. Ces fonds de carte sont le support de la représentation cartographique des résultats statistiques. Leur interdépendance est maintenue dans une table de la métabase.



*Tableau des fonds de carte actifs pour une carte*

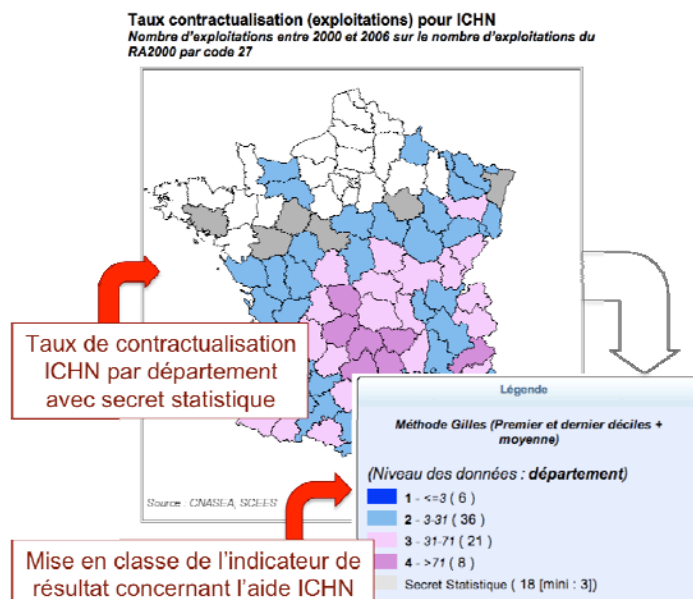
Les fonds de carte sont gérés par l'équipe d'administration. Toutefois il est possible d'ajouter sur demande d'un utilisateur titulaire un nouveau fond de carte créé facilement à partir d'une simple table de correspondance entre un fond existant et la cible (exemple création du fond des petites régions agricoles (PRA) à partir de la liste des communes qui les décrivent).

#### 3.2 Les données géocodées

Les données **géocodées** sont le plus souvent administratives ou statistiques. Ces données doivent obligatoirement posséder un attribut géographique connu du serveur (en lien avec un fond de carte). Pour tout nouvel import de données, une vérification des géocodes est obligatoire, un outil de correction est disponible pour aligner les données géographiques importées sur les données géographiques de référence du serveur (le codage d'une commune peut varier dans le temps)

Ces données, possédant un attribut géographique, permettent la production de résultats, le plus souvent des statistiques descriptives ordinaires sous forme d'indicateurs, et leurs représentations cartographiques. Ces données peuvent être strictement géocodées (une seule valeur par géocode) ou individuelles (plusieurs valeurs pour un même géocode). Dans le cas des données individuelles, un identifiant est nécessaire pour différencier les groupes

d'individus (individus MSA, individus RDR etc.) et il doit être renseigné dans la métabase de Carto-dynamique afin de piloter le rapprochement correct de ces données lors du traitement.



Exemple de traitement de données géocodées

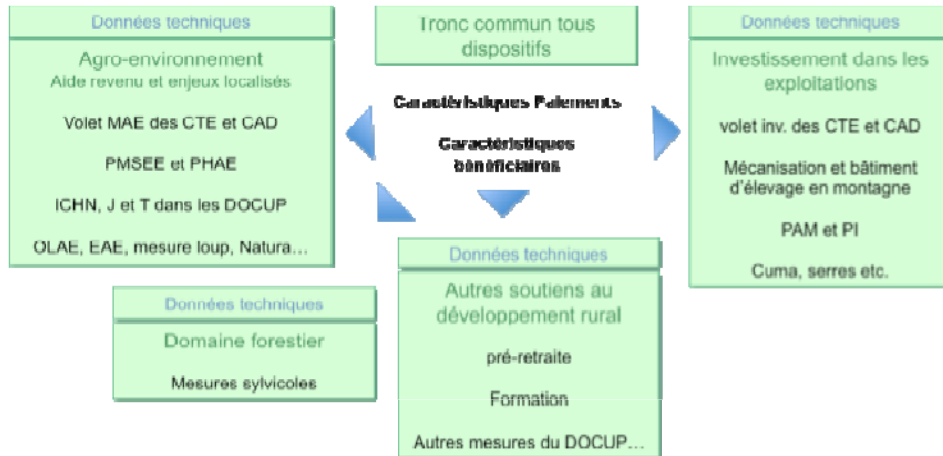
### 3.2.1 Données géocodées gérées par les administrateurs

Majoritairement les données présentes sur le serveur concernent la réalisation des mesures européennes du règlement de développement rural en France, les indications géographiques (AOC, IGP), le nombre et le revenu des exploitants, les structures des exploitations agricoles, l'utilisation du sol etc., ainsi que des données de référence, de contexte ou d'impact dans différents domaines (démographie, revenus et emplois, environnement et systèmes de production)

La gestion des données se rapportant à l'évaluation du second pilier<sup>5</sup> de la PAC est déléguée à l'équipe d'administration par les partenaires de l'ODR.

Les données fournies par l'ASP concernent plus particulièrement l'ensemble des bénéficiaires des aides du Règlement de développement rural pour l'ancien programme (2000-2006) et le programme actuel (2007-2013). Localisées à la commune et agrégables à d'autres niveaux géographiques, ces données regroupent par type d'aide des informations sur les caractéristiques des bénéficiaires, leur localisation, les caractéristiques des actions engagées et les paiements reçus.

<sup>5</sup> Le premier pilier de la PAC concerne le soutien des marchés agricoles, le second pilier de la PAC concerne le développement rural au sens large.



Exemple des dispositifs de RDR1 (2000-2006) présents dans l'ODR

Le statut de ces données n'est pas public, les utilisateurs titulaires doivent déposer une demande auprès du **comité de pilotage** de l'ODR pour obtenir le droit d'utiliser ces données dans un projet Carto-dynamique.

Formulaire de demande de données réservées

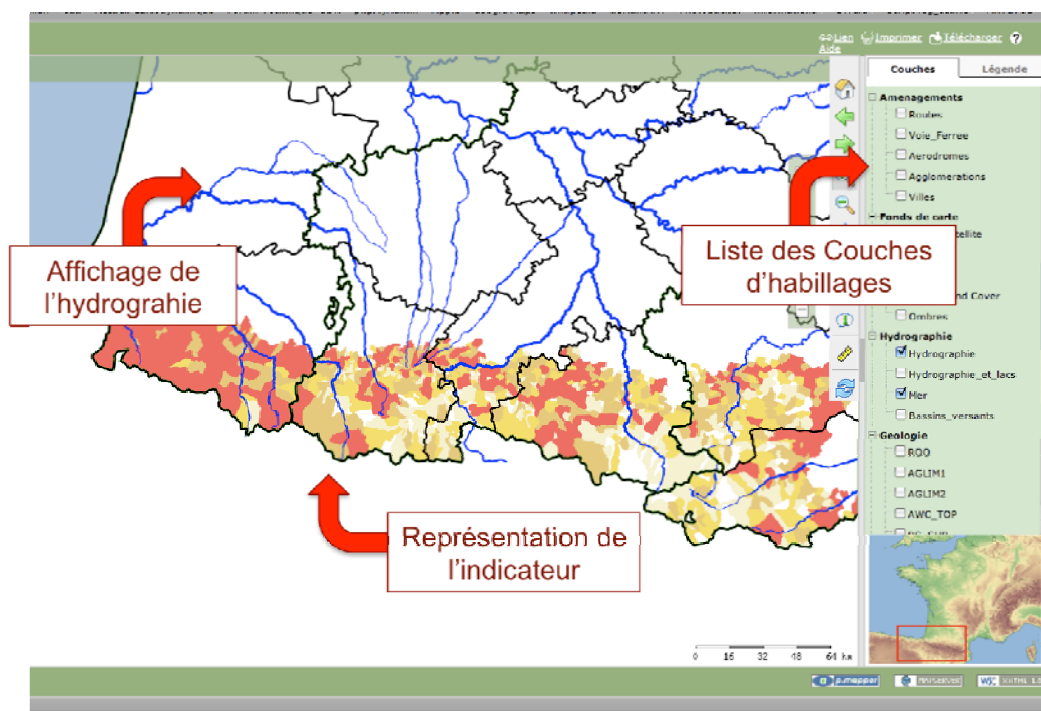
### 3.2.2 Données géocodées gérées par les utilisateurs titulaires

Un des intérêts du serveur est d'offrir la possibilité à tout titulaire de rapprocher ses données personnelles des données déjà présentes sur le serveur auxquelles il a droit. L'utilisateur titulaire peut ainsi transférer ses données (export Excel par exemple ou fichier texte) dans sa zone de dépôt et les référencer dans un projet pour de futurs traitements.

Même sans avoir à déposer et à référencer ses données, tout utilisateur peut avec la fonction « carte rapide » qui permet par copier/coller d'obtenir rapidement un résultat cartographié d'un lot de données ayant un géocode reconnu par le serveur.

### 3.3 Les couches d'habillages

Dernier grand type de données, les données géographiques (raster et vecteur) ou **couches d'habillages** permettent d'ajouter une information complémentaire visuelle à un traitement statistique en superposant son contenu à la carte obtenue (exemple : localisation des cours d'eau, dénivellation du terrain). Ces données n'interagissent pas sur le résultat statistique cartographié.



*Les couches d'habillages*

## Conclusions

Les difficultés à rassembler des données homogènes et de qualité pour l'évaluation en France des politiques européennes de développement rural ont incité l'Inra et ses partenaires à investir dans l'observatoire et son outil Carto-dynamique. Aujourd'hui, cet outil, en constante évolution, favorise l'accessibilité de ces données dans des programmes de recherche. Il offre un ensemble de **services** aux utilisateurs : produire des traitements cartographiques, visualiser et naviguer dans des cartes et des tableaux d'indicateurs dynamiques à différentes échelles spatiales, ainsi que des **produits** : impression/export de carte sous forme d'image, export d'indicateurs en différents formats de fichiers, voire en fichiers géographiques sous format mif/mid.

Le développement de Carto-dynamique a été rendu possible par l'utilisation d'outils libres et les technologies web permettant facilement d'atteindre le public ciblé. Par la centralisation des données sur le serveur et l'incrémentation d'une métabase, Cartodynamique permet de piloter des appariements complexes de données qui demanderaient à l'utilisateur un investissement de temps important, et rend possible le traitement en ligne de grands volumes

d'informations comme les paiements destinés aux bénéficiaires des aides du RDR1 (table avec plus de 4 millions de lignes), ou les engagements des exploitations agricoles dans les mesures agro-environnementales (plus de 2 millions de lignes pour les MAE et MAET).

L'enjeu à venir pour l'observatoire est la normalisation des nomenclatures appliquées aux données et la standardisation des métadonnées dans le cadre d'une démarche de qualité des données.

Enfin, L'ODR a un rôle patrimonial de conservation dans le temps des données servant à produire des résultats de recherche.

**Remerciements** : Ce travail a reçu le soutien financier du ministère de l'Alimentation de l'agriculture et de la pêche ainsi que celui du département SAE2.