

Démarche qualité appliquée à la préparation des données

Cédric Gendre (cedric.gendre@inra.fr)

SAE2, département des sciences économiques et sociales de l'INRA.

Les données (enquêtes, statistiques ou de sources administratives et de gestion) sont une composante stratégique des recherches menées dans le département SAE2 :

- **Données mutualisées** pour les équipes de recherche (mission base de données).
- des **plateformes de traitement de données** telles que le PAP ou l'ODR
- **Données dites « personnelles »** (def. CNRS) à l'usage d'un chercheur, mais pouvant par la suite être pérennisées et partagées.

En 2009, demande du Conseil de Gestion du Département SAE2 :

- mettre en place une démarche qualité afin d'assurer la fiabilité et la traçabilité des processus de traitements des données au sein des unités.

Sources pour les travaux « démarches qualités » :

- Démarche empirique de partage d'expériences, et d'échanges avec la délégation qualité de l'INRA.
- Retours d'expériences d'organismes gestionnaires de données : Eurostat, INSEE, IGN.
- Documents et articles concernant la reproductibilité, le management de l'information.
- Les normes (Dublin Core), guide de bonnes pratiques numériques (CNRS).

L'observatoire du Développement Rural (ODR)

L'ODR est une unité de service INRA du département SAE2. De par ses missions de centralisation et de diffusion de données, elle participe à la mise en œuvre de la démarche qualité.

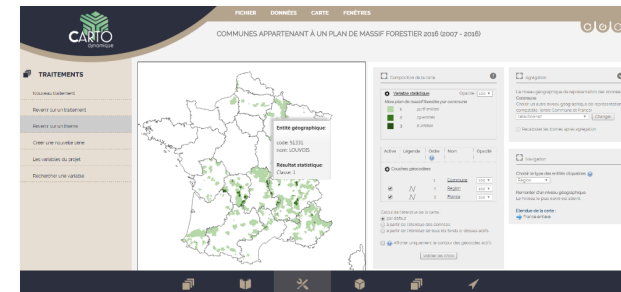
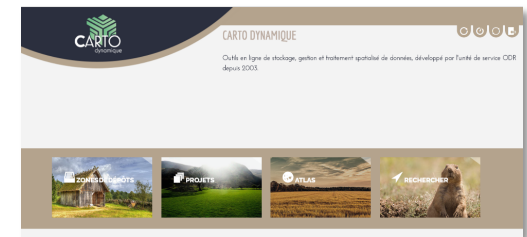
L'ODR contribue :

- ❑ au développement d'une base d'indicateurs pour le suivi et l'évaluation en France des politiques européennes de développement rural (fournisseur ASP)
- ❑ à la centralisation et diffusion du registre parcellaire graphique (RPG) au sein de l'INRA
- ❑ à l'élaboration de tableaux « emplois agricoles » issus des données MSA
- ❑ à la centralisation et gestion des données sur les signes officiels de qualité et de l'origine (SIQO, fournisseur INAO)

L'ODR s'appuie :

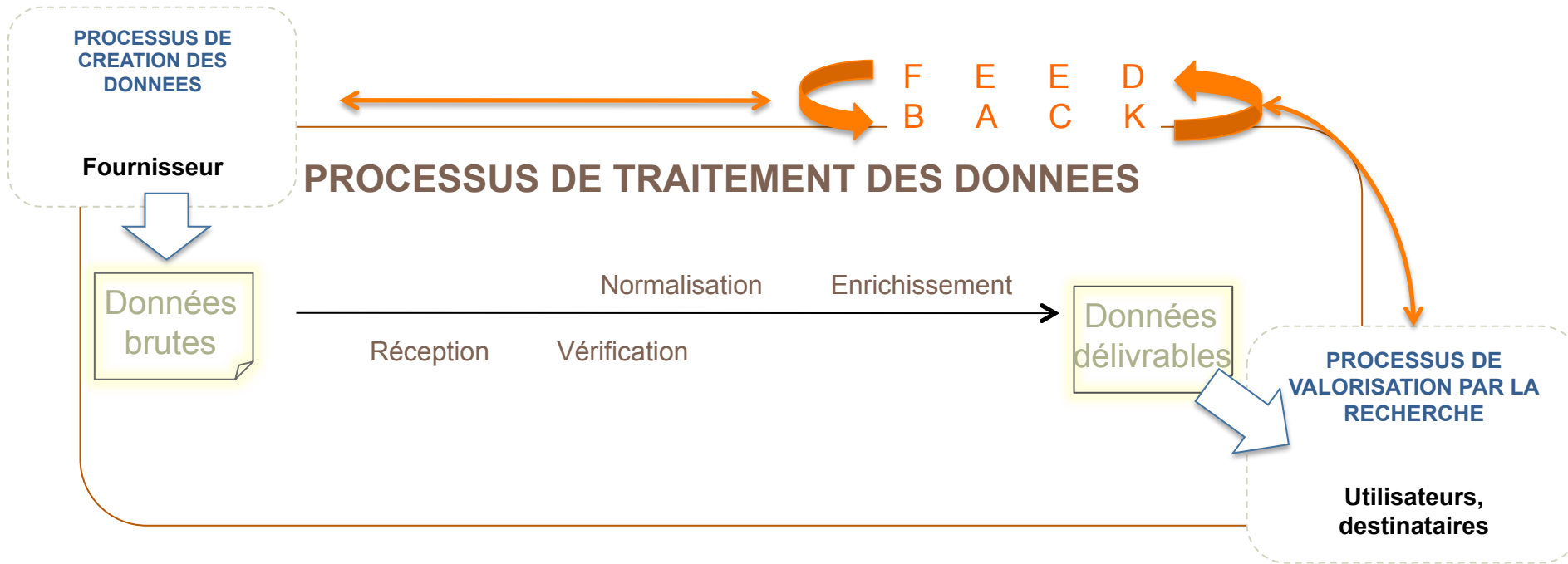
Sur un ensemble organisé de **ressources matériels** (serveurs), **logiciels** (CMS Joomla, Carto-dynamique, MySQL etc.), **personnels** (équipe ODR), **procédures** (d'autorisation d'accès, de traitement des données...), pour centraliser et diffuser les données sous forme d'indicateurs dans des **projets** (*fiches* et *tableaux*) aux équipes de recherches et autres utilisateurs (Ministère, bureau d'étude, etc.).

Remarque : Les données gérées dans l'ODR ne sont pas la propriété de l'ODR (importance de la définition des droits d'accès aux données).



- Périmètre du processus entrant dans la démarche qualité
- Livrables
- Rappel de la charte
- Le guide des bonnes pratiques, une application dans l'ODR
- Pense bête
- Conclusion

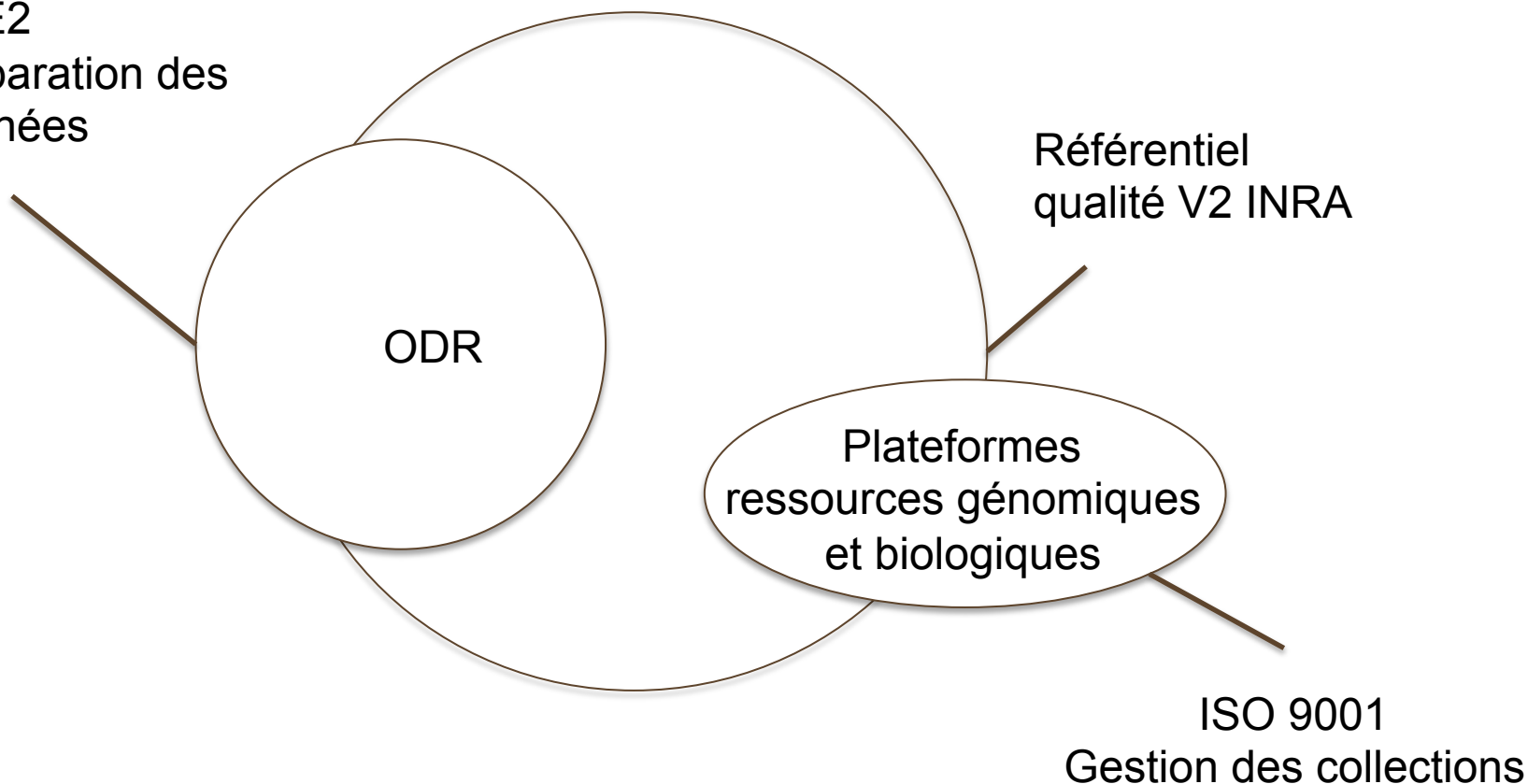
Périmètre du processus entrant dans la démarche



La démarche qualité concerne tous les processus de prétraitements en aval du processus de création des données et en amont de leurs utilisations dans des traitements de recherches. On peut qualifier ces processus sous forme de « chaîne de traitement ».

Politique qualité à l'INRA

Démarche Qualité
SAE2
préparation des
données



- « *Le management de la qualité est efficace quand il sait se faire oublier* » (pleinement intégré au fonctionnement de l'unité).

Pour faciliter la diffusion aux unités du département SAE2

- **Trois documents :**
 - Charte**
 - Guide des bonnes pratiques**
 - Pense-bête**
- **Les objectifs :**
 - Educatifs
 - Harmonisation des pratiques
- **Destinataires :**
 - aux plateformes, équipes, projets ou individus.

La Charte de la démarche qualité

- C'est une introduction aux principes concernant **la fiabilité, la traçabilité et la pérennité** des données.
- C'est un **engagement**
- C'est un **document de référence** qui garantit la mise en place d'une **démarche qualité** au sein d'une entité.

La charte est constituée de **5 grands principes**, appuyés par des exemples de pratiques dans le guide de bonnes pratiques.

Recours aux métadonnées

S'engager à utiliser des normes de métadonnées pour faciliter la diffusion des données

- Ajouter et normaliser des informations sur les données permet une meilleure compréhension de leurs contenus en vue de leurs diffusions
- Destiné aux utilisateurs finaux ou aux systèmes d'informations.

Traçabilité des traitements

S'engager à conserver les scripts et les procédures effectuées sur les jeux de données, à ne pas recourir à des opérations manuelles ne laissant pas de traces.

- Capacité de suivre les états d'une donnée tout au long du cycle de vie.
- « *Moins grave d'avoir une donnée erronée que de ne pas savoir pourquoi elle l'est.* » Kung Fu Panda

Organiser le stockage

S'engager à mettre en place des procédures de stockage et de conservation, à respecter des règles de nommage et d'archivage

- But patrimonial, permettre une utilisation ultérieure par d'autres personnes.
- Que doit-on conserver ?
 - - Tous les éléments associés à une donnée (logs, script, métadonnées, originaux ...).
- Où et comment doit-on les conserver ?
 - Hardware et Software
 - Plusieurs copies
 - Sécurité/Accès

Règles de diffusion

S'engager à appliquer les règles de citation d'auteur, de confidentialité et de conservation des données

- En accord avec les conventions passées avec les propriétaires des données ou instances décisionnelles , respecter :
 - Droits d'auteur.
 - Confidentialité.
 - Règles de diffusion du producteur.

Les principes de la charte

« feedback »

S'engager à faciliter les retours des utilisateurs sur les données, à signaler les erreurs auprès des utilisateurs et à les intégrer si elles sont validées.

- La qualité des données est un processus continu et itératif.

Objectifs

- Définit et décrit des **exemples de pratiques** en accord avec la charte qualité.
- Il s'adresse aux **personnes manipulant des données** en amont du processus de recherche.
- Il est composé de **9 fiches**, chacune décrivant un type de pratiques entrant dans le processus de prétraitement des données.

Le guide des bonnes pratiques

contenu

	Principes de la charte					A quel moment du cycle de vie ?
	Métadonnées	Traçabilité	Stockage	Droit	Feedback	
Fiches disponibles						
Fiche n°1 . Définir et suivre un cahier des charges		X	X	X		<i>Réception/Vérification/Normalisation/Enrichissement</i>
Fiche n°2 . Se fixer des règles de stockage et d'arborescence		X	X			<i>Réception</i>
Fiche n°3 . Utiliser des métadonnées	X	X		X		<i>Tout au long</i>
Fiche n°4 . Suivre l'évolution des programmes et des fichiers		X	X			<i>Tout au long</i>
Fiche n°5 . Se donner des règles de nommage	X	X				<i>Tout au long</i>
Fiche n°6 . Maîtriser l'enchaînement des étapes		X				<i>Tout au long</i>
Fiche n°7 . Organiser le "feedback"		X			X	<i>Distribution</i>
Fiche n°8 . Assurer la sécurité et la sauvegarde			X	X		<i>Réception</i>
Fiche n°9 . Assurer la confidentialité				X		<i>Réception</i>

Définir et suivre un cahier des charges

Objectif : Décrire les attentes des utilisateurs, dans l'idéal

- **Format des données à délivrer**
- **Traitements à effectuer**
- **variables utiles, celles à créer ou modifier**
- **vérifications à effectuer**
- **corrections à apporter.**
- **Définit les métadonnées et les statistiques à fournir**

Au sein de l'ODR

- Organisation de réunions avec les utilisateurs et fournisseurs
- Production d'un cahier des charges (même succinct).
- Mise en place d'outils web facilitant la rédaction et la lecture de cahiers des charges dans l'ODR.

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
X	Organiser le stockage, la conservation
X	Respect des règles de diffusion
	Organiser le feedback

The image shows a screenshot of the ODR (Open Data Request) interface. On the left, a menu for 'PDRH 2010' is visible, with a yellow arrow pointing to the 'Accéder au cahier des charges' option. The main content area displays the 'Edition du cahier des charges : [CDC] PDRH2010' document. The document text includes a date '18/11/2011 : AXE IV (CR du 18/11/2011)' and discusses 'Zonages GAL Leader, deux types : contractualisé et non contractualisé'. It mentions that an analysis of contractualization was performed in Grenoble and that the important part is that the work of comparison was done in Grenoble. It also states that the US-ODR will add GAL indicators (beneficiary commune) to all thematic dossiers in the PDRH evaluation framework, and that the US-ODR with ASP will look at the quality of data from the 6-month evaluation period, emphasizing the need to anticipate needs for technical data on devices.

Définir et suivre un cahier des charges

Autre laboratoire SAE2 (TSE)

Les outils collaboratifs (du type SilverPeas, Agora, Sharepoint ...) facilitent l'élaboration écrite et collective du cahier des charges, ainsi que son suivi.

SilverPeas : Plateforme TCAO (Travail Collaboratif Assisté par Ordinateur).

Site officiel : <http://www.silverpeas.com/>

Site INRA : <https://collaboratif.inra.fr/silverpeas>

Agora : Plateforme TCAO non opensource.

Site officiel : <http://www.online-agera.com/>

Sharepoint : Plateforme TCAO de Microsoft non opensource.

Site officiel : <http://sharepoint.microsoft.com/fr-be/Pages/default.aspx>

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
X	Organiser le stockage, la conservation
X	Respect des règles de diffusion
	Organiser le feedback

Se fixer des règles de stockage

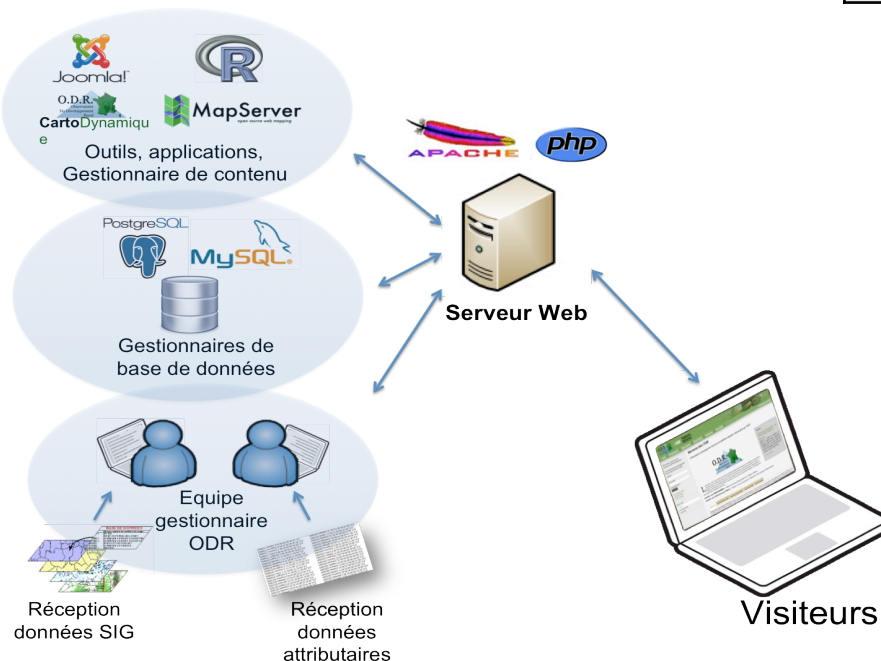
Objectif : Unicité matérielle et temporelle des données, programmes et procédures.

Dans l'ODR

• Intégration et centralisation des données et traitements sur un serveur (SI) qui embarque un système de gestion de base de données (SGBD).

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
X	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback



Se fixer des règles de stockage

Autre laboratoire SAE2 (Grenoble) :

- Générer automatiquement (avec un fichier BAT) l'arborescence de stockage

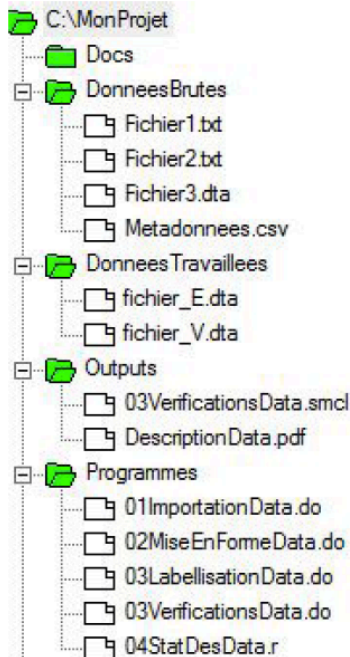


Figure 3 : Exemple d'arborescence

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
X	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback

Script permettant de créer une arborescence.

Le programme « CreateArborescence.bat » (code ci-dessous) permet de créer automatiquement l'arborescence de la Figure 3 :

```
REM "Programme de création des répertoires de base conformément à la démarche qualité"
REM " Ce programme devra être exécuté au niveau ou se situera la racine du projet (l'ensemble des répertoires)"
echo off
cls
REM "On demande le nom du projet qui sera le nom de la racine du projet"
set /p rep= Nom du nouveau projet :
REM "Création du répertoire racine"
mkdir "%rep%"
cd "%rep%"

REM "Création des répertoires de nécessaires à tout traitement suivant (les noms peuvent être adaptés à chacun ci-dessous)"
mkdir "DonneesBrutes"
mkdir "DonneesTravaillees"
mkdir "Programmes"
mkdir "Docs"
mkdir "Outputs"

REM " Un fichier portant la date de création est généré"
set Auto=Creation-%%Date:~-4%%-%%Date:~-7,-5%%-%%Date:~-10,-8%%.txt
echo Création de l'arborescence de ce projet %rep% le %%Date:~-4%%-%%Date:~-7,-5%%-%%Date:~-10,-8%%> %Auto%
cd ..
```

Utiliser des métadonnées

Objectif : décrire les données en leur associant des documents de type dictionnaire et métadonnées normalisées.

Au sein de l'ODR

- Métadonnées pour décrire les jeux de données brutes et jeux d'indicateurs.
- Mise en place de la norme Dublin Core ® et outil web de saisie pour faciliter son utilisation.

Charte

X	Recours aux métadonnées
	Traçabilité des opérations
	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback

Contenu

Titre [Titre] Ensemble des bénéficiaires engagés du PRDH maj janv 2013 Titre développé de la table de données

Sujets ou mot clés [Subject] engagement:PDRH;rdr2 Sujets ou mot clés

Source URL [Source] http://odr.supagro.inra.fr URL

Source enquete [Source] RDR2 ensemble PDRH Enquête ou dispositif de recueil de l'information

Type de ressource [Type] dataset Dataset

Type de données [Type] Données brutes Si pas dans la liste : Doi Doi Doi

Relation [Relation] Remplir ces informations si la table est issue d'une autre table : [Zone de dépôt :] [Table :] issi [Z [T

Description générale [Description] [Liste des variables :] annee_engagement [année d'engagement] bio [bénéficiaire d'une aide en agriculture dans le cadre du RDR1 ou RDR2] code_departement [Code du département du dossier] code_dispositif [Code du dispositif]

Description géographique [Description] commune Description individus

Description individus [Description] bénéficiaires PDRH Description individus

Couverture géographique [Coverage] France Couverture géographique

Couverture temporelle [Coverage] 2007-2012 Couverture temporelle

Propriété Intellectuelle

Créateur [Creator] ASP Fournisseur pour l'ODR, créateur primaire de la donnée (MSA, CNASEA)

Editeur [Publisher] ODR Si pas dans la li.

Contributeur [Contributor] ODR Si pas dans la li.

Droits propriété [Rights] privées Si pas dans la li.

Droits publication [Rights] publiable Si pas dans la li.

Instanciation

Date de l'enquête [Date] 2007-2012 date de l'enquête

Date de mise en ligne [Date] sélectionner : Si pas dans la liste : date de mise en ligne

Format [Format] dataset MySQL

Identifiant Ressource [Identifier] Identification de la ressource : [Zone de dépôt : rdr2_ensemble] [Table : engagement_j9] Nom complet de la table : [Zone de dépôt : nom de la base] [Table : nom de la table]

Langue [Language] français français

SDA

Contacteur l'auteur: Genre cedric

Voir les membres ayant accès

Voir la licence de publication

Consulter les métadonnées

Signaler un bug

Aide sur les dossiers

Valider : Revenir à la liste des imports Valider : Revenir aux metadonnées simplifiées

Utiliser des métadonnées

Autre laboratoire SAE2 (PAP)

- documentation complémentaire de fichier

Charte

X	Recours aux métadonnées
	Traçabilité des opérations
	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback

Fichier	Contenu	Nb. Obs.	Clé
composition.dta	table nutriments (enerkc...)	N=1 342	id : code aliment (codal)
individu.dta	table individus (sexe_ps, v2_age...) pas poids, taille!!!	N=4 079	id : num individu (nomen)
individu_inca2_poids.dta (*)	idem avec poids, taille		
individu_nouvel_envoi.dta (*)	idem avec poids, taille + questionnaire fumeur, nb, nb maladies + infos sport		
consoindiv.dta (*)	questionnaire	N=4 079	id : num individu (nomen)
compl_carnet.dta (*)		N=4 305	
nomenclature.dta	table aliments	N=1 343	id : code aliment (codal) et libellé (libal)
consorep.dta	table des consommations et détails (qté, marque...)	N=541 526	id : numéro ligne (numlig) par individu (nomen), jour (nojour), repas(tyrep)
consorep_nouvel_envoi.dta (*)	idem + var "avecqui" en +		

* 2ème envoi

Figure 4 : Exemple de documentation des fichiers

Suivre l'évolution des fichiers et des programmes

Objectif : Gestion de version, datation des programmes et des fichiers

Dans l'ODR

- Outils web permettant d'exécuter des opérations sur les données (SQL) et de les commenter.
- Sauvegarde pour le suivi d'un log des traitements effectués dans l'ordre avec datation par jeu de données.

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
X	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback

Pense bête (traçabilité) des modifications déjà apportées à la table `rdr2_ensemble.paiement_j9` @ - Exemples (commandes SQL) - cliquer ici-

```
/* indexation facultatif */
#create index num_adm_doss on beneficiaire_j9 (num_adm_doss);
#create index num_adm_doss on paiement_j9 (num_adm_doss);

/* ----- */
/* AJOUT DU CHAMP ID_IND */
/* ----- */

/* ajout des id_ind depuis la table beneficiaire_j9*/

#ALTER TABLE `paiement_j9` ADD `id_ind` VARCHAR(20) NOT NULL COMMENT 'cedric 24/01/12';
#create temporary table rdr2_ensemble.b (index(`num_adm_doss`)) SELECT `num_adm_doss`,`id_ind` FROM `beneficiaire_j9` WHERE 1 group by `num_adm_doss` ;
#update paiement_j9 a left outer join b on a.num_adm_doss = b.num_adm_doss set a.id_ind = b.id_ind where a.num_adm_doss = b.num_adm_doss ;
drop temporary table rdr2_ensemble.b ;

/* ----- */
/* CREATION DE LA TABLE PAIEMENT_PREPARE_[numéro de version] */
/* ----- */
/* nécessite la table paiement_externe_j...b */

#create table paiement_prepare_j9 SELECT `code_programme`,`lib_programme`,`code_mesure`,`lib_mesure`,`code_dispositif`,`lib_dispositif`,`code_presta_dispo`,`lib_presta_dispo`,`code_mesure`,`lib_mesure`,`code_dispositif`,`lib_dispositif`,`code_presta_dispo`,`lib_presta_dispo`,`code_mesure`,`lib_mesure`,`code_dispositif`,`lib_dispositif`,`code_presta_dispo`,`lib_presta_dispo`
```

code_programme
lib_programme
code_mesure
lib_mesure
code_dispositif
lib_dispositif
code_presta_dispo
lib_presta_dispo

Seulement enregistrer la
ou
Exécuter la fonction

Voir la
Voir le log
Statistiques

log des commandes déjà exécutées

Vider la fenêtre Enregistrer le log

Log pour la table `paiement_prepare_j9` de la zone de dépôt `rdr2_ensemble`
Log déjà enregistré

```
[1] - Le : 2013-03-29 16:07:57
[2] - 1 - update rdr2_ensemble.paiement_prepare_j9 set code_action_phae = trim(code_action_phae)
[3] - Modification réussie

[1] - Le : 2013-03-29 15:42:34
[2] - 1 - /* 2009 - 2012 (partiel) */create temporary table rdr2_ensemble.b (index(mae_num_fonctionnel_dossier_osiris)) SELECT `mae_num_fonctionnel_dossier_osiris`, group_concat(distinct(substring(`act_code_action`,7,3)) order by substring(`act_code_action`,7,3) separator ' ') actions FROM rdr2_214a.214a_09_declarations_j9 WHERE 1 group by `mae_num_fonctionnel_dossier_osiris` union all SELECT `mae_num_fonctionnel_dossier_osiris`, group_concat(distinct(substring(`act_code_action`,7,3)) order by substring(`act_code_action`,7,3) separator ' ') actions FROM rdr2_214a.214a_10_declarations_j9 WHERE 1 group by `mae_num_fonctionnel_dossier_osiris` union all SELECT `mae_num_fonctionnel_dossier_osiris`, group_concat(distinct(substring(`act_code_action`,7,3)) order by
```

Suivre l'évolution des fichiers et des programmes

Objectif : Gestion de version, datation des programmes et des fichiers

Autre laboratoire SAE2 (TSE)

- Rappeler en entête les évolutions.
- Ou utilisation de logiciel type « subversion »

```
/*====*/
/* ===== PROGRAMME DATAMAKERNF345 =====*/
en sortie : fichiers d'achats du produit `prod' : p`prod'NF"

/*====*/
/* Ce programme NE TRAITE PAS LES FICHIERS MENAGES (voir MenagesMaker567.do) */
/* DataMakerNF56 créé à partir de DataMakerNF345.do */
/* Version 1.1 29/01/08 sa3 = range.appellation et sa2 = range (pour cohérence
avec les fichiers ancienne formule) via un rename*/
/* Version 1.2 04/02/08 Initialisation des locales (NbAnPasObs AnSansObs) pour
test sur les années manquantes (r(N)>0) */
/* Version 1.3 13/02/2008 Modification du fichier ProduitNF56.dta si produit sans obs. */
/* Version 1.4 7/03/2008 initialisation locale varlistannee_sans_prem */
/* Version 1.41 26/03/2008 drop de an */
/* Version 1.42 22/04/2008 FusionMarques56 intégrée et MN2NF corrigée (moda de Leader Price) */
/* Version 2.0 17/12/2008 Nouvel envoi 2005 et 2006 */
/* Version 2.1 15/06/2009 2005 2006 et 2007 */
/* Version 2.2 8/06/2010 Fichier des caractéristiques produits (Product_Desc_laN.txt) modifié
pour 2005 */
/*
(avant c'était le fichier 1er envoi 2005,
now celui 2ème envoi 2003-2006) */
/* Version 2.3 25/08/2010 Pour nouvelles données 2006n, 2007n et 2008 */
/*
correction labellisation fichier*/
/*Version 3.0 27/07/11 ajout condition pr liste des produits : besoin que l'info panel
soit renseignée*/
/*ex: ds Produits678.dta , pdt 538 n'a pas d'info panel donc
prog plante (V) */
/* Version 3.1 30/08/11 : Changement de pu en Pu et qu en QU (conforme à notre règle
typographique) */
/*Version 4.0 1/09/11 : version générique; chgt de boucle : 1 produit par année*/
/*
: liste des produits à partir du fichier ProduitsNFXXX.dta
DANS la boucle année*/
/*====*/
local version "4.0"
```

Figure 7 : Exemple d'en-tête d'un programme Stata

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
X	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback

Se donner des règles de nommage

Objectifs : Retrouver et identifier la nature d'un fichier (nommage en lien avec le jeu de données, la version) ou d'une variable (origine ou variable crée).

Au sein de l'ODR

- Normaliser les noms de champs avec quelques règles simples. Privilégier des minuscules etc.
- Rester en conformité avec les noms des fichiers et variables d'origine fournis par les producteurs de données (trouver un équilibre avec la compatibilité logicielle).
- Suivre les versions de données grâce à leur nommage.
- Rendre les règles de nommages visibles (ex dans cahier des charges).

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
X	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback

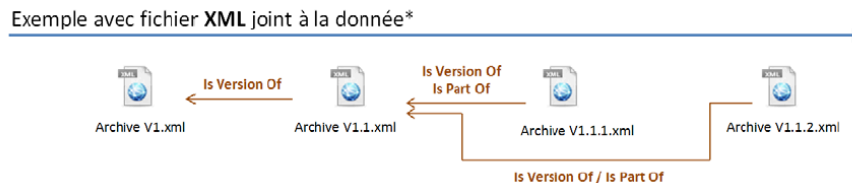
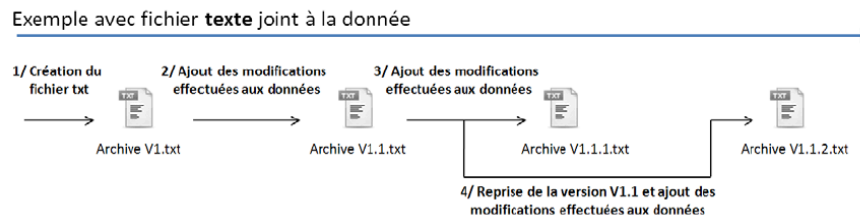
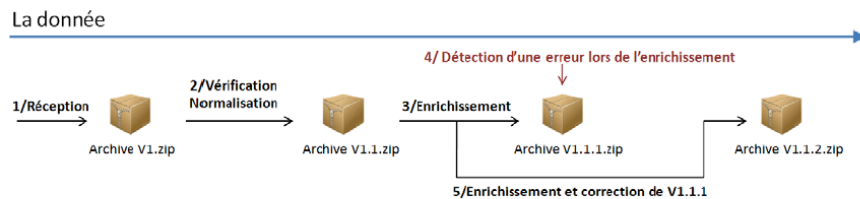
Se donner des règles de nommage

Au Laboratoire SAE2 (Grenoble)

- convention pour identifier les variables d'origines des créées (préfixe etc.)
- ajouter des numéros de versions, homogénéiser les noms des fichiers (données avec documentations associées)

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
X	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback



* Mêmes étapes que le pour le fichier texte, avec en plus les dépendances (notées en orange) au format XML/Dublincore

Figure 11 : XML (eXtensive Markup Language)

Maitriser l'enchaînement des étapes

Objectif : Maitrise de l'ordre des traitements et leurs contenus

Au sein de l'ODR

Regroupement des différents outils de prétraitement dans une seule interface pour faciliter le suivi et leur utilisation :

- **Variable active** : définition des champs référentiels individus et géographique. Définition des droits sur chaque variable (disponibilité, privée/publique).
- **Métadonnées** : saisie des métadonnées.
- **Libellé** : saisie des libellés des variables et informations (type d'unité etc.)
- **Pense bête (log)** : traitements SQL d'ajout ou de correction des données et Log de ces traitements.
- **Comparer** : Normalisation sur une référence (géographique) pour les rapprocher ultérieurement d'autres jeux données ayant la même référence.
- **Statistique** : premier contrôle des données (statistiques descriptives).

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback

No		Nom de la Table	Variables actives	Metadonnées	Libellés des Variables	Pense bête	Comparer les géocodes importés aux géocodes référence	Statistiques de base	Export	Refaire l'import	Suppression en série de la liste	Fichier Source
1	I	Voir la table: caract_generales_j16	Retenir	Créer metadonnees	Créer Libellés	Créer script	Comparer	-Statistiques ou -Utiliser R	Menu export	- Refaire l'import - Depuis sauvegarde	Supprimer	

Maitriser l'enchaînement des étapes

Au laboratoire SAE2 (PAP)

- Utilisation de logiciels de réalisation de diagramme (DIA est libre, ou freemind etc.), avec des conventions simples au niveau des formes et couleurs.

Charte

Recours aux métadonnées	
X	Traçabilité des opérations
	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback

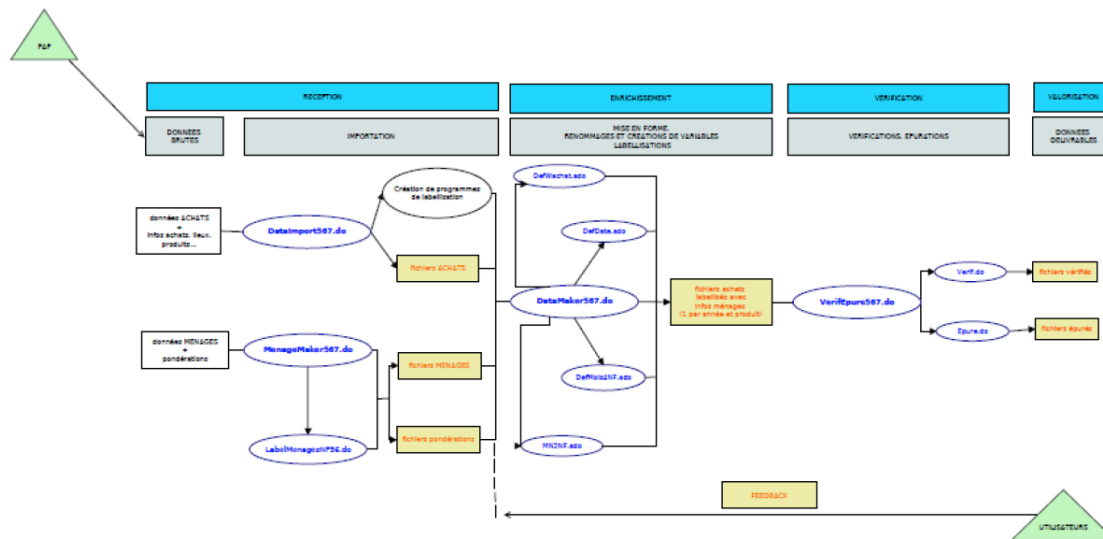


Figure 13 : Exemple de chaîne de traitements illustrée avec DIA

Organiser le feedback

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
	Organiser le stockage, la conservation
	Respect des règles de diffusion
X	Organiser le feedback

Objectif : assurer la bonne communication autour des données et améliorer les traitements.

Au sein de l'ODR

- Contact pour les utilisateurs.
- Utilisation d'outil de suivi des erreurs et corrections.
- Notes de mise à jour dans les publications ODR (fiches, tableaux) permettant de télécharger des jeux d'indicateurs.

Introduction | 2007 | 2008 | 2009

Ensemble des aides du RDR:

Etendue géographique: **France entière**

Présentation

Note de mise à jour (3/05/2013)
 Note concernant les engagements et paiements pour l'axe 3 :
 - Les montants affichés concernent les paiements effectués par l'ASP ainsi que les paiements nationaux dissociés remontés par l'ASP. Ces paiements dissociés sont différés -le paiement FEADER étant versé en premier par l'ASP- et peuvent être...
 - Dans les montants engagés (de l'ASP) les financements dissociés n'apparaissent pas. Ceux-ci sont calculables selon le...
 prochaine mise à jour.

Note de mise à jour (2/05/2013)

Données mise à jour RDR2 janvier 2013, source ASP
 Etendue par défaut, 6 départements pyrénéens. Possibilité de restreindre l'étendue à la zone massif Pyrénées ou aux autres départements...
 Ventilation par région en cours de modification

Note de mise à jour (17/12/2010)
 En 2010 le montant négatif pour les engagements de la PHAE2 (dispositif 214) est dans les faits un désengagement des engagements...
 ne sera visible dans l'ODR qu'à la prochaine mise à jour.

Note de mise à jour (04/11/2010)
 Les résultats correspondent maintenant à la nouvelle version des engagements et des paiements (2007-2010). La version précédente...

mantis
BUG TRACKER

Logged in as: cedric (reporter)

Main | My...

Reporter:	Monitored By:
any	any
Status:	Hide Status:
any	none
Show:	View Status:
50	any
Platform:	OS:
any	any
Navigateur:	any
any	
Note By:	Sort by:
any	Category Ascending

Search [] Apply Filter [] [Advanced Filters]

P	ID	#	Category	Severity	Status	Updated	
	0000161	1	[CMS Joomla] General	minor	resolved (blaise_developpeur)	2013-05-06	un utilis...
	0000157		[Fiches descriptives et Tableaux d'Indicateurs] General	minor	assigned (nledoux)	2013-05-03	fiche de...

Etude stratégique pastoralisme pyrénées

- Editer les règles de publication
- ★ Configurer les profils d'accès
- Contacteur l'auteur: Genre cedric
- Voire les membres ayant accès
- CC Voir la licence de publication
- Consulter les métadonnées
- Signaler un bug**
- Aide sur les dossiers

Organiser le feedback

Autre laboratoire SAE2

- Organiser régulièrement des réunions avec les différents utilisateurs (chercheurs etc.)
- Organiser un stockage centralisé des questions
- Liste de diffusion

Charte

	Recours aux métadonnées
X	Traçabilité des opérations
	Organiser le stockage, la conservation
	Respect des règles de diffusion
X	Organiser le feedback

The screenshot shows the Mantis Bug Tracker interface. At the top, there is a logo for 'mantis BUG TRACKER' and a user login 'cedric (reporter)'. Below the navigation menu, there are filter options for Reporter, Monitored By, Assigned To, Category, Status, Hide Status, Show, View Status, Show Sticky Issues, Changed (hrs), Platform, OS, OS Version, and Taas. A search bar and an 'Apply Filter' button are also visible. The main content area displays a table of issues with columns for ID, #, Category, Severity, Status, and Updated. Two issues are visible: one with ID 0000161 (resolved) and one with ID 0000137 (assigned).

P	ID	#	Category	Severity	Status	Updated
✓	0000161	1	[CMS Joomla] General	minor	resolved (blaise_developpeur)	2013-05-06
□	0000137		[Fiches descriptives et Tableaux d'indicateurs] General	minor	assigned (nledoux)	2013-05-03

Assurer la sauvegarde et la sécurité

Objectif : mettre en place des procédures (système physique/software) pour assurer sécurité et pérennité de toutes les données. Se prémunir des risques de détérioration ou perte de données

Au sein de l'ODR

- serveur sur Montpellier géré en interne INRA (règle de sécurité).
- sauvegarde journalière sur le serveur et hebdomadaire à Toulouse.
- concernant les données : copie de sauvegarde automatique des données brutes au moment de l'import

Charte

	Recours aux métadonnées
	Traçabilité des opérations
X	Organiser le stockage, la conservation
	Respect des règles de diffusion
	Organiser le feedback



Assurer la confidentialité

Objectif : Respect des règles (utilisation/accès/conservation/citation)

Au sein de l'ODR

- Gestion fine des profils utilisateurs et de leurs accès aux données (indicateurs dans l'ODR).
- Outils pour le contrôle du secret statistique, et agrégation des données à des niveaux géographiques exportables.
- Clause d'engagement des utilisateurs avant le téléchargement d'un jeu d'indicateurs (signature numérique).

Charte

Recours aux métadonnées	
	Traçabilité des opérations
X	Organiser le stockage, la conservation
X	Respect des règles de diffusion
	Organiser le feedback

The screenshot shows a web interface for the ODR. A dropdown menu is open for the document 'Etude stratégique pastoralisme pyrénées'. The menu items are: 'Editer les règles de publication', 'Configurer les profils d'accès', 'Contacter l'auteur: Gendre cedric', 'Voir les membres ayant accès', 'Voir la licence de publication', 'Consulter les métadonnées', 'Signaler un bug', and 'Aide sur les dossiers'. Below the menu, a table of indicators is visible. The table has columns for indicator ID, description, and a status letter 'S'. The data rows are:

122A	Amélioration des forêts	S
125A	Desserte forestière	29
125B	Soutien retenues collinaires	S
131	Identification Ovin Caprin	S
132	Qualité alimentaire	77

Demande de confirmation

Vous allez procéder au téléchargement des données présentes sur cette fiche, cette action sera enregistrée sur nos bases de données à des fins statistiques. Vous pourrez à tout moment consulter vos téléchargements dans la rubrique *Vos informations* du menu utilisateur.

Vous avez déjà signé la charte d'engagement (le 2012-03-28 11:03:08)

Votre inscription sur l'ODR est personnelle, votre login et le mot de passe ne doivent pas être transmis.

Il est conseillé de lire attentivement la note méthodologique associée au dossier.

Toute publication utilisant les données téléchargées du site ODR devra mentionner les sources ainsi que le traitement par l'INRA (US-ODR) et les auteurs, tels que figurant sur le document téléchargé. Tout usage commercial des données est interdit.

Les « dossiers thématiques », « fiches descriptives » et « tableaux d'indicateurs » sont des produits de l'ODR, sous la licence « creative commons ». [Voir la licence](#)

Ok

Assurer la confidentialité

Autre laboratoire SAE2

- Préciser en réunion les règles d'utilisation.
- Signature de lettre d'engagement aux utilisateurs (parfois imposée (SSP))
- Vérification ex-post : obligation d'avoir un retour sur la valorisation des données dans les travaux. Les sources exactes et règles d'utilisation doivent être citées.

Charte

	Recours aux métadonnées
	Traçabilité des opérations
X	Organiser le stockage, la conservation
X	Respect des règles de diffusion
	Organiser le feedback

LETTRÉ D'ENGAGEMENT
concernant le respect des textes relatifs à la protection des
données personnelles
.....

Je soussigné XXX

m'engage, dans le cadre de la convention 2015/19 entre l'Organisme demandeur et le ministère de l'agriculture, de l'agroalimentaire et de la forêt, Service de la statistique et de la prospective (SSP),

- à ne transmettre à quiconque les informations qui seront mises à ma disposition,
- à ne pas utiliser les données à d'autres fins que celles qui sont prévues dans le cadre de mes recherches citées en annexe I,
- à m'assurer que les règles de diffusion en vigueur seront respectées dans les publications qui seront faites des travaux et études engagés.

Je déclare :

- avoir été informé que je n'ai qu'un droit d'usage des données fournies qui restent la propriété du Ministère chargé de l'agriculture ;
- et avoir pris connaissance de ce que tout manquement aux obligations mentionnées m'expose :
 - ⇒ à des poursuites au titre des articles 226-13 à 226-24 du code pénal (atteintes aux droits de la personne résultant des fichiers ou des traitements informatiques)
 - ⇒ à des poursuites en responsabilité civile, avec toutes les conséquences pécuniaires que cela comporte au titre des dommages causés.

J'ai bien noté :

Que les obligations mentionnées ci-dessus ne cessent pas à l'expiration de la convention me liant au SSP, et qu'elles ne se substituent pas aux autres dispositions concernant la confidentialité des données, notamment celles régissant le devoir de discrétion (article 26 de la loi du 13 juillet 1983) et celles traitant du délit d'initié (article L465-1 du Code monétaire et financier).

Date et signature

Outil d'auto-évaluation (pense bête)

Outil permettant de s'auto évaluer en repérant les étapes qui mériteraient d'être améliorées. Il permet aussi un suivi des procédures (check-list)

SAE²

Pense-bête pour des bonnes pratiques de gestion et de préparation des données

Ce pense-bête, présenté sous forme de questions chronologiques, se donne pour objectifs de :

- 1) Vérifier que l'on n'a rien oublié dans la démarche de gestion des données
- 2) Donner des repères pour la mise en œuvre de bonnes pratiques permettant d'auto-évaluer nos pratiques de gestion des données.

Les réponses peuvent être les suivantes O : oui, N : non, SO : sans objet, NSP : ne sait pas.

Une progression dans la démarche qualité se mesure notamment par la réduction, voire la disparition, du nombre de réponses NSP à ce pense-bête.

Réception		O	N	SO	NSP
1	Ces données ont-elles été fournies avec un cahier des charges ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Avez-vous reçu des informations complémentaires à ces données ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Type de données	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Sur le Contexte du recueil ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Sur l'échantillon ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

La démarche qualité appliquée à la préparation et à la gestion des données

Première étape vers une démarche plus globale de qualité des données intégrant à l'avenir :

- ❖ D'autres étapes et types de données : données expérimentales (en cours dans dept SAE2).
- ❖ Une démarche vers une certification pour les plateformes de données (de quel type ?).
- ❖ La conception d'un indice de qualité des données (réflexion à venir au sein de l'ODR, ref. gestion de données spatiales).

Les résultats de ces travaux ont été intégrés dans le référentiel qualité V2 de l'INRA sur l'aspect données

La démarche qualité est une garantie à faire valoir, lors de réponses à appel à projet et lors de publication d'articles scientifiques.

Elle est indispensable pour diffuser les données à un public élargi.

Limite

La démarche décrite ici ne garantit pas la qualité interne (intrinsèque) des données en tant que telle, qui dépend des producteurs / fournisseurs de données, mais elle participe à l'amélioration et au suivi de l'évolution de ces données.