



cirad

LA RECHERCHE AGRONOMIQUE
POUR LE DÉVELOPPEMENT

L'inventaire de jeux de données scientifiques: une démarche d'aide à la décision vers l'ouverture de données

Sophie Fortuno - Cirad / UMR Tetis

Open Access & Open Science
Agropolis International, Montpellier
23 octobre 2015

SEMAINE
Internationale
du **LIBRE**  **ACCÈS**

Sommaire

- Un inventaire de jeux de données scientifiques, pourquoi ?
- Objectif et périmètre
- Démarche
- Résultats
- Perspectives

Un inventaire jeux de données scientifiques pourquoi ?

Pour (re)découvrir un capital de données scientifiques produit depuis des années par les unités de recherche du Cirad

capital utile, utilisé, utilisable ; patrimoine passé/courant/à venir

➤ Objectifs :

- Dresser un **premier panorama** et un **premier diagnostic** sur ces données
- **Gérer, valoriser et sécuriser ce capital**
- **Anticiper** les demandes des **tutelles**, des **éditeurs** , se préparer à des **politiques** d'état ou de **bailleurs** (H2020)
- **Aider à la prise de décision** sur :
Favoriser l'ouverture de données vers la communauté scientifique, vers la société...
Comment ? open data ou non, accès sous condition...

Objet et périmètre

- « Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics, avril 2007 »
- "L'Organisation de coopération et de développement économiques (OCDE) définit les « données de la recherche » comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche. "

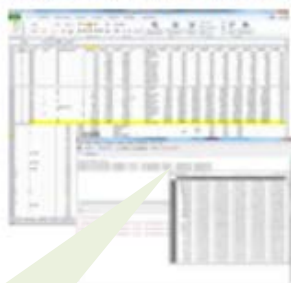
Objet et périmètre

Ressources numériques considérées

Bases de données, SIG

Cartes (format numérique)
cartes de sols, pâturage, historiques...

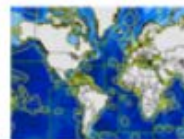
Jeu de données



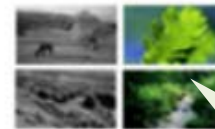
Bases
de données



Cartes



Photos



Exemples

Collection de photos
(numérique/ à
numériser ?)

Tableaux de données :
texte, excel, csv ...

Images



Audio



Vidéos



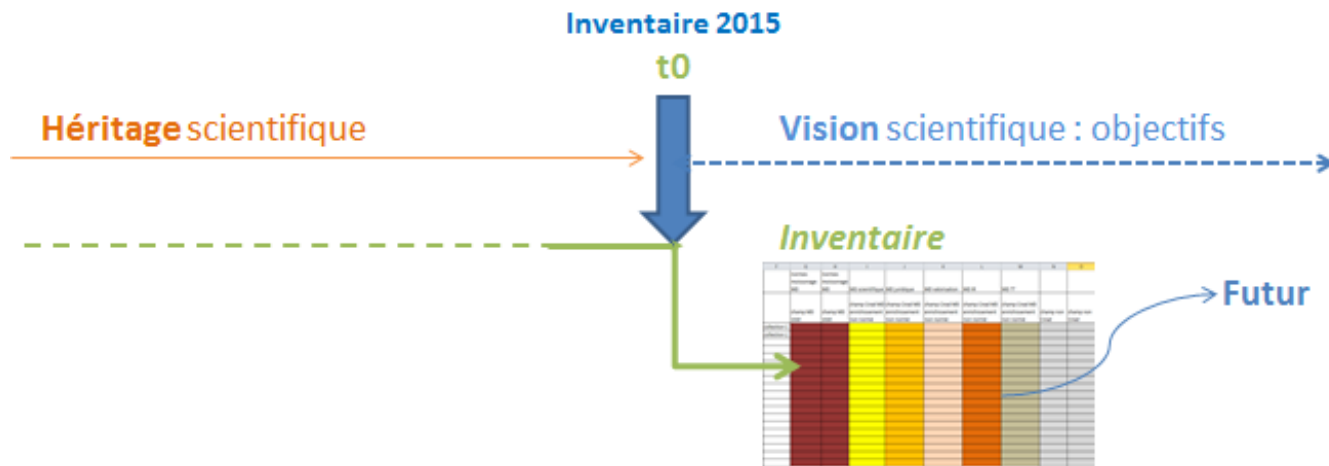
Film,
vidéo virtuelle, cartographie
dynamique etc.

Image virtuelle générée, document graphique...

Enregistrement numérique
[/analogique]

Objet et périmètre

*Fenêtre temporelle inventaire
unité de recherche*



Avril 15

Signalement de jeux de données en continu >>>

Unités pilotes

Démarche de mise en œuvre de l'inventaire

- 1) Identifier l'ensemble des jeux de données d'intérêt scientifique, in situ (UR) et ex situ (en partenariat) et/ou présentant un intérêt patrimonial - *UR pilotes* -
- 2) Décrire leurs caractéristiques

*Description à l'aide de métadonnées (MD)
Choix DublinCore + MD Cirad + MD qualitatives*

- 3) Etablir des notices descriptives des jeux de données
- 4) Valider les jeux de données ayant un intérêt scientifique/patrimonial, évaluer leur vulnérabilité et les éventuels besoins de protection
- 5) Etablir un premier bilan des jeux de données validés (*famille/thématiques/géographie/temporalité...*)
- 6) Vers une généralisation de la démarche...

Résultats

- Un base de données **catalogue d'inventaire**
- Un **annuaire web** de jeux de données consultable en intranet

Rechercher par mots clés, naviguer par critères

Annuaire des données scientifiques
Prototype

(mot clé : pays, lieu, thème, filière, format, ...)
aliment

Rechercher

Rechercher Résultats 1 - 10 sur environ 33 pour **aliment**. Recherche effectuée en 0.03 secondes.

Naviguer

Famille

Essais/Experimentations /Analyses laboratoire (mesures instruments...) (22)

Informatique scientifique originale (logiciels, modèles, scripts...) (15)

Enquêtes (socio éco, épidémiologie, zootechnie, pratiques culturelles...) (9)

Séries temporelles (cohortes + séries chrono + mesures répétées) (9)

Ressources Genétiques/Collections biologiques (8)

Données Omiques (3)

Thématique

Qualité et sûreté des aliments et des produits tropicaux (21)

Echanges, marchés, labels,

Base de données commune et partagée sur la caractérisation des ressources (site d'étude, parcelle sol, animal, aliments ...)

Résumé Base de données existantes et nouvelles mises en commun par les 24 partenaires du projet ARChE_Net. Plus de 10 000 enregistrements. Base en cours de remplissage.

Auteurs Paulo SALGADO ; Emmanuel TILLARD ; Mathieu VIGNE ; Yvane MARBLE

Couverture temporelle 2000-

Pays d'exécution Comores, Australie, Réunion...

Géographie d'intérêt Portée générale

Mots clés Ruminants ; Adaptation systèmes production ; Changements globaux ; Développement d'outils ; Echange de compétences ; Formation ; Océan Indien

Famille de données Données Omiques
Essais/Experimentations/Analyses laboratoire (mesures instruments...)
Observations de terrain (relevés, photos, imagerie satellite, film...)
Enquêtes (socio éco, épidémiologie, zootechnie, pratiques culturelles...)

Lien d'accès http://www.arche-net.org/archenet_db/

Format MYSQL

Base de données commune et partagée sur la caractérisation des ressources (aliments, rendement biomasse, rendement ...)

Résumé Base de données nouvelles du projet BVPI SEHR. Environ 1 000 enregistrements.

Auteurs Paulo SALGADO

Couverture temporelle 2010-2013

Pays d'exécution Madagascar

Thématique Qualité et sûreté des aliments et des produits tropicaux (21)

Guide de saisie détaillée des champs

DESCRIPTION DU CONTEXTE DE LA RESSOURCE	5
1. DATE DE MISE A JOUR	5
2. NOM DU PROJET	5
3. DESCRIPTION DU PROJET	5
4. DEPARTEMENT	5
5. UNITE	5
6. UMR : ORGANISME PARTENAIRE DECLARANT	5
7. RESPONSABLE PROJET	6
8. DATE DE DEBUT (PROJET/RESSOURCE)	6
9. DATE DE FIN (PROJET/RESSOURCE)	6
10. GEOGRAPHIE D'INTERET	6
11. PAYS D'EXECUTION	6
PRESENTATION DE LA RESSOURCE	7
12. TYPE DE RESSOURCE	7
13. NOM DE LA RESSOURCE	7
14. TITRE DE LA RESSOURCE EN FRANÇAIS	7
15. RESUME SUR LE CONTENU DE LA RESSOURCE EN FRANÇAIS	8
16. LANGUE DU CONTENU	8
17. LIEU(X) VILLE, VILLAGE	8
CARACTERISATION DE LA RESSOURCE	8
18. THEMATIQUES CIRAD	8

PROJET PATRIMOINE NUMERIQUE SCIENTIFIQUE

- Une méthodologie
- Une organisation
- Un guide inventaire
- Un outillage informatique
- Un accompagnement (groupes de travail, ateliers...)

Perspectives

■ Mise à jour en continu de l'inventaire

L'inventaire des jeux de données est un **inventaire continu** (modification, retrait, ajout de notices descriptives...).

L'inventaire continu vise ainsi à garantir des informations fiables et actualisées

■ Valorisation des données

Etudier le(s) trajectoire(s) de réutilisation possible(s) pour ces jeux de données, pour quel public cible?

Enseignement | Recherche | Partenariat | Privé | Citoyen

OPEN DATA | ACCES SOUS CONDITION | EMBARGO

SANCTUARISATION...

Des questions ...





Retrouvez toutes les informations concernant les photos présentes dans ce document, ainsi que l'ensemble des images de la photothèque du Cirad auprès du Pôle Image du Cirad

polimage@cirad.fr