

Partager les données pour créer de la valeur en agriculture

La vision d'IRSTEA

Véronique BELLON-MAUREL

Directrice du Département
ECOTECHNOLOGIES

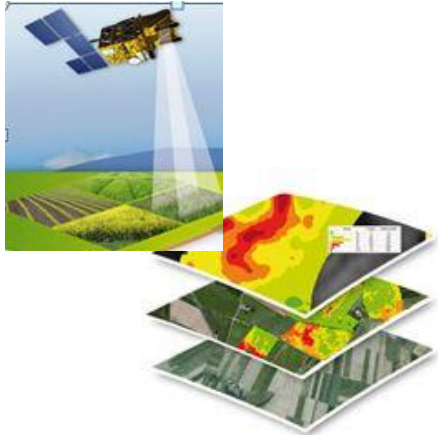


Pour mieux
affirmer
ses missions,
le Cemagref
devient Irstea

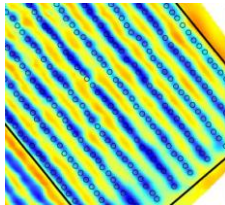
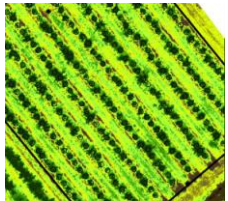


www.irstea.fr

Des données de plus en plus massives



- Données de la ferme: enregistrements automatiques, saisies manuelles
- Données des prestataires de services, données agrégées
- Données des centres de recherche et d'essais => GEOSUD et pôle THEIA (Spot, Pleiades, Sentinel, radar...)
- Données sur l'environnement (Copernicus)
- (Données économiques et génomiques)





**Est-on en train de créer
un BIG DATA AGRICOLE?**



Les « 3 V » du big data

Volume : tera (10^{12}) et zetta (10^{21}) octets.

Variété : croisant des données différentes et auparavant incompatibles, extrêmement **hétérogènes**, à la fois quantitatives et qualitatives

Vélocité



Le Big Data agricole est-il Big ?

« Facebook génère plus de données en deux jours, que nous en un an. Avec 10 millions d'hectares, nous commençons tout juste à avoir besoin de ces technologies »

Stéphane Marcel, Pdg de Smag (Agra-Presse, Juillet 2015)

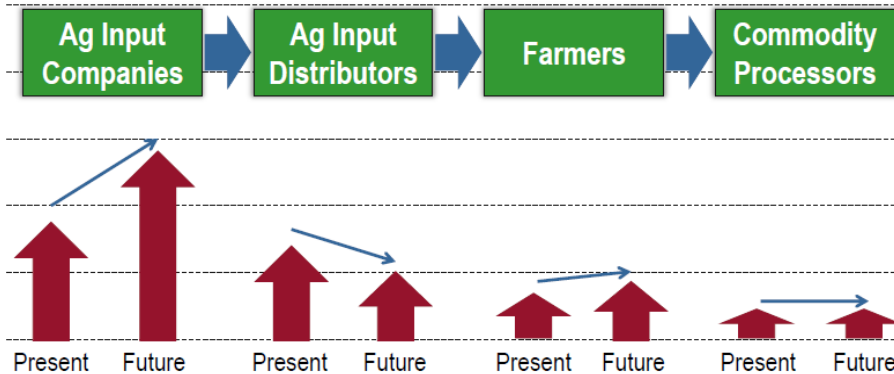
Le big data agricole est-il vraiment big ?

Qq chiffres issus de <http://soilbigdata.org/sites/soil/media/471.pdf>

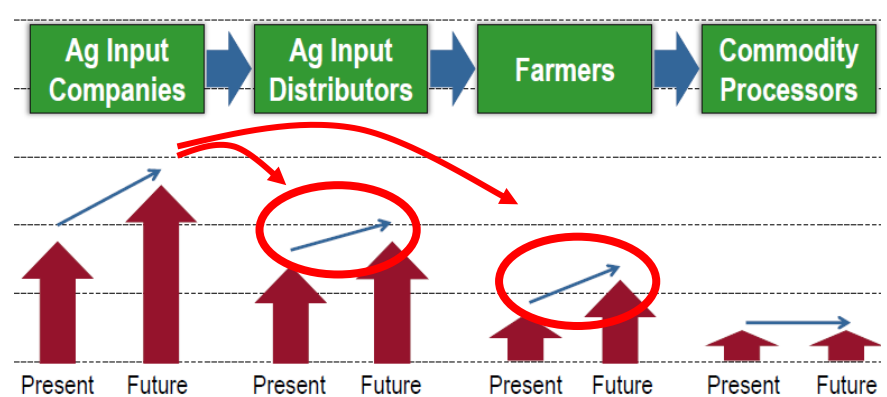
- Exemple d'imagerie drones (Buschermohle, U of Tennessee)
Environ: 7-12 MB/ha.
- Autres capteurs et données de prescription (Shearer, tOSU)
 - Spraying 0.7 MB/ha
 - Planting 13 MB/ha
 - Yield data 10 MB/ha
 - Soil /Fertility Data 1.6 MB/ha
 - Prescription files 0.02 MB/ha

vs tera (10^{12}) et zetta (10^{21}) octets.

Valoriser ces données, quelles options?



Multinationales



Agriculteurs

Répartition de la valeur ajoutée selon le type de gouvernance.

Source : The Hale Group & LSC

Ex MONSANTO: 20 – 25 Milliards/ an
(sur 400 Millions d'ha)

Mettre en place une gouvernance par les agriculteurs?

Portail de données agricoles pour l'Open innovation

- **Recenser l'ensemble des données d'intérêt agricole disponibles**
 - Cataloguer et décrire les données disponibles fournies par apport volontaire de leur producteur
 - Mettre à disposition les données par téléchargement sous conditions (licences, données ouvertes...)
- **Présenter des services, OAD et applications développés par les utilisateurs de la plateforme + offres connexes**
 - Cloud de stockage des données agriculteurs;
 - Applications, tutoriels en ligne, test des produits, référentiels de production...
 - Outils (capteurs) et services nouveaux, travaux de recherche et innovations...
- **Servir d'espace d'échange (type forum ou autre)**
 - Pour rechercher des données spécifiques ou de savoir-faire,
 - Pour mettre en lumière les innovations des agriculteurs (bottom up) ;
 - Pour échanger des expériences, des OADS, des outils ...

Les questions

Stratégiques

- Créer une gouvernance dominée par la profession agricole
- Convaincre les agriculteurs d'ouvrir leurs données, en direct ou via leurs fournisseurs de service de gestion
- Créer les conditions d'une valeur ajoutée française

Techniques

- Concilier fluidité d'accès (avec demande d'accès) et sécurité
- Assurer l'anonymisation via l'agrégation (sas d'agrégation)
- Assurer l'interopérabilité des données, et la création de référentiels

Organisationnelles

- Animer le portail et permettre son renouvellement
- Lien avec les autres portails (ETALAB, API-Agro...) et plus...

Conclusion: et les organismes de recherche ? L'exemple d'Irstea

Comment valoriser les données de la recherche dans des démarches d'innovation ouverte (portail de données) ?

Une politique Irstea

« Valoriser la donnée dans ses usages multiples »

- De l'équipe de recherche à la terre entière.
- Le portail de données fait partie des voies de valorisation.
- Intégrer ces différents niveaux, gérer la temporalité et l'accès pérenne aux données (réutilisation).

Un objectif (sur 4) à 5 ans de la Direction de l'IST

« Une gestion cohérente et pérenne des cycles de vie des documents et des données scientifiques et techniques d'Irstea : traitement et diffusion en vue d'une réutilisation »

Conclusion: Les challenges Irstea

- L'interopérabilité entre les différents silos de données produites
- La qualité des données : métadonnées, procédures de validation, démarche qualité (processus R2 « Acquérir et gérer les données »)
- Acculturer les chercheurs (publications dans des « data journals », métadonnées, processus « linked open data »)
- Faire reconnaître l'importance du partage de données dans l'évaluation des chercheurs
- Collaboration étroite avec la Direction des Systèmes d'Information et autres Directions => évolution du métier de l'IST?
- Implication dans les instances nationales et internationales (Research Data Alliance)

... à partager avec les autres organismes de recherche