

Données de la recherche et recherche agronomique

Esther Dzalé Yeumo - Chef de projet - Délégation à l'IST



Stage ENSSIB – 07 juin 2016



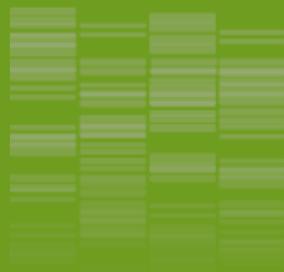


Données de la recherche et recherche agronomique : faire adopter et mettre en place une politique de gestion, valorisation, préservation et partage

- ❖ De quelles données parle-t-on ?
- ❖ Quels sont les enjeux de leur gestion, valorisation et préservation pour la recherche ?
- ❖ Comment s'est organisée la réflexion dans l'institut pour faire adopter une politique de conservation des données ?
- ❖ Quelle dimension stratégique ?
- ❖ Quelles orientations/réalisations acquises ?
- ❖ Comment s'inscrivent les métiers de l'information/documentation dans ces réalisations aujourd'hui ?
- ❖ Quelles compétences sont requises ?

Plan

- ❖ Gestion et partage des données à l'INRA
 - ❖ Contexte et enjeux pour l'Inra
 - ❖ Le chantier gestion et partage des données de l'INRA
- ❖ De quelles données parle-t-on?
 - ❖ Essai de définition des données de la recherche
 - ❖ Les données de la recherche à l'Inra
- ❖ L'offre de service pour la gestion et le partage des données
- ❖ Place de l'IST



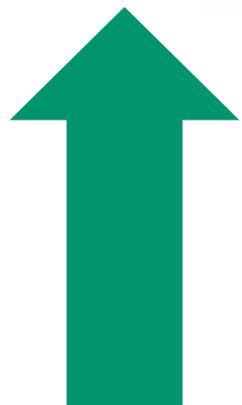
_01

Gestion et partage des données

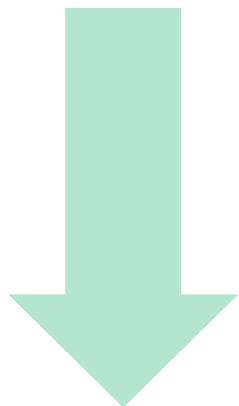
CONTEXTE ET ENJEUX POUR L'INRA



Contexte général



Se conformer à la réglementation sur l'ouverture des données



Protection, intelligence économique



Transparence / preuve
Création de valeur / Innovation



Politique – Juridique

Scientifique

Se conformer à la réglementation (open data)

❖ Loi CADA

- ✓ Les jeux de données produits par la recherche publique sont assimilables dans certains cas à des documents administratifs
- ✓ Les documents administratifs achevés doivent être communiqués

❖ Loi Valter (modifie la loi CADA)

- ✓ Incitation à mettre à disposition publiquement les documents achevés sans attendre une demande
- ✓ Les documents administratifs mis à disposition sont utilisables librement, gratuitement (sauf exceptions) et à toutes fins

Code de la recherche modifié par la loi ESR de juillet 2013

Chapitre II : Objectifs et moyens institutionnels de la recherche publique.

Article L112-1 [En savoir plus sur cet article...](#)

Modifié par [LOI n°2013-660 du 22 juillet 2013 - art. 16](#)

La recherche publique a pour objectifs :

- a) Le développement et le progrès de la recherche dans tous les domaines de la connaissance ;
- b) La valorisation des résultats de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie ;
- c) Le partage et la diffusion des connaissances scientifiques en donnant priorité aux formats libres d'accès ;

c bis) Le développement d'une capacité d'expertise et d'appui aux associations et fondations, reconnues d'utilité publique, et aux politiques publiques menées pour répondre aux défis sociétaux, aux besoins sociaux, économiques et du développement durable ;

d) La formation à la recherche et par la recherche ;

e) L'organisation de l'accès libre aux données scientifiques.

Les établissements publics de recherche et les établissements d'enseignement supérieur favorisent le développement des travaux de coopération avec les associations et fondations, reconnues d'utilité publique. Ils participent à la promotion de la recherche participative et au développement des capacités d'innovation technologique et sociale de la Nation. Ces coopérations s'exercent dans le respect de l'indépendance des chercheurs et, en l'absence de clauses contraires, dans un but non lucratif. Les travaux de recherche menés dans le cadre de ces coopérations sont, en l'absence de clauses

Nouveau décret sur les missions de l'INRA

Décret n° 2015-1517 du 23 novembre 2015 relatif à l'Institut national de la recherche agronomique

NOR : MENR1517685D

« Dans son domaine de compétences, l'institut :

« a) Produit, publie et diffuse les connaissances scientifiques résultant de ses travaux de recherche et d'expertise ;

« b) Organise, en l'absence de clauses contraires, l'accès libre aux données scientifiques et aux publications ;

« c) Contribue à l'élaboration de la politique et de la stratégie nationale et européenne de recherche ;

« d) Apporte son concours à la formation à la recherche et par la recherche ;

« e) Contribue, par la valorisation de ses compétences, de ses savoir-faire et des résultats de la recherche, à la conception d'innovations technologiques et sociales ;

« f) Contribue au développement de la capacité d'expertise scientifique et d'appui aux associations et fondations reconnues d'utilité publique et aux politiques publiques menées pour répondre aux défis sociétaux, notamment aux besoins économiques, sociaux et environnementaux liés au développement durable ;

« g) Contribue aux débats qui concernent la place de la science et de la technologie dans la société. »

Charte de déontologie

Les données, les savoir-faire et collections, la propriété industrielle : un patrimoine de qualité à préserver et valoriser

9. Les données de la recherche constituent l'une des composantes identifiables des productions de l'Institut, qu'elles soient issues des équipes de recherche, des infrastructures de recherche, des services administratifs. Elles sont la propriété de l'INRA, ou la copropriété, si elles sont obtenues dans le cadre d'accords de partenariat. La production, l'archivage, le traitement, la gestion des données de la recherche, obtenues par l'Institut ou transmises par des tiers, doivent obéir à des procédures explicitées, respectant les règles juridiques et déontologiques qui prévalent, notamment dans l'expérimentation ou dans l'utilisation des données personnelles. Ces procédures garantissent leur qualité, leur sécurisation, en particulier informatique, leur traçabilité, et fixent les règles de leur mise à disposition.

<http://prodinra.inra.fr/record/188870>

Avis du comité d'éthique

RECOMMANDATIONS À L'ADRESSE DES DIRECTIONS ET DES PERSONNELS DES DEUX ÉTABLISSEMENTS

1• Le Comité d'éthique ne peut que conforter l'Inra et le Cirad dans leur volonté de partager les données issues de leurs recherches. Le Comité d'éthique recommande de définir une politique d'établissement concernant la gestion et le partage des données qui affiche clairement les engagements des organismes et précise les rôles et responsabilités des différents acteurs. La politique d'établissement devra prévoir un important volet d'information et de formation continue des chercheurs à la mise en ligne des données et à leur réutilisation responsable.

2• Il paraît essentiel que la définition de cette politique d'établissement soit étroitement concertée avec les autres organismes français de recherche publique (CNRS, Inserm, IRD) afin d'aboutir à une réelle mutualisation d'expertise.

3• Il convient de garantir l'harmonisation de la sémantique, la standardisation, le contrôle qualité, etc. Ce travail devra être mené notamment au sein de la communauté internationale *ou du Global Open Data for Agriculture and Nutrition (GODAN)*.

4• La politique d'établissement doit aussi évaluer précisément et dégager les moyens humains et techniques nécessaires pour en assurer le succès. En particulier, s'imposent la création et la reconnaissance de nouveaux métiers spécialisés dans le contrôle qualité, l'édition, la gestion et la mise en ligne de données (conservateurs de données), ainsi que dans l'exploitation de ces données (fouille de textes et de données, FTD).

5• La qualité des données constitue un aspect essentiel de leur fiabilité. Les chercheurs devront y consacrer une attention particulière, mais les organismes devront renforcer et généraliser les procédures de contrôle de cette qualité, ce qui pourra nécessiter des actions spécifiques de formation.

6• Les activités de production et de partage des données de qualité devront faire l'objet d'une reconnaissance des institutions et être prises en considération dans l'évaluation des chercheurs, des équipes, unités et organismes.

Extrait du [rapport du comité d'éthique](#)

La recherche a pour objectif premier de produire des données à l'origine de nouvelles connaissances. Cependant, toutes les données issues de la recherche ne peuvent être partagées sans discernement.

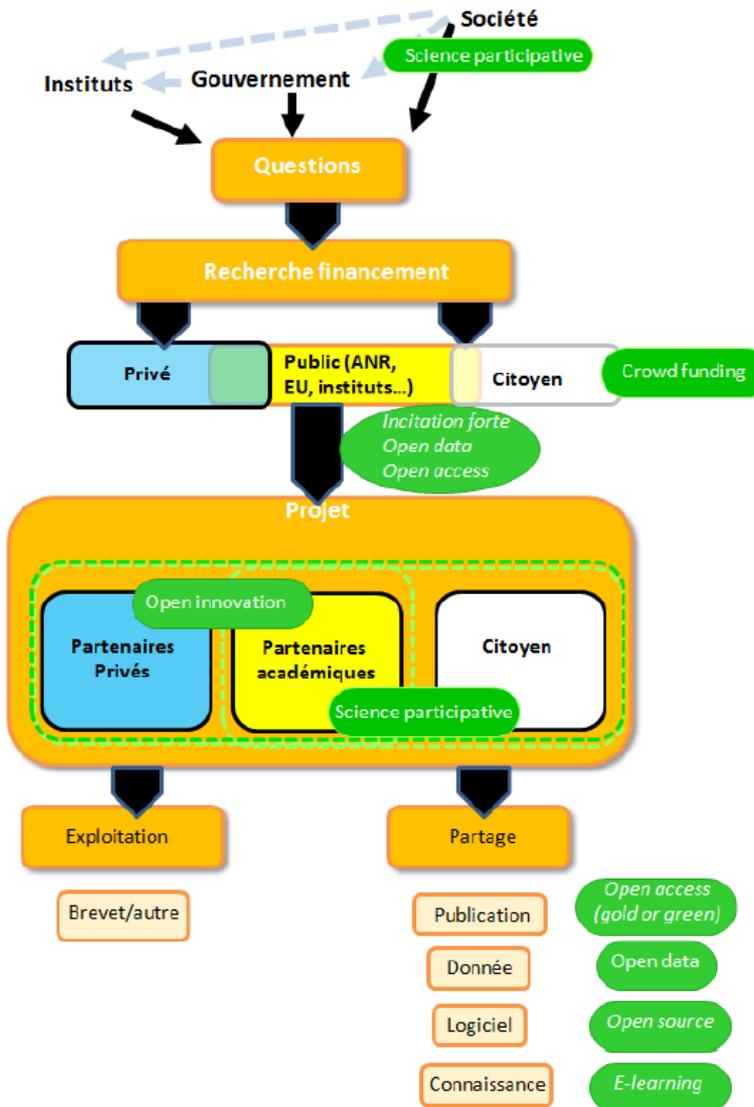
Compte tenu de l'extrême diversité des activités de l'Inra et du Cirad, de la complexité des choix qui en résulteront (de la mise en ligne sans restriction à la conservation hors ligne, en passant par le partage des données selon des conditions très spécifiques), le Comité estime qu'un processus de décision, appliqué au cas par cas, est inévitable.

Vers une dynamique nécessaire de partage et d'analyse des données scientifiques

- Some Societal challenges :
 - Feed the world
 - Climate change
 - Sustainable agriculture
 - Health and nutrition
- Imply to deal with data :
 - Data driven science
 - Big data
 - Data management, sharing, and re-use
- With different point of views :
 - Politics
 - Technics (IT)
 - Scientific Disciplines
 - Intellectual property, ethics
 - Economics



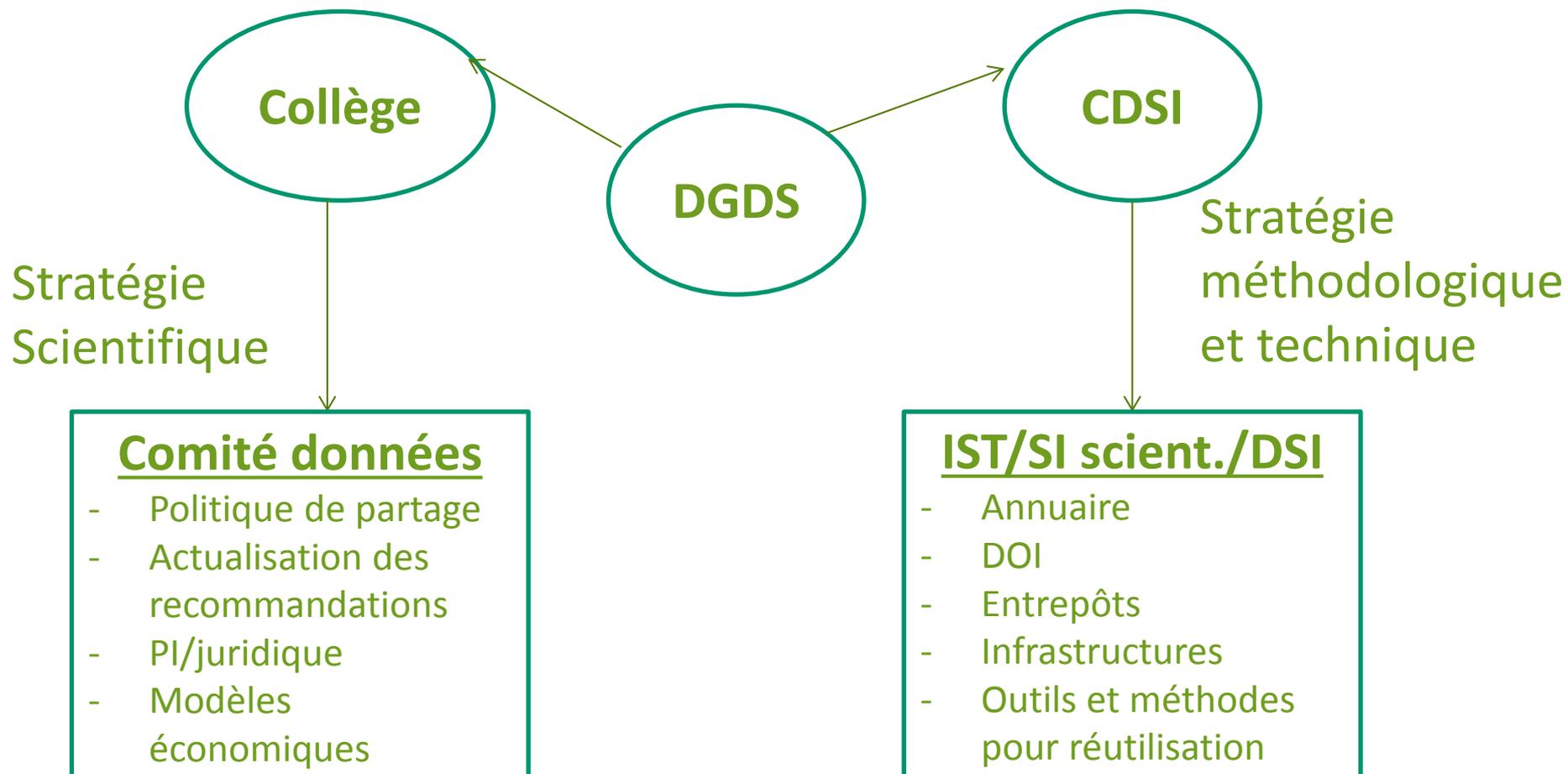
Dans le contexte d'une science ouverte



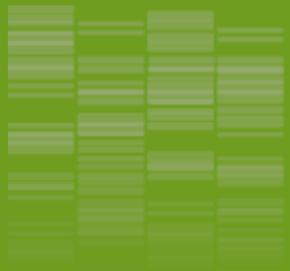
Diapo O. Hologne, INRA

Gouvernance – organisation

R2 du CS : dispositif d'évaluation des données



Diapo O. Hologne, INRA

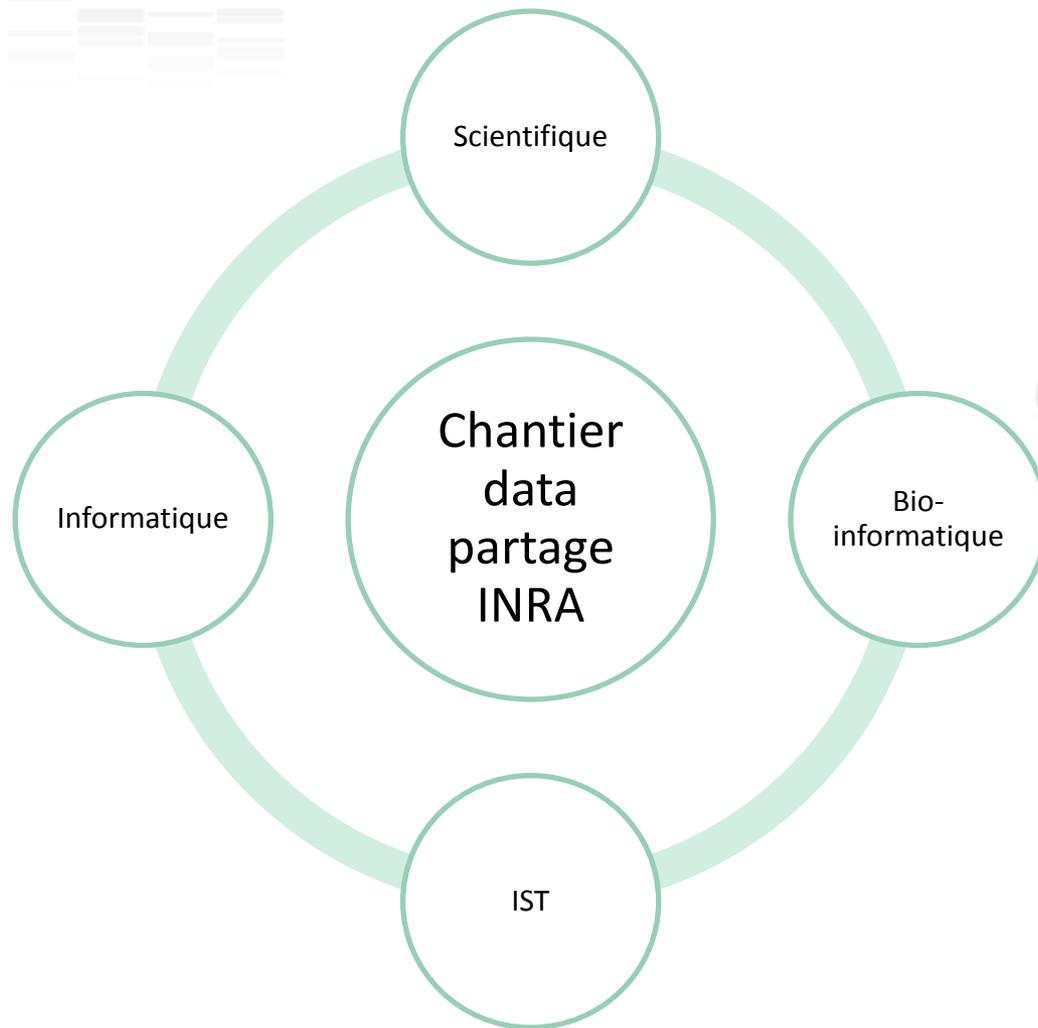


_02

Gestion et partage des données

LE CHANTIER INSTITUTIONNEL

Une démarche institutionnelle



Partage des données à l'Inra : étapes clés

2011-2012 : le CS instruit la question

- Groupe de travail piloté par D. Pontier
- 9 recommandations pour l'Inra (rapport Juin 2012)

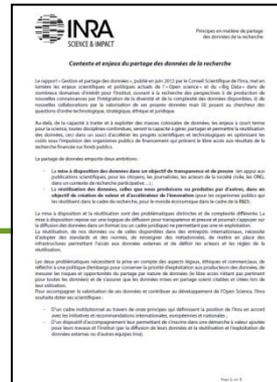


Rapport CS

<http://prodinra.inra.fr/record/206746>

Fin 2012 : élaboration d'une politique

- 11 principes pour mieux gérer et partager les données de la recherche



Note de cadrage :

[Principes en matière de partage des données de la recherche](#)

Avril 2013 : lancement des chantiers de mise en œuvre

- 3 chantiers disciplinaires / familles de données
- Chantier juridique
- Chantiers techniques (outils et méthodes)
- Chantier compétences

Chantier

« Data Partage »

Diapo O. Hologne, INRA

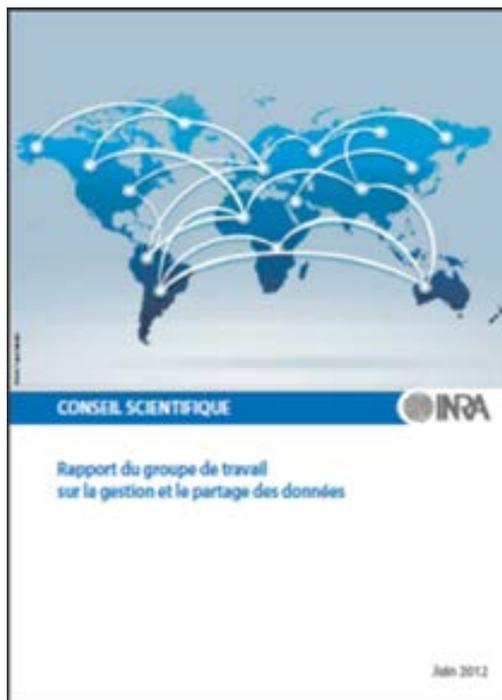
Reco du rapport du CS : fil rouge des actions

RESUME

Depuis quelques années, la biologie et les sciences humaines et sociales font face à un accroissement exponentiel des données, provenant de l'adoption en masse des nouvelles technologies, et du développement des sciences et techniques de l'information d'une ampleur et à une échelle sans précédents. Une telle rupture nécessite des transformations stratégiques majeures pour assurer le stockage, la préservation, l'exploitation de ces masses de données, mais aussi leur partage. Elle nécessite également une prise de conscience et une modification des pratiques des ingénieurs et chercheurs de l'institut, pour lesquels ces évolutions constituent un défi culturel.

Au terme de son analyse, le groupe de travail propose les recommandations suivantes :

- 1) Définir la politique de l'établissement et la communiquer.
- 2) Mettre en place un comité d'évaluation des données produites par l'Inra.
- 3) S'impliquer dans les comités internationaux de standardisation.
- 4) Développer un portail d'accès à un ensemble de ressources distribuées.
- 5) Prendre en compte le cycle de vie des données dès l'élaboration des projets de recherche.
- 6) Définir un cahier des charges pour les plateformes.
- 7) Doter l'Inra d'infrastructures dimensionnées pour les stockages et les calculs hautes performances.
- 8) S'engager dans une politique de gestion des compétences répondant aux besoins en émergence.
- 9) Conduire une réflexion inter-organismes pour promouvoir une politique nationale et locale en matière de gestion et partage de données.



Diapo O. Hologne, INRA

Partage des données: pratiques actuelles

❖ Données –omiques (ressources génomiques et génétiques)

- Culture de partage importante
 - Partage réalisé en général après publication scientifique
 - Entrepôts internationaux

❖ Données d'enquêtes, analyses textuelles

- Culture de partage très faible + problèmes anonymisation et de droits sur données
 - Pas de difficultés pour le partage des données d'économie expérimentale
 - Partage de données ayant servi à des publications dans le respect de l'anonymat, du secret statistique de la propriété intellectuelle et du cadre juridique (données achetées)

❖ Données d'expérimentation, observations

- Culture de partage faible mais globalement ok
 - Dépend de l'origine, de la nature et du public visé
 - Partage « *réfléchi* » des protocoles expérimentaux ou analytiques qui peuvent constituer un avantage compétitif

Attentes des scientifiques de l'INRA

❖ Services, accompagnement

- Accompagnement juridique et SSI : quelles données partager, quelles conditions, quelles précautions?
- Accompagnement IST : politiques éditoriales, publication des données (data papers, choix d'un entrepôt, doi), lien données et publications



❖ Outils informatiques

- Portail de données
- Accès sécurisé aux données soutenant une publication par les relecteurs
- DOI
- Environnement collaboratifs, plateformes de logiciels



❖ Politique, RH

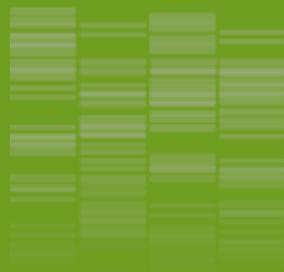
- Reconnaissance de tous les contributeurs à la gestion et au partage des données
- Gestion prévisionnelle des emplois et des compétences



Attentes des scientifiques de l'INRA

Extrait de l'avis du comité d'éthique :

« Le Comité constate que nombre de chercheurs expriment leurs incertitudes sur le statut des données qu'ils produisent, en particulier au regard de cette question du partage et des missions de valorisation de l'Inra et du Cirad. Ils formulent clairement un besoin d'informations et de concertation et attendent de leur institution qu'elle les guide dans ce nouveau domaine, vœu qui paraît d'autant plus légitime que la responsabilité de toutes les données produites appartient, de fait, aux organismes de recherche. »



_03

Données de la recherche

DÉFINITION(S)



Plusieurs définitions des données de la recherche

« Providing an authoritative definition of research data is challenging, as any definition is likely to depend on the context in which the question is asked »

<http://ands.org.au/guides/what-is-research-data.html>

- ❖ La notion de « données de la recherche » est différente selon
 - les disciplines scientifiques
 - Le contexte juridique
 - Les choix politiques

Définition de l'Australian National Data Service (ANDS)

*“Some of the data might be **raw data**, the unprocessed observations of particular phenomena. Some might be **processed data**, the data produced when raw data has been calibrated or corrected. Some might be **derived data**, which present a summary or specific view of the raw data. Some might be **textual data**, the publications which result from a research project or the textual data (texts, bibliographies, surveys, etc.) which forms the basis of a research project”.*

(Research Data Strategy Working Group, 2011)

Guide en ligne

<http://ands.org.au/guides/what-is-research-data.html>

Définition

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.**

(OCDE, 2007)

Principes et Lignes directrices pour l'accès aux données de la recherche financée sur fonds publics

I. Objectifs

II. Champ d'application et définitions

Données de la recherche

Données de la recherche financée sur fonds publics

Dispositifs d'accès

III. Principes

A. Ouverture

B. Flexibilité

C. Transparence

D. Conformité au droit

E. Protection de la propriété intellectuelle

F. Responsabilité formelle

G. Professionnalisme

H. Interopérabilité

I. Qualité

J. Sécurité

K. Efficience

L. Responsabilité de rendre compte

M. Pérennité

Extrait – Rapport pour le conseil scientifique de l'Inra

« En l'absence d'une définition juridique, nous appellerons une donnée scientifique ou donnée de la recherche : information qui représente le matériau de base d'une activité de recherche ayant bénéficié d'un financement sur fonds publics. » (Gaspin & Pontier, 2012)

Extrait - Rapport du groupe juridique du chantier gestion et partage des données de l'INRA

«... l'expression données de la recherche ne s'applique pas aux éléments suivants: carnets de laboratoires, analyses préliminaires et projets de documents scientifiques, programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues, et objets matériels (par exemple les échantillons de laboratoires, les souches bactériennes et les animaux de laboratoires tels que les souris »

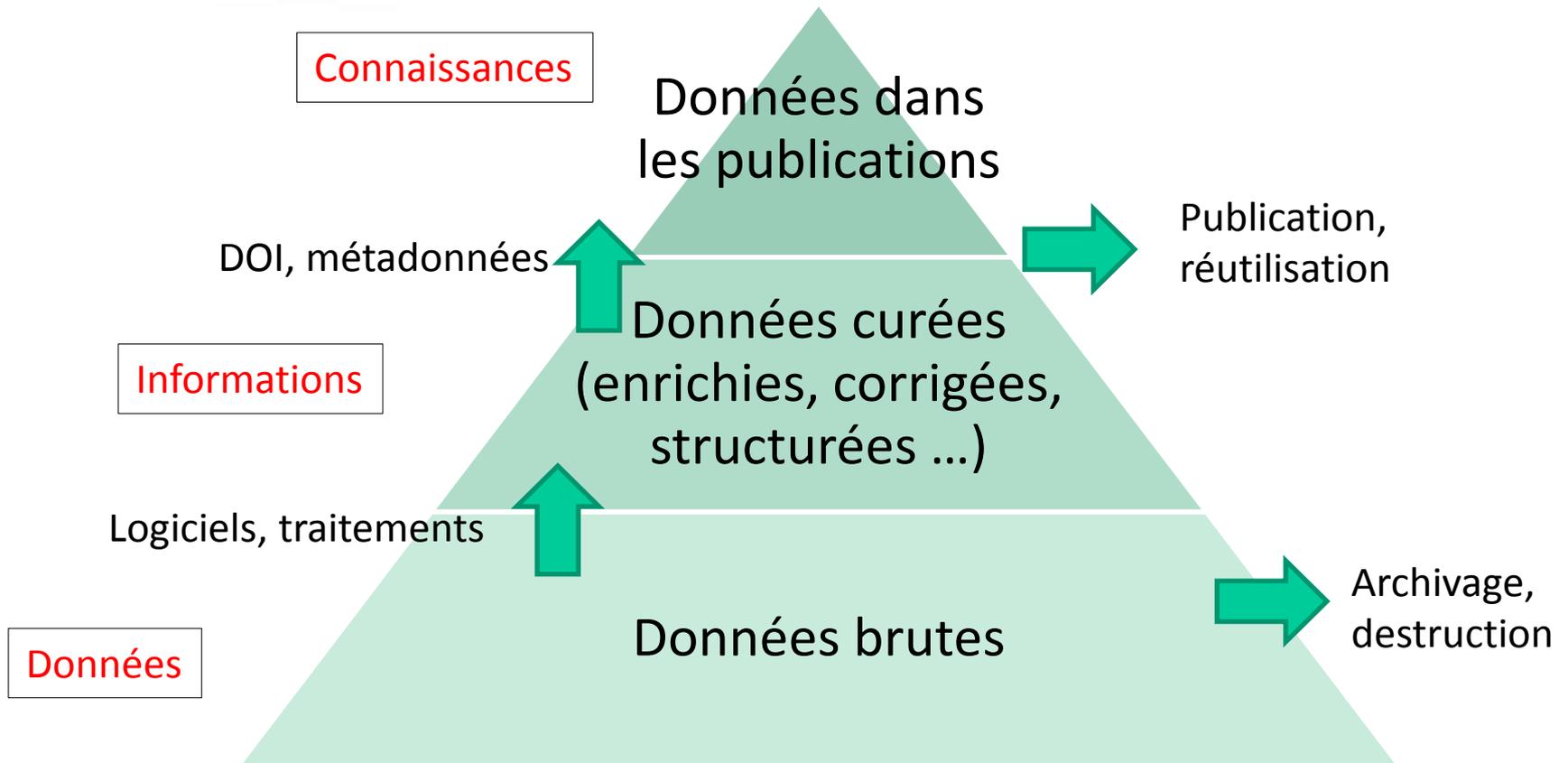
Données, informations, connaissances

Le qualitatif : données, informations et connaissances

Données	Description élémentaire d'une réalité	<i>Mesures de températures dans une station météo</i>
Informations	Données avec un sens (pour construire une représentation de la réalité)	<i>Une courbe donnant l'évolution des minimas & maximas moyens en un lieu suivant le mois de l'année</i>
Connaissances	Informations avec une vérité, plus généralement une loi qui est considérée comme vraie	<i>Le fait que la température sur terre augmente du fait de l'activité humaine</i>

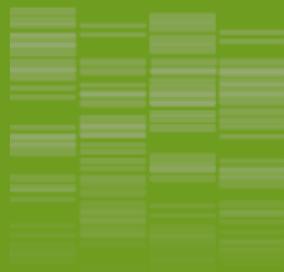
(Abiteboul, 2012)

Les états de la donnée – vision processus



D'un point de vue juridique

- ❖ Pas de distinction entre données brutes, élaborées ou métadonnées
- ❖ Les données sont des « documents administratifs »
- ❖ Elles sont considérées comme une information « de libre parcours ». A ce titre, l'établissement du producteur de la donnée peut décider de sa diffusion ou non.
- ❖ Deux exceptions où une « propriété » peut s'exercer :
 - les publications (droit moral + patrimonial), le chercheur « propriétaire »
 - Les bases de données « droit sui generis » du producteur de BdD protège l'investisseur



_04

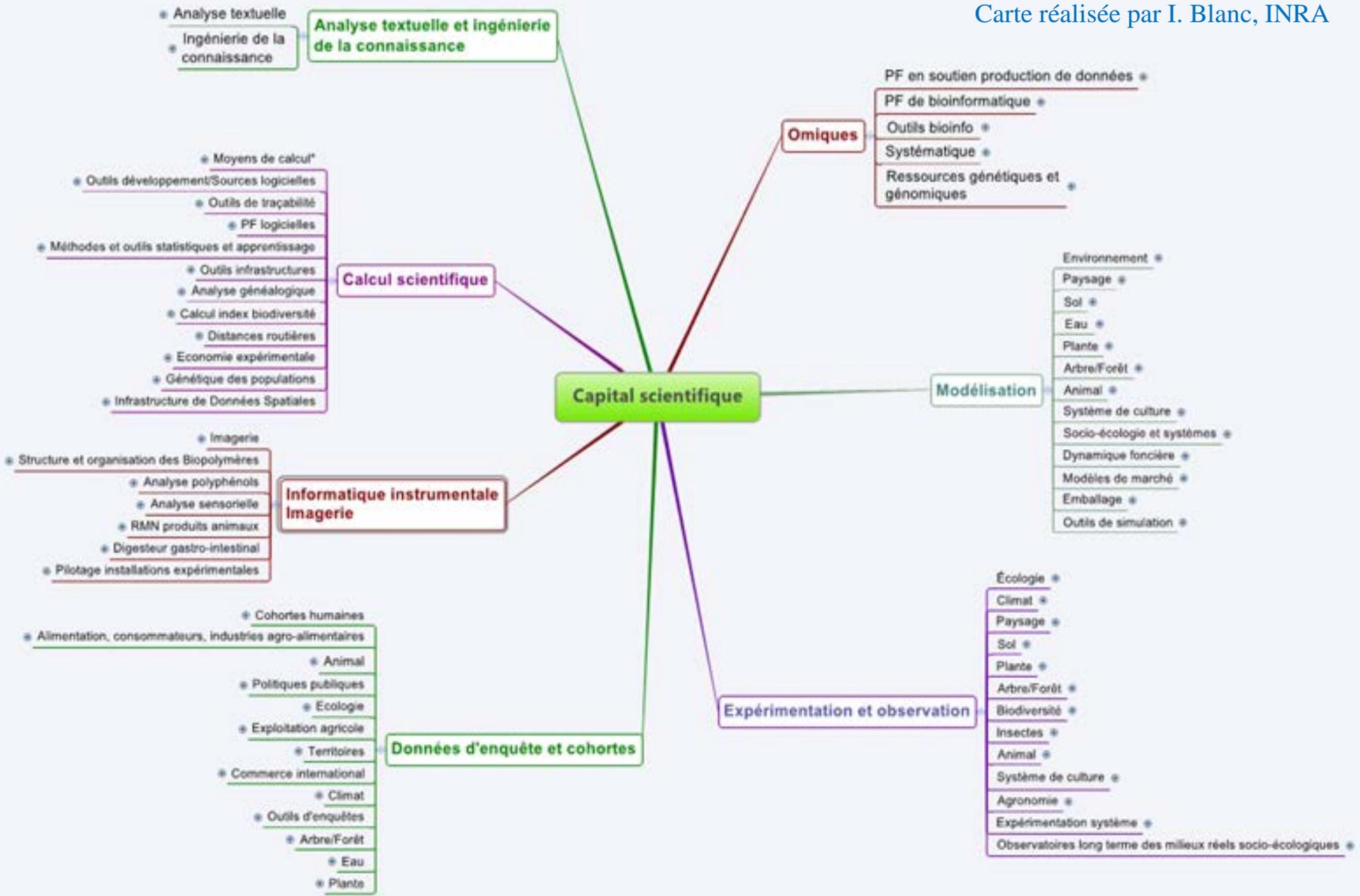
Les données de la recherche à l'INRA

Quelles données à l'INRA?

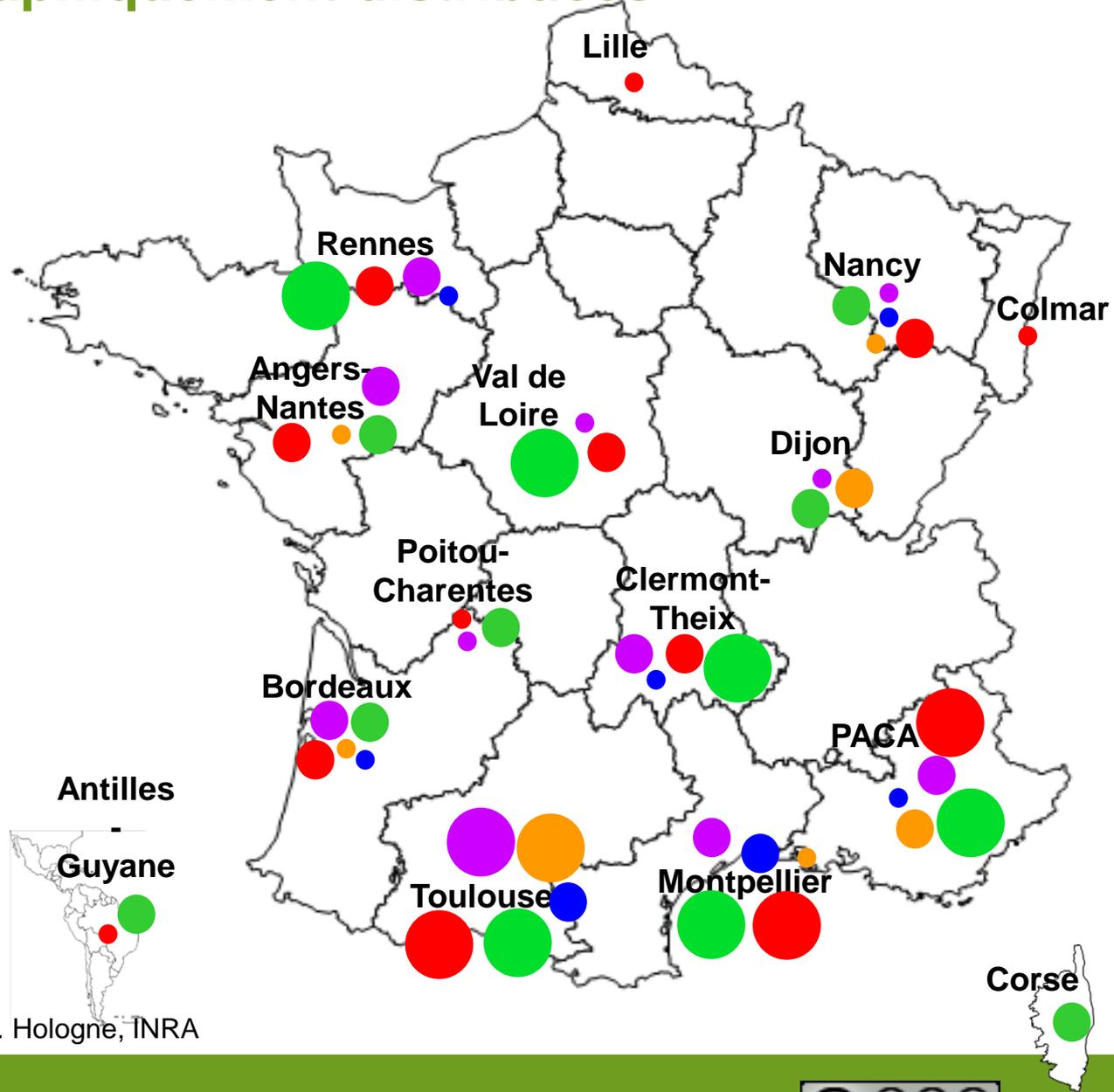
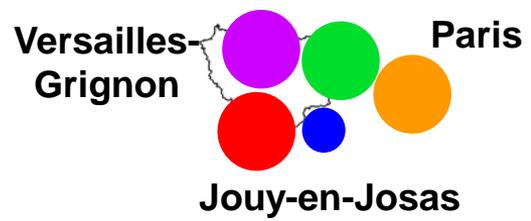
Genomic
Observation
Experimental
Simulation
Social **Genetic**

Text
Numeric
Video
Audio
Images

Purchased: ∞
Small
Secondary
Historical
Active
Processed
Raw
Primary



Des données géographiquement distribuées



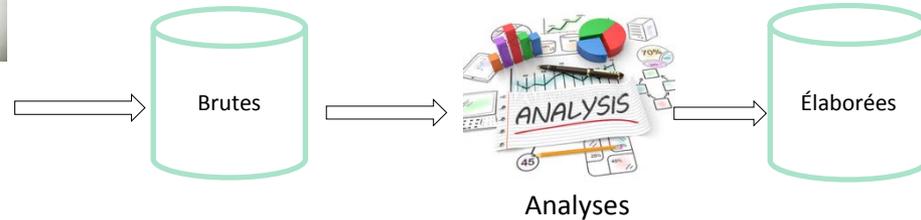
- -omics, ressources génétiques
- Expérimentation – observation
- Modélisation
- Enquête et cohortes
- Données extraites de l'externe

Diapo O. Hologne, INRA

La famille -omique



Collecte / Acquisition



Principales caractéristiques

- 19 types de données: 6 types de données brutes et 13 types de données élaborées
- Volumétrie croissante
- Données brutes plus volumineuses que données élaborées
- Existence de standards de formats et de métadonnées (pour 16 types sur 19)
- Souvent produites en collaborations avec des partenaires publics ou privés
- Durée moyenne de valorisation = 3 ans (plus si espèces modèles)

brute	séquences lues ADN-ARN
brute	génotypes SNP
brute	génotypes SSR
brute	données d'expression (arrays, qPCR)
brute	données métabolome
brute	profils protéiques (quantitatifs)
élaborée	séquences protéines
élaborée	séquences alignées/assemblées
élaborée	données d'expression RNAseq
élaborée	polymorphismes SNP
élaborée	polymorphismes SSR
élaborée	variants structuraux
élaborée	patrons de méthylation
élaborée	annotations des gènes
élaborée	orthologues, paralogues, familles de gènes
élaborée	cartes (génétiques, QTLs, physiques)
élaborée	données passeport populations/souches
élaborée	données passeport croisements temporaires
élaborée	données passeport banques génomiques

La famille -omique

Données relatives aux ressources génétiques et génomiques

Problématiques

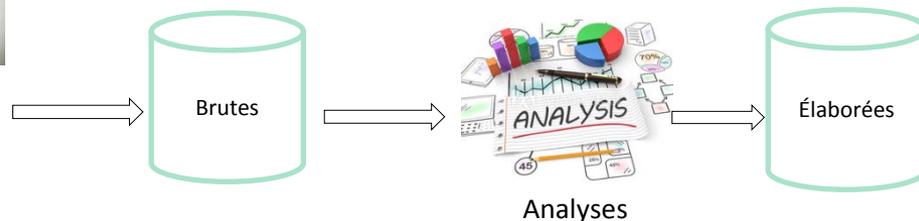
- Stockage distribué
- Coût du stockage, parfois plus que celui de la production (séquençage)
- Pérennité des entrepôts
- Traitements de gros volumes
- Méta-analyses

brute	séquences lues ADN-ARN
brute	génotypes SNP
brute	génotypes SSR
brute	données d'expression (arrays, qPCR)
brute	données métabolome
brute	profils protéiques (quantitatifs)
élaborée	séquences protéines
élaborée	séquences alignées/assemblées
élaborée	données d'expression RNAseq
élaborée	polymorphismes SNP
élaborée	polymorphismes SSR
elaborée	variants structuraux
elaborée	patrons de méthylation
elaborée	annotations des gènes
elaborée	orthologues, paralogues, familles de gènes
elaborée	cartes (génétiques, QTLs, physiques)
elaborée	données passeport populations/souches
elaborée	données passeport croisements temporaires
elaborée	données passeport banques génomiques

La famille – enquêtes, cohortes



Collecte / Acquisition



Principales caractéristiques

- Susceptibles de contenir des informations sensibles
- Peuvent concerner de petits échantillons
- Importantes données secondaires (données obtenues après enrichissement de données primaires)
- Importantes données achetées

brutes , originales	Enquêtes auprès de petits échantillons
Brutes, originales	Enquêtes sociologiques auprès d'échantillons plus importants
Brutes, originales	Données expérimentales, en générale économiques
Brutes, originales	Données de relevés d'étiquetage ou d'analyse
Élaborées, secondaire	Données d'achats harmonisées ou agrégées (ex: Kantar)
Élaborées, secondaires	Séries historiques constituées à partir d'autres données informatisées ou non
Élaborées	Données textuelles
Élaborées	Données soutenant une publication
Élaborées	Résultats (tableaux statistiques, élasticités, ...)

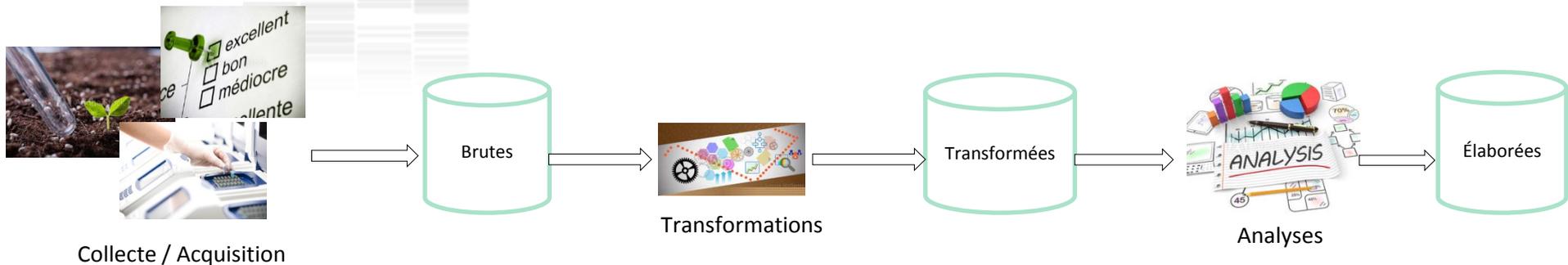
La famille – enquêtes, cohortes

Problématiques

- Anonymisation, respect du secret statistique
- Aspects longitudinal, panel historique
- Partage des données secondaires
- Respects des recommandations et des obligations réglementaires en matière de destruction des données
- Pérennité des entrepôts / plateformes
- Accès par les relecteurs aux données soutenant des publications dans le respect de la PI et du cadre légal

brutes , originales	Enquêtes auprès de petits échantillons
Brutes, originales	Enquêtes sociologiques auprès d'échantillons plus importants
Brutes, originales	Données expérimentales, en générale économiques
Brutes, originales	Données de relevés d'étiquetage ou d'analyse
Élaborées, secondaire	Données d'achats harmonisées ou agrégées (ex: Kantar)
Élaborées, secondaires	Séries historiques constituées à partir d'autres données informatisées ou non
Élaborées	Données textuelles
Élaborées	Données soutenant une publication
Élaborées	Résultats (tableaux statistiques, élasticités, ...)

La famille – expérimentations, observations



Principales caractéristiques

- Données d'expérimentation : données acquises dans le cadre de dispositifs dans lesquels certains facteurs sont contrôlés ou manipulés
- Données d'observation : données acquises en milieu naturel sur des objets non manipulés (phénotypage, essais agronomiques, etc.)
- Données de simulation : données produites par des expériences de calcul numérique en utilisant des algorithmes et des équations

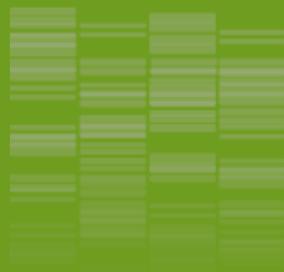
Brutes non transformées	Données acquises par les systèmes de mesures.
Brutes transformées	Données brutes converties en grandeurs physiques et unités
Données élaborées	Données obtenues par combinaison / analyse de plusieurs types de données brutes ou élaborées. Exemples : carte de température interpolée sur un domaine spatial, flux de chaleur dans un organisme, résistance à un stress, etc.

La famille – expérimentations, observations

Problématiques

- Stockage distribué
- Pérennité des entrepôts
- Traitement / analyse de gros volumes (en particulier images)
- Standardisation des formats, des métadonnées et des vocabulaires
- Enjeux stratégiques liés à la publication des métadonnées

Brutes non transformées	Données acquises par les systèmes de mesures.
Brutes transformées	Données brutes converties en grandeurs physiques et unités
Données élaborées	Données obtenues par combinaison / analyse de plusieurs types de données brutes ou élaborées. Exemples : carte de température interpolée sur un domaine spatial, flux de chaleur dans un organisme, résistance à un stress, etc.



_05

Exemple concret

Exemple 1 : données d'observation de chenilles processionnaires

Protocole de l'expérimentation

- 1 Pose des pièges à phéromones sur les sites test
- 2 Comptage des papillons piégés sur les sites test
- 3 Comptage des nids de chenilles sur les sites test et témoin

→ Période de piégeage des papillons

→ Période de dénombrement des nids de chenilles



Description des sites

Données géographiques

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Communes	Saint-Fargeau-Ponthierry	Obernai	Angers							Saint-Martin-de-Ré	Marenne
2	Description Sites Témoins	Type de configuration	alignement	pas de site témoin déterminé	îlot boisé							îlot boisé	îlot boisé
3		Description des sites	parc de loisir		Echangeur de la Baumette							Parcelle forestière	à proximité la piscine Marenne
4		Géolocalisation	48°31'56.85"N, 2°33'47.42"E		47°27'53.61"N, 0°34'14.82"O							46°15'3.24"N, 1°31'0.58"O	45°49'28.21"N, 1°8'19.54"O
5		Surface	56 m		25100 m²							0,13 ha	10 ha
6		Historique du site	non communiqué		non communiqué							Comptage des nids par l'ONF depuis 2003 (données transmises)	non communiqué
7	Nombre de pins et/ou de cèdres	20	62							116	non communiqué		
8	Nombre de sites tests	1	1	7							1	1	
9	Type de configuration	alignement	îlot boisé	alignement	îlot boisé	alignement	alignement	alignement	alignement	îlot boisé	alignement	îlot boisé	îlot boisé
10	Description des sites	parc de loisir	jardin privé	Roseraie I bord de route	Roseraie II Parc	Roseraie III butte antibruit	Roseraie IV quartier résidentiel	Arboretum V Falun d'Orgemont	Arboretum VI cours d'école	Arboretum VII Plaine de jeux	parc	bosquet	
11	Géolocalisation	48°31'59.30"N, 2°33'29.47"E	48°27'34.36"N, 7°30'18.04"E	47°26'55.62"N, 0°34'21.40"O	47°26'48.09"N, 0°33'51.29"O	47°26'33.64"N, 0°33'28.63"O	47°26'38.59"N, 0°33'14.78"O	47°26'43.17"N, 0°32'29.08"O	47°27'10.16"N, 0°32'47.32"O	47°27'45.99"N, 0°31'39.55"O	46°10'54.61"N, 1°22'26.48"O	45°49'14.1"N, 1°8'10.14"O	
12	Surface	500 m	< 15 ares	3000 m²	17939 m²	8650 m²	8650 m²	3250 m²	5270 m²	1077 m²	4,6 ha	1 ha	

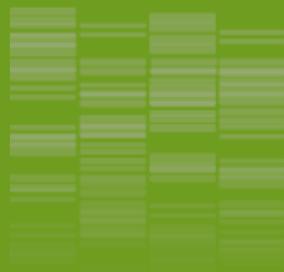
Protocole

Relevés

Dispositifs pièges à phéromones été 2011	Relevés des nids réalisés à l'hiver 2011	non réalisés	non réalisés	non réalisés							non réalisés
	Nombre de pièges installés	10	15	14 a) 1 b)	9	12 a) 1 b)	12 a) 2 b)	8 a) 1 b)	6 a) 1 b)	8	18
	Type de pièges	Procerex	Mastrap	a) Mastrap b) Nufarm	Mastrap	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	Mastrap	Mastrap
	Type de phéromones	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex
	Date de la pose des pièges	25 juillet (sem 30)	01 juin (sem 22)	01 juillet (sem 26)			01 juillet (sem 26) et 06 juillet (sem 27)	31 mai (sem 22) et 19 juillet (sem 29)	19 juillet (sem 29)	19 juillet (sem 29)	12 juil (sem 28)
	Premiers vols de papillons constatés	non communiqué	Sem 26 et 27	sem 25							non réalisés
	Relevés intermédiaires réalisés	non réalisés	19 juin 4 juillet	non réalisés							30 juil (sem 29)
	Nombre de papillons piégés au total	non communiqué	8	340	305	430	201	390	285	155	580

Bibliographie – Qu'est ce que les données ?

- ❖ Arend, D., Lange, M., Chen, J., Colmsee, C., Flemming, S., Hecht, D., & Scholz, U. (2014). eIDAL--a framework to store, share and publish research data. *Bmc Bioinformatics*, 15, 214. [10.1186/1471-2105-15-214](https://doi.org/10.1186/1471-2105-15-214)
- ❖ Borgman, C. L. (2014). [Keynote: Data, Data, Everywhere, Nor Any Drop to Drink (slides)]. <http://works.bepress.com/borgman/322>
- ❖ Fayet, S. (2013). « Données » de la recherche, les mal nommées Retrieved from <http://urfistinfo.hypotheses.org/2581>
- ❖ Gaspin, C., & Pontier, D. (2012). Rapport du groupe de travail sur la gestion et le partage des données. Conseil Scientifique de l'Inra (Ed.), (pp. 1-62). http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf
- ❖ Knight, G. (2013). [Data Management for Librarians : an introduction]. <http://fr.slideshare.net/GarethKnight/data-management-for-librarians-an-introduction>
- ❖ OCDE. (2007). Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics (pp. 1-29). <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>
- ❖ Simukovic, E., Kindling, M., & Schirmbacher, P. (2014). *Unveiling Research Data Stocks: A Case of Humboldt-Universität zu Berlin*. iConference 2014, 2014/03/04-07, Berlin - DEU. <http://hdl.handle.net/2142/47259>



_06

Gestion et partages des données à l'INRA

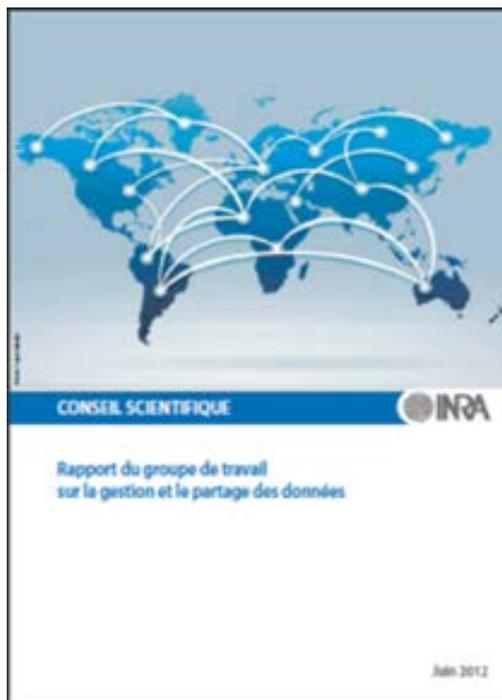
OFFRE DE SERVICE

Reco du rapport du CS : fil rouge des actions

À faire 

Commencé 

Fait 



RESUME

Depuis quelques années, la biologie et les sciences humaines et sociales font face à un accroissement exponentiel des données, provenant de l'adoption en masse des nouvelles technologies, et du développement des sciences et techniques de l'information d'une ampleur et à une échelle sans précédents. Une telle rupture nécessite des transformations stratégiques majeures pour assurer le stockage, la préservation, l'exploitation de ces masses de données, mais aussi leur partage. Elle nécessite également une prise de conscience et une modification des pratiques des ingénieurs et chercheurs de l'institut, pour lesquels ces évolutions constituent un défi culturel.

Au terme de son analyse, le groupe de travail propose les recommandations suivantes :

- 1) Définir la politique de l'établissement et la communiquer. 
- 2) Mettre en place un comité d'évaluation des données produites par l'Inra. 
- 3) S'impliquer dans les comités internationaux de standardisation. 
- 4) Développer un portail d'accès à un ensemble de ressources distribuées. 
- 5) Prendre en compte le cycle de vie des données dès l'élaboration des projets de recherche. 
- 6) Définir un cahier des charges pour les plateformes. 
- 7) Doter l'Inra d'infrastructures dimensionnées pour les stockages et les calculs hautes performances. 
- 8) S'engager dans une politique de gestion des compétences répondant aux besoins en émergence. 
- 9) Conduire une réflexion inter-organismes pour promouvoir une politique nationale et locale en matière de gestion et partage de données. 

Attribution de DOI

DOI

Le DOI est un identifiant unique et pérenne qui permet de référencer et citer des données dans vos publications scientifiques.

[Obtenir un DOI »](#)

[Modifier un DOI »](#)

[Inactiver un DOI »](#)

Obtenir un DOI

Avant de demander un DOI, il convient de vous assurer que :

- vous pouvez rendre la ressource concernée publique.
- vous avez choisi une licence en accord avec tous les co-auteurs de la ressource.
- la ressource concernée est complète

S'il s'agit d'une nouvelle version d'une ressource déjà identifiée par un DOI, nous vous recommandons d'indiquer ce lien dans le pavé "Related Identifier(s)", en fournissant le DOI de la précédente version, le type d'identifiant (DOI en l'occurrence) et le type de la relation (isNewVersionOf dans ce cas).

Il est possible d'automatiser les attributions des DOI via l'appel d'un API.

Support

Pour obtenir de l'aide, contactez-nous : doi at versailles.inra.fr

[Vos questions, nos réponses»](#)

Ou [posez votre question.](#)

© INRA 2015

Digital Object Identifier (DOI) : la synthèse



Helping you to find, access, and reuse data

Inist-CNRS
Member Institute

10.5072
Préfixe

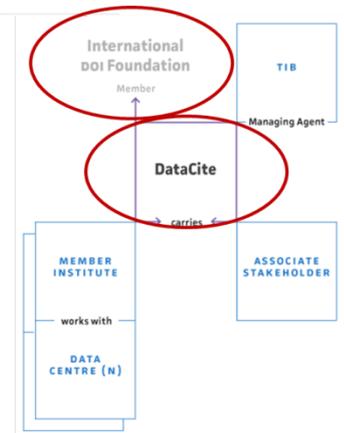
xfwugk
Suffixe

DOI: 10.5072/xfwugk

RESOLVER

Page descriptive publique

INRA
Data center



Site Web de communication et d'information



Créer

Plan de gestion

- > Pourquoi rédiger un plan de gestion ?
- > Comment rédiger un plan de gestion ?
- > Exemples de plans

Collecte de données

Formats de fichiers

Décrire ses données

Propriété intellectuelle

Gérer

Partager



🏠 Accueil • Créer • Plan de gestion

Plan de gestion des données

Un plan de gestion des données est un document formalisé, **établi au démarrage d'un projet de recherche**, et qui décrit la façon dont les données produites seront obtenues, documentées, utilisées, gérées et disséminées, pendant la phase de recherche et une fois le projet terminé.

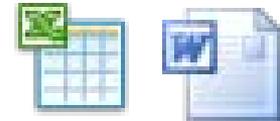
Le plan de gestion des données **prend en compte toutes les étapes du cycle de vie des données et doit être mis à jour tout au long du projet** de recherche.

Créer

- > Plan de gestion
 - > Pourquoi rédiger un plan de gestion ?
 - > Comment rédiger un plan de gestion ?
 - > Exemples de plans
- > Collecte de données
- > Formats de fichiers

Plans de gestion des données

❖ Trame type testée sur les « groupes données »



- Présentation succincte du Projet
- Données du projet : type, origine, méthodes d'acquisition/de traitement, format, métadonnées associées, publications...
- Accès et partage des données au cours et à l'issue du projet
- Stockage et sauvegarde des données au cours du projet
- Archivage et conservation des données à long terme (après la fin du projet)
- Droits de propriété intellectuelle
- Ethique et confidentialité

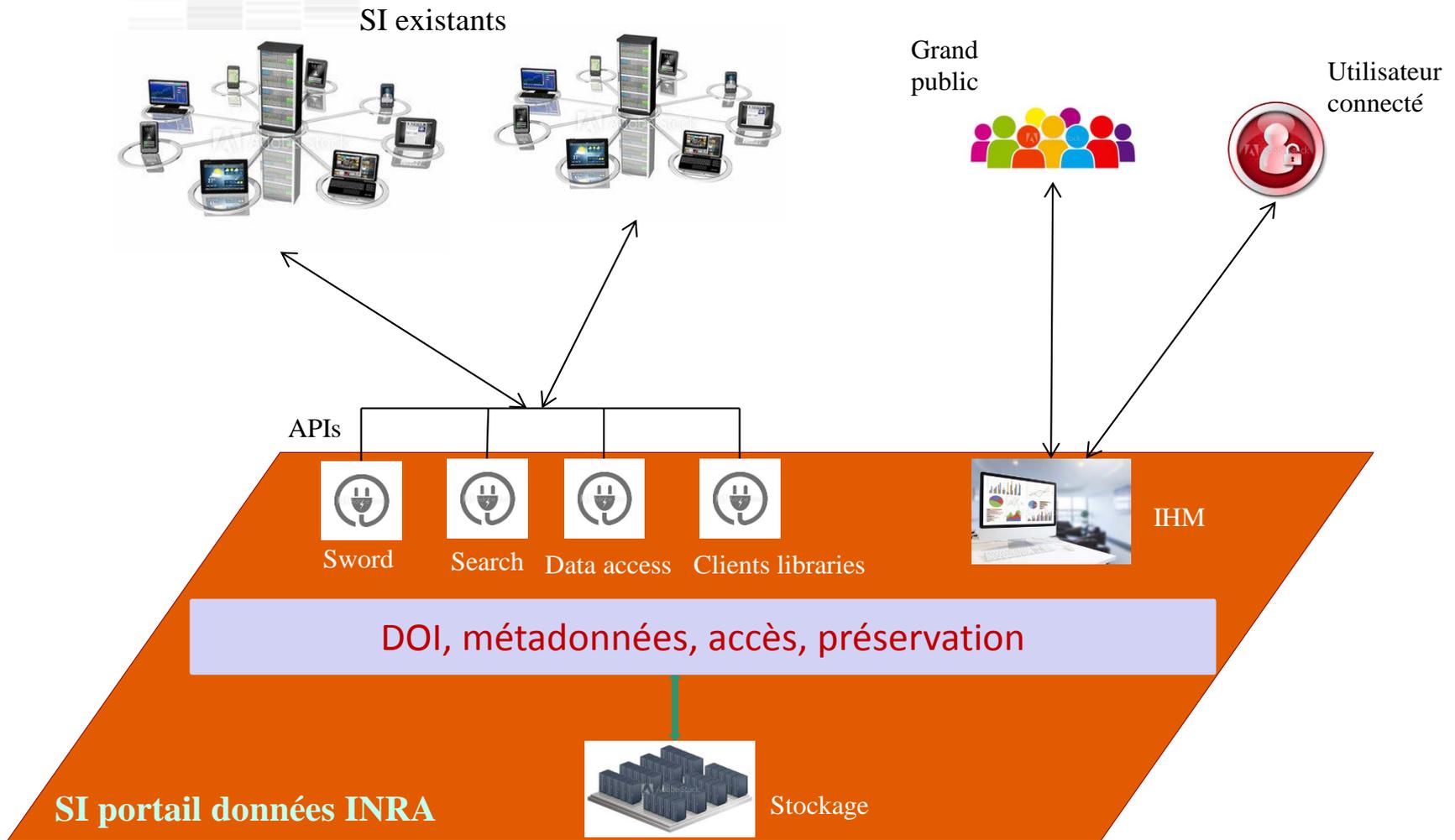
❖ Aide à la saisie : [DMPOnline](#)

- Outil Open source créé par le DCC, permet de créer un PGD en fonction des exigences d'un financeur et/ou d'une institution.
- Intégration en cours de la trame type et des aides à la saisie propres à l'Inra
- Collaboration en cours avec l'INIST qui va héberger une instance de DMPOnline pour l'ESR



Diapo S. Cocard, INRA

Portail des données INRA





Service de Q/R

[Posez votre question](#) [Toutes les réponses](#) [Focus sur...](#) [Comment ça marche ?](#) [Votre avis](#)

Archives de catégorie : [Données de la recherche](#)

Catégorie créée en juin 2015

Attribution d'un DOI

Publié le [22 avril 2016](#) par [le service Q/R](#)

Bonjour, Est-il possible d'attribuer un DOI à un document qui a été rédigé par 3 auteurs extérieurs à l'Inra, à la demande de chercheurs Inra, à partir de contenus scientifiques et techniques qu'ils ont fournis ? Précisions : Les ingénieurs ... [Continuer la lecture →](#)

Publié dans [Données de la recherche](#), [Ressources Inra](#) |

Déposer ses données dans Zenodo : quelques questions

Publié le [7 avril 2016](#) par [le service Q/R](#)

J'envisage de déposer des données, liées à une publication soumise encore en cours de correction, sur Zenodo. Je me pose des questions : Une fois déposés, peut-on modifier à tout moment les fichiers? (notamment si un reviewer nous demande de ... [Continuer la lecture →](#)

Publié dans [Données de la recherche](#) |

[Rechercher](#)

Q/R toutes fraîches !

- [Attribution d'un DOI](#)
- [Déposer ses données dans Zenodo : quelques questions](#)
- [Où déposer des données liées à une publication soumise ?](#)

Q/R par domaine

- [Bibliométrie](#) (13)
- [Données de la recherche](#) (8)
- [Droits d'auteur](#) (20)
- [Gestion bibliographique](#) (31)
- [Gestion des connaissances](#) (1)
- [Open access](#) (20)
- [ProdInra](#) (6)
- [Publication scientifique](#) (40)
- [Recherche documentaire](#) (25)
- [Réseaux sociaux](#) (2)
- [Ressources Inra](#) (31)
- [Technologies en IST](#) (12)

[S'abonner](#)

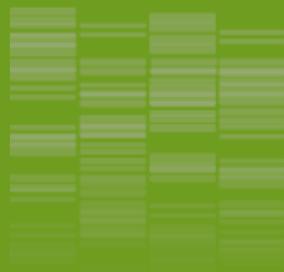
Formations

- ❖ Introduction à la gestion et au partage des données de la recherche
- ❖ PGD
- ❖ DOI
- ❖ Data papers
- ❖ Choix d'un entrepôt
- ❖ Questions juridiques et éthiques
- ❖ Etc.



infodoc
express





_07

Place de l'IST

How are libraries engaging in RDM?

The library is leading on most DCC institutional engagements.

www.dcc.ac.uk/community/institutional-engagements

They are involved in:

- defining the institutional strategy
- developing RDM policy
- delivering training courses
- helping researchers to write DMPs
- advising on data sharing and citation
- setting up data repositories...

Information specialists and data librarians

Who are information specialists and data librarians?

In the context of eResearch, information specialists such as librarians, archivists, curators and records managers may play key roles

Data librarians are professional library staff engaged in managing research data, using research data as a resource, or supporting researchers in these activities.

The screenshot shows a webpage with the following content:

- The Role of Librarians**
Librarians have a wealth of skills and experience to bring to data management.
 - Resource description and discovery
 - Scholarly publishing, open access and metrics
 - Delivering training and information literacy programs
 - Working with research staff and students to understand information needs... and more
- Research Data Portals**
 - Australia**
 - Research Data Australia
 - Atlas of Living Australia
 - Australian Data Archive
 - Topical Data Hub
 - Australian Open Data Network
 - CSIRO Data Access Portal
 - data.gov.au
 - Your institutional data repository
 - ... more
 - International**
 - Government data portals - data.gov.uk, data.gov.au, ...
 - Discipline specific - Pangea, Gene Expression Omnibus, Protein Data Bank, CASP ...
 - Organizational - NASA, World Health Organization, JPL/Caltech, ...
 - Lots of data repositories - DataCite, Zenodo, ...
 - General - Dryad, Figshare, ...
- Partner**
 - Link your institutional data repositories to key web pages and your library catalogue
 - Connect data with published articles in institutional repositories and data sets
 - Include data repositories in federated searches
 - Get your data repositories registered with publishers
 - Include data collections and repositories in your collection building strategies
- Outreach**
 - Include data in your FAQs, resources and chatbot guides, as well as training sessions
 - Talk about data resources in reference queries and your guest lecture slots
 - Encourage your researchers to deposit their data into a repository; work with them to achieve this
 - Ensure your library staff are "data savvy" and ready for the next wave of the information revolution

At the bottom, it says: "Investigate the Research Data Librarians playlist on our YouTube channel" with the URL www.youtube.com/andsdata.

Research Data and Librarians

Australian National Data Service



ands.org.au

<http://www.ands.org.au/guides/dmframework/dmskills-information.html>

What do information specialists do?

The role may include supporting researchers or institutional initiatives in the following areas:

- Data management
 - data management planning
 - issues such as copyright, intellectual property, licensing of data, embargoes, ethics and re-use, privacy
 - storing and managing data during the research project (curation)
 - depositing data in archives at the end of the project, determining retention and disposal
 - open access and publishing of data
 - research organisation policies affecting data
- Metadata management
 - creating and maintaining metadata
 - developing and applying metadata standards
- Using data (data as a resource)
 - finding or obtaining data for re-use
 - citing data
 - data analysis tools and support services
 - data literacy (an extension of information literacy to include the ability to "access, assess, manipulate, summarize and present data". [Read more](#))

The role may also include:

- developing, delivering or arranging
 - resources such as data management checklists
 - training sessions, on topics such as data management planning, data literacy, use of statistical and analytical tools
 - awareness sessions or materials
- referral to sources of information and advice, either within or external to the organisation

[Reading list for information specialists beginning data management projects](#) [PDF 76.9KB]



Données de la recherche



L'équipe Valorisation des données de la recherche met ses connaissances et compétences au service des chercheurs, laboratoires, groupements ou équipes de recherche de l'ESR, en proposant un accompagnement personnalisé pour la gestion et valorisation de leurs données de recherche.

Cet accompagnement est réalisé soit à distance (visio- ou audio-conférence), soit en présentiel (d'une durée de quelques jours, répétées si nécessaire).

L'accompagnement porte sur les actions suivantes :

- ◆ bilan des pratiques de gestion des données de recherche ;
- ◆ élaboration d'un modèle et/ou plan de gestion des données ;
- ◆ élaboration d'un guide de bonnes pratiques adapté à vos besoins et usages ;
- ◆ assistance/conseils pour l'intégration des bonnes pratiques de gestion dans les processus/workflow de recherche et dans le choix d'outils de gestion ;
- ◆ valorisation des collections de données existantes : curation des métadonnées (standardisation/normalisation, enrichissement de la description des données) ;
- ◆ assistance/conseils pour la diffusion et valorisation de vos données : aide au dépôt dans des entrepôts, à la publication dans des data papers (publication de données) ;
- ◆ attribution d'un identifiant pérenne aux objets issus de la recherche (DOI DataCite).

En résumé : apports possibles de l'IST

- ❖ Contribuer à l'émergence d'une politique institutionnelle
- ❖ Aider à définir les stratégies de publication / valorisation
- ❖ Recenser et valoriser les sources de données
- ❖ Intervenir dans la gestion des données : DOI, métadonnées associées, interopérabilité sémantique
- ❖ Sensibiliser les scientifiques à la gestion et à la valorisation des données
- ❖ Etre le relais des bonnes pratiques...

Trois axes de travail



Les compétences



• Traditionnelles

- Connaissance des entrepôts de publications
- Mécanismes de citation et d'évaluation (métries)
- Connaissances juridiques (CC, PI, BdD)
- Maintenance de référentiels
- Interopérabilité (OAI-PMH)
- Métadonnées documents (Dublin core, etc.)

• Nouvelles

- Identifiants (DOI ...)
- Métadonnées datasets (DataCite, standards spécifiques)
- Compréhension des nouvelles voies de publication (Data Journals ...)
- Connaissance des entrepôts de données
- Linked Open Data, ontologies et standards associés



Challenges...



- ❖ S'engager dans de nouvelles activités
- ❖ Etre crédible dans un nouveau domaine d'expertise
- ❖ Acquérir de nouvelles compétences, « *getting techie* »
- ❖ «Essayer les plâtres» sur de nouveaux objets
(Data, Datasets ...)

Se former pour ensuite pouvoir sensibiliser – accompagner – former...



Home Thesaurus Data Management Data Literacy Scholarly Communication Trends & Tech Science Resources

http://esciencelibrary.umassmed.edu/lib_roles

Next Steps: DIL Toolkit



- A guide for librarians seeking to develop DIL Programs of their own
- Developed from the shared experiences of the 5 project teams
- Comprised of:
 - User Guide
 - Case Studies
 - Program Materials



(Carlson, 2014)



Home About Acknowledgements DIY Training Kit for Librarians Feedback Contact Us

Do-It-Yourself Research Data Management Training Kit for Librarians

Provided by EDINA and Data Library, University of Edinburgh in association with the UK Data Archive, Digital Curation Centre (DCC), and Distributed Data Curation Center at the Purdue University Libraries

Page contents

- Introduction
- Training kit contents
- Downloadable contents by session
- Independent study: Data Curation Profiles
- For further study and engagement
- How to Reuse and attribute this content

<http://datalib.edina.ac.uk/mantra/libtraining.html>



(Jones, Guy, & Pickton, 2013)



23 Things: Libraries for Research Data - by RDA's Libraries for Research Data Interest Group

[Home](#) » [Reports & Publications](#) » 23 Things: Libraries For Research Data - By RDA's Libraries For Research Data Interest Group

21
AUG



23 Things: Libraries For Research Data - By RDA's Libraries For Research Data Interest Group

21 August 2015

An overview of practical, free, online resources and tools that you can begin using today to incorporate research data management into your practice of librarianship

23 Things: Libraries for Research Data was presented on August 17 at the International Federation of Library Associations' annual conference in Cape Town, South Africa, in a session that was jointly sponsored by the RDA.

The presentation was based on a **web resource** of the same name that was developed by the RDA's Libraries for Research Data Interest Group that

RDA News

RDA Data Share Announces Fellowship Awards

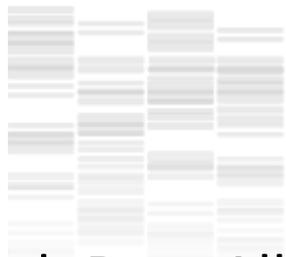
25 May 2016

RDA Data Share, an early career engagement program for the Research Data Alliance/US (RDA/US), is... [Read more](#)

"Funding research data management and related infrastructures" - a Science Europe & Knowledge Exchange report



9- Place de l'IST à l'INRA



À l'international

- ❖ Research Data Alliance (RDA)
 - Groupe d'intérêt « Agricultural Data »
 - Groupe de travail « Wheat Data Interoperability »
 - Groupe de travail « Agrisemantics »
- ❖ Global Open Data for Agriculture and Nutrition (GODAN)
- ❖ Coherence in Information for Agriculture Research (CIARD)
- ❖ Science Europe

Au sein de l'INRA

❖ Chantier « Data Partage »

- Co-pilotage du chantier / O. Hologne – Directrice IST
- Implication dans des groupes de travail <https://wiki.inra.fr/wiki/donneesrechercheist>
 - Données et publications
 - Plan de gestion des données
 - Entrepôt/annuaire
 - Familles de données
 - Communication

❖ Pôle DigitalIST (Données de la recherche & IST)

❖ Pôle GeCo (Gestion des connaissances)

Scientifiques – IST: les points d'entrée

- ❖ Partage de données soutenant une publication
 - L'éditeur me demande de déposer mes données dans un entrepôt et de fournir un lien d'accès pérenne
 - Où déposer mes données?
 - Comment donner accès uniquement aux reviewers tant que mon article n'est pas publié?
 - Que se passe-t-il pour mes données si mon article n'est pas accepté?
 - Quelle licence?

- ❖ Besoin de DOI
 - Comment obtenir un DOI?
 - Quelles données?
 - Quelles conditions particulières?

- ❖ Intérêt pour les Data papers
 - Qu'est-ce qu'un data paper?
 - Quelle revue pour publier mon data paper?
 - Quelles précautions?
 - Par où commencer?

Gestion

Tout ce qui relève d'un des axes

Valorisation

Réutilisation

Information/Formation

Les membres du pôle

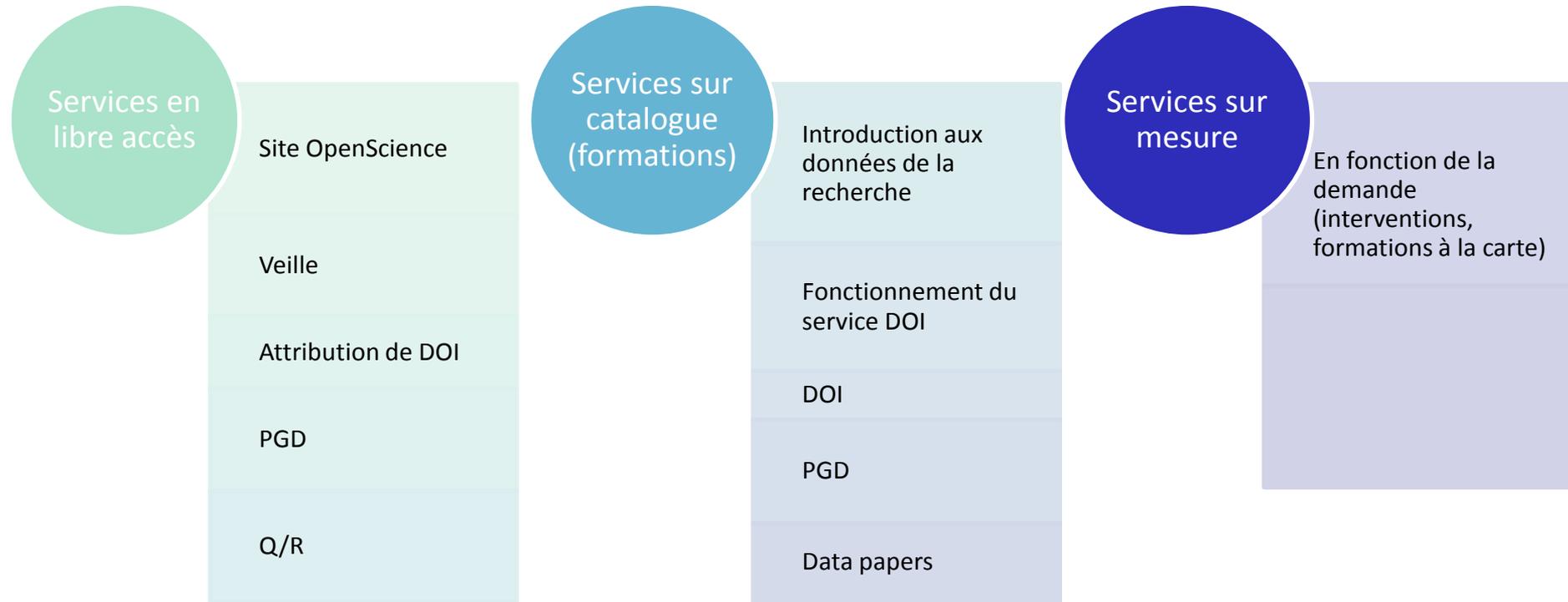
Di – Pôle Données de la recherche

!!!! Ce s... n !!!! Il sera un outil pour l'animation du pôle

Les membres du pôle

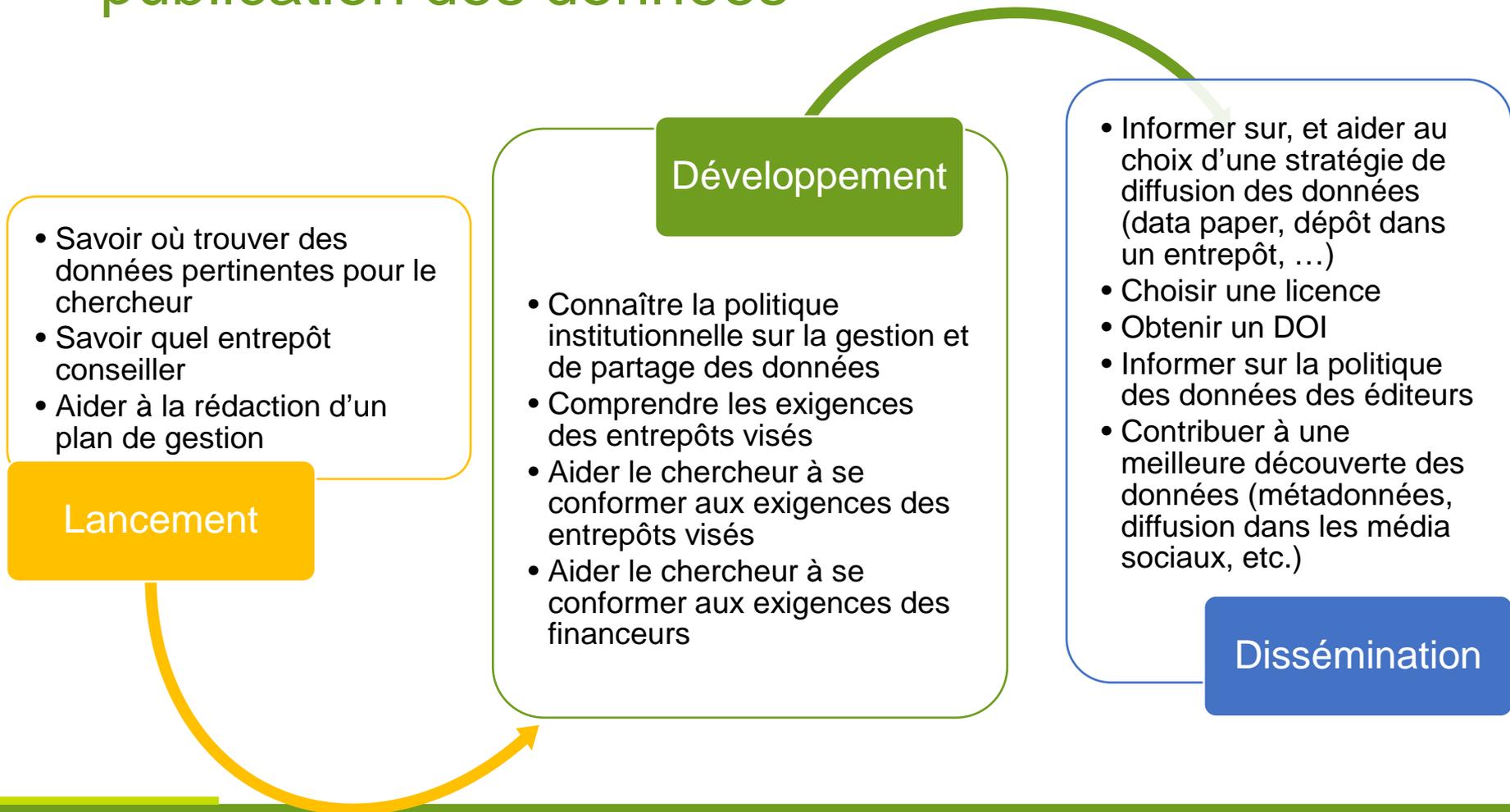
- Agnès Girard (**Gestion**)
- Anne-Sophie Martel (Gestion)
- Bruno Pierrel (Réutilisation, **Formation/Information**)
- Christine Molé (Gestion)
- Diane Le Hénaff (Formation/Information)
- Dominique Fournier (Gestion, Valorisation, Formation/Information)
- Dominique L'Hostis (Gestion, **Formation/Information**)
- Eric Marchoux (Gestion, Valorisation)
- **Esther Dzalé Yeumo** (animatrice du pôle)
- Hugues Leiser (Valorisation)
- Magalie Weber (Gestion)
- Marianne Peiffer (Gestion)
- Marie-Violaine Tatry (Valorisation)
- Marjolaine Hamelin (Valorisation, Formation/Information)
- Mathieu Andro (Valorisation, Réutilisation)
- Monique Delabuis (Gestion)
- Nathaële Wacrenier (Valorisation, Formation/Information)
- Odile Hologne (Formation/Information)
- Odile Viseux
- Pascal Aventurier (Gestion, **Valorisation**, Formation/Information)
- Sophie Aubin (Gestion, Réutilisation)
- Sylvie Cocaud (**Gestion**, Valorisation)
- Véronique Decognet (Gestion)
- Virginie Lelièvre (Valorisation, Formation/Information)

Offre de service DigitalIST



A l'INRA. Une vision possible du futur ?

Aide aux chercheurs pour la gestion et la publication des données





Conclusion

- ❖ Beaucoup d'actions possibles
 - ✓ Publication, inventaire, gestion
 - Lesquels cibler ?
 - Qui sont vos alliés ?
- ❖ Des atouts et un socle de compétences recyclables
- ❖ De nouvelles compétences à développer
- ❖ Quels conseils ?
 - ✓ Comprendre sa place dans le système
 - ✓ Aller voir ailleurs
 - ✓ Ne pas rester seul



Questions?

Bibliographie – « Données et IST »

- ❖ Ball, A. (2013). [Les métiers liés aux données de la recherche : Data Librarian].
- ❖ Carlson, J. (2014). [The data information literacy project]. <http://www.duraspace.org/esi-logistics>
- ❖ Clark, D. (2014). [Getting onboard the data training: How librarians fit in]. http://fr.slideshare.net/di_clark/getting-onboard-the-data-train
- ❖ Hügi, J., & Prongué, N. (2014). Le virage Linked Open Data en bibliothèque : étude des pratiques, mise en œuvre, compétences des professionnels. *Ressi*, (Décembre). http://www.ressi.ch/num15/article_100http://www.ressi.ch/num15/article_100
- ❖ Lyon, L. (2012). *The informatics transform : re-engineering libraries for the Data Decade*. Paper presented at the VALA2012 PLENARY 4. <http://www.vala.org.au/vala2012-proceedings/vala2012-plenary-4-lyon>
<http://webcast.gigtv.com.au/Mediasite/Play/90077d28660e4c0487f8cf293d639e071d>
- ❖ Lyon, L. (2012). The Informatics Transform: Re-Engineering Libraries for the Data Decade. *International Journal of Digital Curation*, 7(1), 126-138. <http://dx.doi.org/10.2218/ijdc.v7i1.220>
- ❖ MacMillan, D. (2014). Data Sharing and Discovery: What Librarians Need to Know. *The Journal of Academic Librarianship*, 40(5), 541-549. [10.1016/j.acalib.2014.06.011](http://dx.doi.org/10.1016/j.acalib.2014.06.011)
- ❖ Martin, V. (2014). *Demystifying eResearch. A primer for librarians*. Santa Barbara (USA): ABC-CLIO, 189 p. (ISBN = 978-1-61069-520-6, e-ISBN = 978-1-61069-521-3)
- ❖ Tenopir, C.; Sandusky, R.J.; Allard, S.; Birch, B., 2014. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 36 (2): 84-90. [10.1016/j.lisr.2013.11.003](http://dx.doi.org/10.1016/j.lisr.2013.11.003)
- ❖ Dzalé Yeumo, E.; L'Hostis D., 2015. Formation données de la recherche Agropolis
- ❖ Hologne O, 2015. Formation données de la recherche ENSSIB



Merci de votre écoute

Esther Dzale Yeumo edzale@versailles.inra.fr



@edzale