

# *Groupe données d'enquêtes et analyse textuelle*

## 1 Composition du groupe

---

**Coordination** : Eric Cahuzac (ODR, SAE2) & Christine Boizot (ALISS, SAE2)

**Membres** : Christophe Bontemps (GREMAQ, SAE2), Pierre Triboulet (AGIR, SAD), Véronique Batifol-Garandel (IST, SAD), Patrice Buche (IAT, CEPIA), Lydie Soler (MIA 518, MIA), Philippe Breucker (SENS, SAE2/SAD), Annie Hofstetter (LAMETA, SAE2), Jean-Marc Rousselle (LAMETA, SAE2) Jean-Loup Dupuy & Cédric Lanu (GAEL, SAE2).

Dans ce panel se dégagent les sensibilités de **4 sous familles** (Analyse textuelle ; Économie expérimentale, Analyse d'enquête [qualitatives/quantitatives], Production d'enquêtes), de **3 Cati** (ICAT, ACTION, CITISES) et de **l'IST**.

## 2 Table des matières

---

1	COMPOSITION DU GROUPE.....	1
3	RÉSUMÉ .....	2
4	L'ÉCOSYSTÈME INFORMATIONNEL.....	2
4.1	Que souhaite-on partager? .....	2
4.2	Importance du niveau de granularité de la donnée .....	3
4.3	Importance des aspects de propriété intellectuelle et du cadre juridique .....	3
4.4	Dans quel objectif et pour qui partager ? .....	3
4.5	L'écosystème des données d'enquêtes .....	5
5	IDENTIFICATION DU PATRIMOINE HISTORIQUE À PÉRENNISER.....	6
6	CRÉNEAUX DE DONNÉES ORIGINALES POUR SE POSITIONNER À L'INTERNATIONAL – UNE VISION STRATÉGIQUE DU PARTAGE .....	6
7	CONCLUSION : ATTENTES EN TERME D'OFFRE INRA .....	8
8	ANNEXE.....	9

## 3 Résumé

---

Ce travail s'est étalé sur dix-huit mois et a demandé plusieurs rencontres et visioconférences. La démarche en a été la suivante :

Sur la base d'un tableau collectant les contributions d'un certain nombre de chercheurs (Cf. annexe), les premières discussions du groupe ont porté sur les données qu'il était souhaitable de partager et les difficultés que l'on pouvait rencontrer.

- Quelles données peut-on diffuser et qu'est-ce qui est intéressant de diffuser ?
  - Les données ayant servi à publication sont intéressantes ;
  - Mais ce ne sont pas les seules.
- Le partage de données s'il s'adresse à un large public doit porter sur les données qui ont été validées, certifiées :
  - Certification par les éditeurs : données qui ont servi à publication dans revues ;
  - Certification par un organisme certificateur : données de plateforme certifiées.

Il existe différents points d'entrée

- Données produites par l'INRA ou venant d'autres sources avec apport de valeur ajoutée par l'INRA ;
- Données ayant ou non donné lieu à publication ;
- Partage qui se pose différemment si grand public ou si chercheurs.

## 4 L'écosystème informationnel

---

### 4.1 Que souhaite-on partager?

Il résulte de nos différentes discussions une position claire qui est que pour publier des données au nom de l'INRA, il faut que les données soient propres, bien décrites et validées. C'est notamment le cas pour les données qui ont servi à publication. L'accent doit être donc mis en priorité sur le partage de données ayant subi des prétraitements et non pas sur des données brutes.

Pour les autres données, qui n'ont pas été utilisées dans des publications ou qui ne sont pas fournies par des plateformes certifiées, la tâche est plus délicate et il faut analyser au cas par cas les difficultés. Un premier stade est de montrer que l'information existe mais ensuite c'est au producteur de la donnée de trancher et de décider de partager ou pas. A lui aussi de définir les modalités de diffusion et de collaboration avec le demandeur.

**→ On commence par partager les données qui ont servi à publication et ensuite on se préoccupe des autres qui ont aussi de l'intérêt.**

## 4.2 Importance du niveau de granularité de la donnée

En SHS on manipule des données individuelles sensibles, qui peuvent porter sur de petits échantillons. Il ne faut pas donner la possibilité de remonter à l'individu, c'est-à-dire pouvoir l'identifier. A priori les données portant sur les individus ne peuvent être diffusées, sauf exception :

- Les données d'économie expérimentale. Celles-ci respectent l'anonymat des sujets puisque dès que le sujet quitte la salle on ne peut plus retrouver l'individu ;
- Les données agrégées (ou floutées, moyennées) selon diverses typologies, et respectant les règles du secret statistique de telle sorte que l'on ne puisse pas remonter à la source.

De manière générale, quand on utilise des données SHS et que l'on se projette dans un processus de partage, il faut savoir à quel niveau d'agrégation on aura le droit de les diffuser, c'est-à-dire à quel niveau elles seront considérées comme publiques. Il faut donc connaître le « grain » le plus fin à partir duquel les données sont diffusables.

Ceci permettrait de caractériser, pour chaque jeu de données, le degré d'agrégation nécessaire à la diffusion. Par exemple : aucun pour les données d'économie expérimentale (cf. supra) ; selon un critère géographique (commune, département,...) pour les données géo-référencées ; par secteur d'activité (APE) pour des données d'enquêtes et/ou des données achetées. Cette information, qu'il faudrait préciser, pourrait être attachée aux données brutes (via les métadonnées) ce qui permettrait de clarifier l'utilisation et la diffusion des données.

Enfin cette notion de « grain » pourrait être multidimensionnelle, c'est-à-dire qu'on pourrait proposer des « grains » différents selon la population amenée à utiliser les données (chercheurs vs grand public).

## 4.3 Importance des aspects de propriété intellectuelle et du cadre juridique

À partir de quand l'enrichissement d'un jeu de données en fait-il un nouveau jeu de données original ?

Dans le cadre de l'analyse textuelle de documents, des données de la littérature sont utilisées afin d'alimenter des outils d'aide à la décision. De vraies problématiques juridiques se posent et un accompagnement doit être mis en place pour permettre aux chercheurs de savoir dans quel cadre ils peuvent partager ces données enrichies et les valoriser.

Au-delà de ces exemples, un réel besoin d'accompagnement se fait ressentir de la part des chercheurs et des producteurs de données secondaires. Ils souhaiteraient savoir ce qu'ils ont le droit de faire, de diffuser en matière de données et surtout comment le contractualiser avec leurs partenaires.

## 4.4 Dans quel objectif et pour qui partager ?

Certains cas ne posent pas de problèmes car ils ont déjà été identifiés et traités par le passé :

- *Pour répliquer les résultats de publication.*

Beaucoup de revues demandent cela, et pas qu'en SHS, pour permettre une meilleure évaluation de l'article et éviter la fraude. En soi, ce n'est pas un problème si le processus est sécurisé. Mais **dans certains cas on ne pourra pas publier dans des revues qui demandent les données brutes car il y aura impossibilité de céder les droits.**

- *Pour partager les résultats des travaux de recherche publiés.*  
Dans la majorité des cas les plateformes IST sont là pour cela et font déjà ce travail (ProdInra, Hal,...)
- *Pour une meilleure dynamique de recherche (en interne à la recherche).*  
Dans ce cas il faut favoriser les plateformes et la diffusion de données à partir de celles-ci. Ce sont des outils adaptés pour cela. Elles gèrent elle-même les droits sur les données, les métadonnées et leurs utilisateurs.
- *Dans le cadre de projets ANR ou EU.*  
La diffusion dans ce cas est encadrée par l'accord de consortium du projet. L'objectif est d'alerter sur l'importance de ce consortium dans les prochains contrats. De même, il faut sensibiliser les chercheurs sur la demande croissante des financeurs de fournir un plan de gestion des données prévoyant de manière explicite la question de la réutilisation des données.

Même s'il subsistera des cas particuliers dans les points ci-dessus, le cadre législatif ou contractuel est présent et pourra s'adapter en fonction des demandes.

Lorsque le partage de la donnée a pour objectif d'informer le citoyen ou de partager la donnée avec une communauté externe, la mise à disposition dépend en premier lieu de l'origine de la donnée (voir tableau en Annexe) :

- *Dans le cadre des données textuelles*, on réemploie, par exemple, des données de publication. L'objectif est que les données de bases (brutes) puissent être utilisées. Les données viennent de sources qui ont des droits mais qui ont été publiées. Les travaux sur des données textuelles permettent d'enrichir ses données avec des métadonnées. Ces travaux ont aussi pour résultats la construction de référentiels, d'ontologies spécifiques aux besoins de l'INRA et de nouvelles bases de données. Le partage est dans ce cas soumis aux conditions juridiques de réutilisation de la donnée qui ne sont pas si évidentes à trancher pour un chercheur (voir 4.3).
- *Dans le cadre de données d'enquêtes* (socio, éco) auprès d'agriculteurs, de sociétés,... on travaille sur de trop petits échantillons pour rester dans l'anonymat. La publication des données brutes paraît impossible, sauf si des agrégations peuvent être réalisées permettant de respecter l'anonymat (cf. supra). Dans les autres cas, la diffusion peut se faire en contactant le chercheur au cas par cas.
- *Dans le cadre de données d'économie expérimentale*, les jeux de données sont anonymes donc il n'y a pas de difficultés pour diffuser les données. Cependant, la diffusion à un large public nécessite la mise en forme et une bonne information sur les données afin qu'elles puissent être utilisées sans difficulté et à bon escient.
- *Dans le cadre de données achetées et/ou pour lesquelles les fournisseurs demandent la confidentialité*, les publications réalisées à partir de ces données sont les seules informations qui puissent être diffusées. Ces publications informent de l'existence de telles bases et à la demande (au cas par cas) des données agrégées peuvent, la plupart du temps, être diffusées (si elles n'existent pas par ailleurs). Mais les données brutes ne peuvent être diffusées sans l'accord de l'auteur qui généralement n'est pas l'INRA. Problèmes de licence et de droits.

Dans tous les cas, la mise en forme, l'enrichissement et tout le travail que fait l'INRA pour faciliter la mise à disposition de données élaborées à partir des données brutes doit être pris en compte et valorisé.

## 4.5 L'écosystème des données d'enquêtes

Par écosystème, nous ne nous limitons pas ici uniquement à la notion d'entrepôts déjà existants, mais nous regardons dans quels domaines (économiques, alimentation etc..) des jeux de données sont déjà publiés et sous quelle forme (données "brutes", vocabulaires etc..). Cette vision de ce qui se fait autour de nous, nous aide à identifier chez nous les données à publier.

L'entrepôt de données le plus proche de nos thématiques est celui du Réseau Quételet (<http://www.reseau-quetelet.cnrs.fr>). C'est le plus ancien et le plus important entrepôt de données dédié aux données quantitatives de la recherche en SHS. Cet entrepôt œuvre dans l'archivage, la documentation et le dépôt des données en vue de leur réutilisation pour la recherche nationale et internationale. Il met à disposition des chercheurs les grandes enquêtes, recensements et autres bases de données issues de la statistique publique française ; les enquêtes françaises provenant de la recherche ; certaines enquêtes internationales. Quételet s'ouvre maintenant, peu à peu sur les données qualitatives notamment de science politique et de sociologie.

Cette plateforme est secondée par des initiatives comme la plateforme universitaire des données de l'université de Lille (<http://www.meshs.fr/PUDL/>) qui lui sert de réseau local

D'autres plateformes généralistes peuvent donner accès dans la thématique à des jeux de données comme Datahub (<http://datahub.io>). Ce sont des outils mis à disposition de communautés de chercheurs (ou pas) pour partager de l'information, pouvant aller jusqu'à la donnée, mais principalement des métadonnées, au niveau international. Dans cette catégorie on retrouve TGIR Huma-Num (<http://www.huma-num.fr/>) qui est une sorte d'UDS (Uniform Data System) visant à favoriser le partage et la diffusion des recherches en SHS. Ils offrent des services comprenant un service d'archivage à long terme et une plateforme de mise en relation de données. Il y a une moissonneuse qui analyse des textes et en retire des métadonnées qui ensuite sont interrogeables. C'est aussi un gros silo de données open (<http://www.rechercheisidore.fr/>).

D'autres initiatives dans le domaine peuvent être citées, car utile dans le domaine de l'échange informationnel, comme un entrepôt pour stocker du patrimoine sonore (<http://phonotheque.mmsch.univ-aix.fr>) ou écrit (<http://www.numes.fr/numes/mainMenu.html>) ou des plateformes de publications d'ontologies, dans le domaine de l'alimentation par exemple (<http://data.lirmm.fr/ontologies/food>).

Enfin, on peut noter les « solutions logicielles » comme Nesstar (<http://www.nesstar.com>) qui permettent la construction d'un environnement web facilitant la diffusion et le partage d'informations sur les données d'enquêtes (métadonnées, statistiques, etc.) dans le monde académique et le secteur public. Cette solution est assez largement utilisée aux états unis et en Europe.

**Le partage le dépôt et la diffusion de nos données sur des plateformes existantes est donc possible. Il pose cependant plusieurs questions, celle de la pérennité, de la visibilité et du financement de ces plateformes, mais aussi celle de l'anonymisation des données individuelles sensibles.**

**Il faut cependant distinguer les grandes infrastructures nationales (Quetelet, Huma-Num,...), soutenues par les ministères et partagées avec d'autres organismes, qui peuvent nous fournir une solution d'hébergement et de services utiles et pérennes, des solutions locales ou privées, plus généralistes dont on ne connaît pas la durée de vie ni la finalité.**

La dimension « entrepôt national de données d'enquêtes » est en conclusion assez peu développée en France pour les SHS (hormis pour les grandes enquêtes nationales) et pas du tout pour les enquêtes relevant de travaux de recherches relatifs aux thématiques INRA. Seuls les établissements publics nationaux (INSEE, SSP,...) fournissent des informations issues de leurs enquêtes, mais ne proposent pas d'héberger les données des tiers.

Ce constat s'explique d'une part car nous mobilisons souvent dans nos travaux de recherche, des données produites par des tiers et qu'il nous est impossible de les partager, et d'autre part parce que la question de la confidentialité des données individuelles est souvent un obstacle à l'intérêt qu'il y aurait à les partager.

## 5 Identification du patrimoine historique à pérenniser

---

Étant donné le faible volume de nos données en SHS au regard des données d'autres disciplines, il semble qu'il n'y a pas lieu d'avoir une politique de gestion très restrictive quant à l'archivage ou la destruction des données produites passé un certain délai. Elles pourront toutes sans problème être archivées aux côtés de données plus récentes et pourront servir pour des travaux de dimension temporelle, notamment pour les historiens.

Toutes les données qui ont servi à publication, toutes les données achetées, toutes les données pour lesquelles un protocole a été déposé/validé pourraient entrer dans le patrimoine historique de l'INRA. Il faudra cependant veiller à respecter les recommandations du CNIS et les obligations réglementaires vis à vis des fournisseurs qui imposent souvent une destruction des données au bout d'une période définie. Il pourrait aussi être bon de sauvegarder des données publiques sur certains domaines en "moissonnant" le web, car les données de prix de certains sites en ligne par exemple, ou de certaines institutions (nouvelles du marché, etc...) sont accessibles et puis sont actualisées et disparaissent...

L'expérience a montré par le passé que certaines enquêtes ou données pouvaient disparaître de chez les producteurs (y compris les plus sérieux), lors de restructuration ou de changement de système informatique. Cette réflexion est à la marge de l'open data à l'INRA mais soulève quand même le fait que la question patrimoniale est une question sensible, autant pour les producteurs de données que pour les chercheurs qui les utilisent, mais qu'elle n'est peut-être pas abordée de la même manière dans les deux cas. En SHS plus particulièrement, la pérennité des données, l'aspect longitudinal, panel ou historique des données sont des critères importants pour les chercheurs qui s'intéressent aux évolutions de comportements de structures ou d'observations. La profondeur historique est une dimension très importante qui entre dans le cycle de vie de la donnée et qui pose la question de l'archivage. La conservation de données à usage interne pour des finalités différentes, pose alors des questions juridiques.

## 6 Créneaux de données originales pour se positionner à l'international – une vision stratégique du partage

---

L'originalité des données ne provient pas seulement du fait que les données soient uniques – critère qui est atténué par le fait qu'il y a de moins en moins de production d'enquêtes originales en SHS – mais surtout de la plus-value apportée par la recherche sur des données qui peuvent être « standardisées ».

Cette plus-value peut provenir de l'enrichissement des données d'enquêtes par croisement/rapprochement d'informations provenant de sources différentes, par le calcul de nouveaux indicateurs, par modélisation ou agrégation.

L'originalité peut provenir également de la profondeur historique de la source ou de son aspect longitudinal/panel (les mêmes individus sont suivis dans le temps). Ces dernières caractéristiques ouvrent la voie à des études plus approfondies, à des modélisations plus riches ou à des indicateurs plus sophistiqués que lors de l'utilisation de données en coupe.

Selon les conclusions des juristes, on peut considérer que ces données « secondaires » – qui peuvent être produites à partir de données « primaires » qui ne nous appartiennent pas – sont des données, ou des constructions de données originales et que nous en sommes propriétaires. Dans tous les cas, des moyens ont été mis en œuvre dans la production et la mise en forme de cette nouvelle information, que l'on peut valoriser de façon stratégique en partageant cette information. Les plateformes et systèmes d'informations à l'INRA sont alors les plus appropriés pour réaliser ce travail.

## 7 Conclusion : attentes en terme d'offre Inra

---

Il va de soi que le partage et la diffusion des données de la recherche est indispensable pour accroître la visibilité de nos travaux en interne comme à l'international. Comme nous avons pu le voir (cf. § 4.4) des structures proposant des services de partage de données avec une ouverture vers l'extérieur existent déjà dans notre famille. Elles sont déjà organisées et référencées pour une parfaite lisibilité tant des données que des métadonnées. Sous réserve du respect des principales conditions liées au partage de données individuelles sensibles, et d'ajustement aux spécificités thématiques de l'INRA, ces structures sont un bon compromis nous permettant de partager nos données dans les conditions les moins onéreuses. Le choix de la plateforme (Quêtelet, TGIR Huma-Num) dépendra alors de la discipline ou de décisions internes mais il ne nous paraît pas opportun à ce stade de développer un système concurrent ou parallèle dédié à nos propres thématiques.

Là où il faudra être vigilant, c'est dans le cadre du partage de la donnée à des fins de reproductibilité des travaux de recherche (cf. § 4.3, point 1). Il n'est pas envisageable de céder aux éditeurs de revues les droits sur les données utilisées dans un travail de recherche. Soit ces données ne nous appartiennent pas et la convention qui a encadré leur utilisation pour la recherche ne le permet pas. Soit le contrat avec le producteur nous l'interdit pour des raisons de propriété intellectuelle ou de droit à la concurrence. Soit enfin, céder les données brutes serait un manquement au respect du secret statistique et aux règles relatives au respect de la personne. De toute façon ce n'est plus possible actuellement dans le cadre de la mise à disposition des enquêtes via le CASD)

**Notre proposition** est donc que l'INRA fournisse un service opérationnel qui réponde à cette attente des éditeurs avant que ces derniers ne proposent une solution qui nous ferait perdre les droits sur les données ou la possibilité de publier dans les revues. Cette solution pourrait être assimilée à un serveur sécurisé, dans le même esprit que le Centre d'Accès Sécurisé à Distance (CASD) du Genes (Groupe des écoles nationales d'économie et statistique). Ce serveur serait dédié aux données qui ont servi dans les publications scientifiques des chercheurs et proposerait les outils nécessaires à la reproductibilité des analyses effectuées dans la publication. Cette solution garantirait la sécurité des données et donnerait la preuve de l'exactitude des résultats. Il faudrait attacher un DOI à chaque jeu de données pour donner un accès aux éditeurs dans un but de reproduction/vérification mais aussi aux chercheurs désireux de refaire la publication pour l'enrichir par des aspects scientifiques ou méthodologiques nouveaux.

Au-delà de ce besoin spécifique **il faut continuer** en interne à développer les plateformes thématiques de mutualisation de données et de services attachés à ces données. En mutualisant les moyens et en favorisant le partage de connaissance autour des données, nous gagnons en efficacité sur le traitement de l'information et la production de résultats novateurs. Ces plateformes sont le complément à la proposition ci-dessus, car elles rassembleront des données qui n'auront pas forcément servi à publication et donc ne seront pas forcément validées. Néanmoins elles pourront bénéficier, tout au long du cycle de vie, de vérifications et d'enrichissements dans le respect de la qualité.



## 8 Annexe

Tableau récapitulatif des données produites en SHS									
Données Questions	Enquêtes auprès de petits échantillons	Enquêtes sociologiques auprès d'échantillons plus importants	Données expérimentales, en générales économiques	Données de relevés d'étiquetage ou d'analyse	Données d'achats harmonisées ou agrégées (comme Kantar)	Séries historiques constituées à partir d'autres données informatisées ou non.	Données textuelles	Jeux de données pour les revues	Résultats (tableaux statistiques, élasticités,...)
Type de données produites	Originales, qualitatives, parfois quantitatives, brutes	Originales, qualitatives, quantitatives, brutes	Originales, quantitatives, brutes	Originales, qualitatives, quantitatives, brutes	De seconde main, quantitatives, élaborées	De seconde main, élaborées	Elaborées	Elaborées	Elaborées
Destinataire et usage initiaux des données	Pour un chercheur, le plus souvent sociologue, démographe, restreint à une recherche ponctuelle	Dans le cadre d'un projet, recherches multiples.	Dans le cadre d'un projet, pour une recherche	Observatoire outil de politique publique	Pour les chercheurs d'une unité	Un chercheur pour une recherche spécifique	Les chercheurs dans le cadre d'un projet	Les éditeurs, la communauté scientifique	La communauté scientifique, les professionnelles, le grand public, les politiques.
Niveau de finesse de la donnée	Individu, entreprise, filrière, exploitation agricole	Ménage, individu	Individu	Référence produit (EAN)	Ménage	Institution, entreprise, population générale	Corpus textuel	Données fines ou agrégées	Ensemble de la population, sous- population, catégories de produits
Qui produit les données	Les chercheurs eux- mêmes	Les chercheurs sont maîtres d'ouvrage, le maître d'œuvre est un prestataire.	Par l'équipe du projet	INRA - Anses	Une unité	Le chercheur lui-même	Les participants au projet	Les chercheurs	Les chercheurs
Etapes de production réalisées en interne	Idée, questionnaire, recueil, traitement	Idée, questionnaire, traitement	Idée, protocole, recueil, traitement	Recueil, traitement	Harmonisation, diffusion	Idée, recueil, traitement	Traitement	Recueil, traitement	Traitement, publication
Support et/ou format de la base de données	Audio, Excel, papier, Stata, Sas, texte	Excel, papier, Stata, Sas, texte.	Excel, Stata, Sas, texte.	Mysql	Postgres, csv, stata	Excel, texte, papier		Texte	Word, Latex, Scientific Word, pdf.
Archivage	Le plus souvent par le chercheur lui-même	Par un ou plusieurs des participants au projet	Par un ou plusieurs des participants au projet	Dans une base de données gérée par des informaticiens	Dans une base de données gérée par des informaticiens	Le plus souvent par le chercheur lui-même		Les chercheurs eux-mêmes et les éditeurs	
Métadonnées	Au bon vouloir du chercheur	Questionnaire, protocole d'enquête.	Protocole d'expérience	Documentation	Documentation	Au bon vouloir du chercheur		Publications	
Données spécifiques ou généralistes	Usage spécifique	Multi-usages spécifique	Usage spécifique	Multi-usages spécifique	Généraliste	Usage spécifique	Usage spécifique	Usage spécifique	Usage spécifique

**Tableau récapitulatif des données produites en SHS (suite et fin)**

Données Questions	Enquêtes auprès de petits échantillons	Enquêtes sociologiques auprès d'échantillons plus importants	Données expérimentales, en générales économiques	Données de relevés d'étiquetage ou d'analyse	Données d'achats harmonisées ou agrégées (comme Kantar)	Séries historiques constituées à partir d'autres données informatisées ou non.	Données textuelles	Jeux de données pour les revues	Résultats (tableaux statistiques, élasticités,...)
<b>Pour qui si élargissement de la diffusion</b>	La communauté scientifique	La communauté scientifique	La communauté scientifique	La communauté scientifique, les professionnels, le grand public	La communauté scientifique INRA	La communauté scientifique INRA		Les chercheurs eux-mêmes et les éditeurs	
<b>Elargissement possible de la diffusion</b>	Si accord des chercheurs producteurs de la données après une période d'embargo, si aucune possibilité d'identifier les entités interrogées, si création d'une documentation et de règles d'utilisation	Si accord des chercheurs producteurs de la données après une période d'embargo, si aucune possibilité d'identifier les entités interrogées, si création d'une documentation et de règles d'utilisation	Si accord des chercheurs producteurs de la données après une période d'embargo, si aucune possibilité d'identifier les entités interrogées, si création d'une documentation et de règles d'utilisation	Données constituées par financement public devant être ouvertes mais en respectant la confidentialité de certaines informations contenues dans la BDD	Pas d'élargissement possible car données privées à usage exclusif de l'INRA	Si accord des chercheurs producteurs de la données après une période d'embargo, si aucune possibilité d'identifier les entités interrogées, si création d'une documentation et de règles d'utilisation		Publications	
<b>Support de diffusion possible</b>	Réseau existant type Quételet, offrir au niveau des unités, des départements et/ou de plateformes existantes la possibilité de déposer des données documentées pour diffusion plus ou moins large (avec différents niveaux d'accès et des règles d'utilisation)	Réseau existant type Quételet, offrir au niveau des unités, des départements et/ou de plateformes existantes la possibilité de déposer des données documentées pour diffusion plus ou moins large (avec différents niveaux d'accès et des règles d'utilisation)	Réseau existant type Quételet, offrir au niveau des unités, des départements et/ou de plateformes existantes la possibilité de déposer des données documentées pour diffusion plus ou moins large (avec différents niveaux d'accès et des règles d'utilisation)	Réseau existant type Quételet, offrir au niveau des unités, des départements et/ou de plateformes existantes la possibilité de déposer des données documentées pour diffusion plus ou moins large (avec différents niveaux d'accès et des règles d'utilisation)	Offrir au niveau des unités, des départements et/ou de plateformes existantes la possibilité de déposer des données documentées pour diffusion plus ou moins large (avec différents niveaux d'accès et des règles d'utilisation)	Réseau existant type Quételet, offrir au niveau des unités, des départements et/ou de plateformes existantes la possibilité de déposer des données documentées pour diffusion plus ou moins large (avec différents niveaux d'accès et des règles d'utilisation)			
<b>Validation et maintenance</b>	Qui valide les données et métadonnées déposées ? Qui fait le "service après- vente" ?	Qui valide les données et métadonnées déposées ? Qui fait le "service après- vente" ?	Qui valide les données et métadonnées déposées ? Qui fait le "service après- vente" ?			Qui valide les données et métadonnées déposées ? Qui fait le "service après- vente" ?			