

## Diplôme national de master

Domaine - sciences humaines et sociales

Mention - sciences de l'information et des bibliothèques

Parcours - archives numériques

# **La gestion des données de la recherche agronomique : de la science ouverte à l'histoire des sciences**

**Guillaume Tuloup**

Sous la direction de Clément Oury  
Conservateur des bibliothèques et Adjoint au chef du service Conservation,  
Restauration, Numérisation – Muséum national d'Histoire naturelle

## **Remerciements**

*Je remercie mon directeur de mémoire Clément Oury pour ses conseils et ses encouragements durant ce travail de recherche.*

*Mes remerciements à Odile Hologne, directrice de la DipSO, Colette Cadiou, responsable du pôle ACDC et Raymond Schiano, archiviste à INRAE, qui m'ont accordé leur temps pour longuement échanger sur leurs activités.*

*J'adresse enfin mes remerciements à Pierre Cornu et Odile Maeght-Bournay qui ont pris le temps de réfléchir sur leurs propres pratiques de recherche et ont su mettre de la perspective à mon étude.*

*A Margaux (bis).*

**Résumé :** *L'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) est engagé depuis plusieurs années dans le développement de la science ouverte. Se positionnant aujourd'hui en fer de lance des mouvements d'ouverture, il œuvre à la rationalisation de la production, de la préservation et de la diffusion des données de la recherche afin de faire des sciences agronomiques des sciences impliquées, dans la société. En effet, confrontée aux problématiques environnementales, sanitaires et alimentaires, la recherche agronomique est engagée dans une restructuration de la production des savoirs sur la nature et le vivant, à l'appui du numérique qui se fait environnement totalisant. L'Institut étant particulièrement sensibilisé aux enjeux mémoriels et historiques, l'histoire des sciences peut restituer les conditions historiques de cette restructuration, en mobilisant elle-même les ressorts du numérique et de la science ouverte.*

*Descripteurs : Recherche agronomique – Données de la recherche – Histoire des sciences - Archivage*

*Abstract : The National Research Institute for Agriculture, Food and Environment (INRAE) has been involved in the development of open science for several years. Today, INRAE is cutting edge the open science movement and is working to rationalize the production, preservation and dissemination of research data in order to make agricultural sciences more socially relevant. Indeed, faced with environmental, health and food issues, agricultural research is engaged in restructuring the production of knowledge on nature and living organisms, with the help of digital technology which becomes a totalizing environment. Particularly aware of the historical and memorial stakes, the history of science can restore the historical conditions of this restructuring, by mobilizing the resources of digital and open science.*

*Keywords : Agronomic research - Research data - History of science – e-Archiving*

### **Droits d'auteur**



Cette création est mise à disposition selon le Contrat : « **Paternité-Pas d'Utilisation Commerciale-Pas de Modification 4.0 France** » disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr> ou par courrier postal à Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

*« Je me suis efforcé de décrire le monde,  
non pas comme il est, mais comme il est quand je  
m'y ajoute, ce qui, évidemment, ne le simplifie pas. »*

Jean Giono, *Voyage en Italie*, 1979.

# *Sommaire*

<b>SIGLES ET ABREVIATIONS .....</b>	<b>5</b>
<b>INTRODUCTION.....</b>	<b>7</b>
<b>LA SCIENCE OUVERTE ET LE LIBRE ACCES : DE NOUVELLES PRATIQUES POUR UNE NOUVELLE SCIENCE .....</b>	<b>14</b>
La fabrique du savoir .....	14
Les ambitions du partage des données.....	26
Une double dynamique internationale et nationale .....	41
<b>INRAE ET LA GESTION DES DONNEES DE LA RECHERCHE .....</b>	<b>55</b>
L’engagement précurseur des instituts français de recherche agronomique .....	55
La gouvernance des données à INRAE.....	71
La gestion des données de la recherche : de l’information scientifique à la technique .....	84
<b>LES SCIENCES AGRONOMIQUES EN PERSPECTIVE : HISTOIRE ET EPISTEMOLOGIE EN ENVIRONNEMENT NUMERIQUE .....</b>	<b>95</b>
Faire archive : intentionnalité et trace de la donnée patrimoniale scientifique ...	95
Ecrire l’histoire de la recherche agronomique en environnement numérique ....	114
<b>CONCLUSION .....</b>	<b>123</b>
<b>ANNEXES.....</b>	<b>125</b>
<b>SOURCES.....</b>	<b>128</b>
<b>BIBLIOGRAPHIE.....</b>	<b>131</b>
<b>TABLES DES ILLUSTRATIONS .....</b>	<b>137</b>
<b>TABLE DES MATIERES.....</b>	<b>138</b>

## *Sigles et abréviations*

AAF :	Association des archivistes français
ADBU :	Association des Directeurs et des personnels de direction des Bibliothèques Universitaires et de la documentation
ACDC :	Accès, capitalisation et diffusion des connaissances
AERES :	Agence d'évaluation de la recherche et de l'enseignement supérieur
ANR :	Agence nationale de la recherche
BnF :	Bibliothèque nationale de France
BSN :	Bibliothèque scientifique nationale
CCSD :	Centre pour la communication scientifique directe
CINES :	Centre Informatique National de l'Enseignement Supérieur
CIRAD :	Centre de coopération international en recherche agronomique pour le développement
CNRS :	Centre national pour la recherche scientifique
CoSO :	Comité pour la Science Ouverte
DipSO :	Direction pour la science ouverte
DOI :	Digital Object Identifier
EOSC :	<i>European Open Science Cloud</i>
ESR :	Enseignement supérieur et recherche
GED :	Gestion électronique de document
INIST :	Institut National de l'Information Scientifique et Technique
Inra :	Institut national de la recherche agronomique ( <i>cf.</i> INRAE)
INRAE :	Institut national de recherche pour l'agriculture, l'alimentation et l'environnement
Inria :	Institut national de recherche en informatique et en automatique
IST :	Information scientifique et technique
Irstea :	Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture
Num4Sci:	Numérique pour la Science
OAI-PMH :	Open Archives Initiatives Protocol for Metadata Harvesting
OAIS:	Open Archival Information System
OCDE :	Organisation de coopération et de développement économiques
PGD :	Plan de gestion des données

- RDA : *Research Data Alliance*
- RGPD : Règlement général pour la protection des données
- SAE : Système d'archivage électronique
- SIAF : Service Interministériel des Archives de France
- URFIST : Unités régionales de formation à l'information scientifique et technique

# INTRODUCTION

---

« *La science est une entreprise essentiellement anarchiste* »<sup>1</sup> écrivait en 1979 le philosophe des sciences Paul Feyerabend. Non au sens trivial du terme, c'est-à-dire désordonné, tumultueux, chaotique, mais bien politique car « *l'anarchisme s'oppose à l'ordre existant, il s'efforce de détruire cet ordre ou de lui échapper* »<sup>2</sup>. Dans l'acte même de produire méthodologiquement des savoirs sur le réel réside donc une part d'idéal de subversion qui, avec une acuité renouvelée dans les moments de crise, permet de faire émerger des concepts, des pratiques et des outils tentant de rompre avec les usages, sinon de les transformer.

Le mouvement d'ouverture qui traverse depuis les années 1990 le monde de la recherche, associé au développement du numérique qui se fait technologie intellectuelle<sup>3</sup>, marque ainsi peut-être un de ces moments de rupture avec l'« *ordre existant* », qu'il soit scientifique, politique, économique ou social, et dont les limites sont depuis longtemps dépassées. Du « *publish or perish* » au concept de « progrès », des problématiques scientifiques, éthiques et épistémologiques aux enjeux politiques et démocratiques sur fonds de crises multiples, les sciences se retrouvent confrontées, dans un élan réflexif inédit depuis la fin du XX<sup>e</sup> siècle, « *à leurs propres produits, à leurs propres insuffisances* »<sup>4</sup>, aux problèmes dont elles sont elles-mêmes la source. Elles sont dès lors enjointes à se renouveler dans leurs fondements pour répondre simultanément aux nouveaux défis de recherche qui se posent à elles et aux attentes de la société.

Ce changement de paradigme est d'autant plus prégnant pour la recherche agronomique que ses propres objets d'étude ne peuvent plus être sereinement pensés indépendamment des problématiques environnementales, sanitaires et alimentaires auxquelles elle est confrontée et dont elle est en partie responsable. Ce sont des objets hybrides de nature et de culture, des animaux d'élevage sélectionnés, des plantes génétiquement modifiés, qui échappent à l'objectivation usuelle du réel par les sciences occidentales<sup>5</sup>, qui sortent du laboratoire et du champ par les sollicitations répétées de nouveaux acteurs extérieurs aux communautés scientifiques. Au cours des années 1990, la crise de la « vache folle » témoigne de cette sortie en trombe des objets confinés et expose certaines des limites intrinsèques des sciences modernes. L'Institut national pour la recherche agronomique (Inra), instigateur de la modernité agricole depuis la fin de la Seconde Guerre mondiale, est alors confronté à la remise en cause de la « *modernité alimentaire* », parallèlement à la proclamation de la fin du monde moderne, voire de sa négation<sup>6</sup>. Les sciences

---

<sup>1</sup> Paul Feyerabend, *Contre la méthode*, Éditions du Seuil, 1979, p. 13.

<sup>2</sup> Paul Feyerabend, « Thèses sur l'anarchisme épistémologique », *Alliage*, n°28, 1996.

<sup>3</sup> Pascal Robert, *Mnémotechnologies. Une théorie générale critique des technologies intellectuelles*, 2010, Lavoisier, 400 p.

<sup>4</sup> Ulrich Beck, *La société du risque. Sur la voie d'une autre modernité*, Paris, Editions Flammarion, 2008, [1986], p. 341.

<sup>5</sup> Philippe Descola, *Par-delà nature et culture*, Paris, Editions Gallimard, 2005, 792 p.

<sup>6</sup> Bruno Latour, *Nous n'avons jamais été modernes : essai d'anthropologie symétrique*, Paris, La Découverte, 2013 [1991], 210 p.

affrontent effectivement une crise de la scientificité : les frontières idéalement étanches entre nature et culture, dans la tradition des Lumières, perdent en consistance face à ces objets indécis et le champ scientifique ne parvient plus à garder son jardin secret. Il se voit de fait vigoureusement interrogé sur les enjeux éthiques de ses démarches et de ses pratiques, de son rapport à la société, à l'environnement et au vivant. Dans un monde défini par les risques et devant apprendre à vivre au rythme de l'Anthropocène, les chercheurs de cette modernité réflexive doivent donc repenser les modèles de communication et de transmission des savoirs *contre* l'ordre existant, ce que numérique et ouverture donnent l'opportunité de faire.

Ceux-ci sont étroitement liés : les deux sont vécus comme des changements décisifs par les communautés scientifiques<sup>1</sup>. Née au cours des années 1990 sous la forme du libre accès des publications, sous l'impulsion notamment des physiciens par l'utilisation d'entrepôts comme ArXiv, l'ouverture s'envisageait initialement par « *la gratuité, la rapidité de publication et une certaine visibilité des résultats de la recherche* »<sup>2</sup>. Les outils numériques nourrissaient alors les ambitions de partage de l'information entre les communautés scientifiques concernées, d'autant que le web offrait de la visibilité et devenait « *un vecteur d'information et de communication* »<sup>3</sup>. Toutefois, au début des années 2000, c'est un partage encore trop peu tourné vers la société civile et les citoyens alors que certains acteurs politiques pointent « *l'urgence de sortir d'un modèle communicationnel qui consiste à penser que les préoccupations et critiques du grand public face à la science [...] pourraient disparaître grâce à la seule "meilleure information scientifique"* »<sup>4</sup>. Le champ politique commence ainsi à s'emparer du sujet qui quitte les laboratoires pour essaimer ici et là dans la société. La philosophe des sciences Isabelle Stengers le soulignait en 2010, « *la production de connaissances est aujourd'hui considérée comme un enjeu trop important pour laisser ce minimum d'autonomie aux chercheurs* »<sup>5</sup>. *Open access, open science, open data, open government* et bien d'autres modèles d'ouverture sont autant de dynamiques marquant une transformation des rapports au savoir, et donc à la donnée.

Au centre de l'ouverture de la science se pose en effet la question des données, de leur création, de leur gestion, de leur partage ainsi que de leur pérennisation, et ce à partir d'outils numériques encore trop souvent impensés comme environnement totalisant<sup>6</sup>. La gestion des données est un enjeu majeur de l'économie et de la politique du savoir qui, depuis le début du XXI<sup>e</sup> siècle, concentrent les attentions

---

<sup>1</sup> Hélène Bosc, « Archives ouvertes : quinze ans d'histoire », C. Aubry, J. Janik, *Les Archives Ouvertes : enjeux et pratiques. Guide à l'usage des professionnels de l'information*, Paris, ADBS, 2005, p. 27-54.

<sup>2</sup> Alain Jacquesson, Jean-Philippe Schmitt, « Les grands éditeurs face au mouvement open access », Joachim Schöpfel, *La publication scientifique. Analyses et perspectives*, Paris, Lavoisier, 2008, p. 125.

<sup>3</sup> Valérie Schafer, « Renouveau de la science et science du renouveau à l'heure numérique », Valérie Schafer (dir.), *Information et communication scientifiques à l'heure du numérique*, Paris, CNRS Editions, 2014, p. 15.

<sup>4</sup> *Ibid.*, p. 15.

<sup>5</sup> Isabelle Stengers, *Au temps des catastrophes. Résister à la barbarie qui vient*, Paris, La Découverte, 2013, p. □□□.

<sup>6</sup> Bruno Bachimont, *Patrimoine et numérique : technique et politique de la mémoire*, Paris, INA, 2017, 246 p.

des secteurs publics et privés. Le numérique produit en effet une relation inédite à la donnée qui se fait *big data* ou, dans sa version eschatologique, *data deluge*. L'inflation est d'ailleurs telle « *que les moyens techniques de traitement traditionnels ne sont plus suffisants et nécessitent une réflexion sur la capture, le stockage, l'analyse et la visualisation des données* »<sup>1</sup>. C'est une nouvelle ère qui s'ouvre aux sciences, une ère des données massives dont leur origine pluridisciplinaire résulte d'un repositionnement de la recherche scientifique, notamment des sciences agronomiques, sur des approches systémiques niant au réductionnisme la légitimité seule de son discours sur la nature. C'est une ère marquée par le numérique qui est « *à la fois une technologie intellectuelle et une ingénierie pour les systèmes physiques* »<sup>2</sup>, et dont on ne peut penser l'usage « *par et dans le monde scientifique* » sans aborder, selon l'historienne Valérie Schafer, « *des thématiques plus larges, qui touchent aux fondements mêmes de la science, au statut des chercheurs et experts, aux disciplines, à la production, diffusion, médiation des connaissances, aux relations sciences/société* »<sup>3</sup>.

La nécessité de cette refondation plurielle de la recherche scientifique a été finalement très bien intégrée par l'Inra au cours des deux dernières décennies. Si durant les années 2000 furent engagées de premières initiatives ciblées, notamment le déploiement progressif d'une archive ouverte locale, c'est surtout ces dernières années qu'ouverture et numérique ont été conjointement pensés et inscrits dans les orientations politiques et scientifiques de l'Institut. Ainsi, dans le document d'orientation « Inra2025 » publié en 2016, la première grande orientation de politique générale définit « *une science ouverte grâce au numérique* »<sup>4</sup>, suivi en second axe d'un « *OpenInra* » engagé dans les territoires. Cinq années plus tard, un nouveau document est produit, à la suite de la fusion de l'Inra avec l'Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (Irstea). La science ouverte est alors pleinement intériorisée dans les orientations politiques du nouvel Institut qui, en cinquième et dernière axe scientifique aux côtés des enjeux environnementaux, agroécologiques, sanitaires et alimentaires, s'engage à « *mobiliser la science des données et les technologies du numérique au service des transitions* »<sup>5</sup>.

Toutefois, cette intériorisation aura tardé tandis que dès 1993, la direction de l'Inra est sollicitée pour définir, « *en liaison avec l'administration des Archives nationales, une politique de collecte et de conservation de ses archives*

---

<sup>1</sup> Clément Mabi, Jean-Christophe Plantin, Laurence Monnoyer-Smith, « Interroger les données en SHS à partir de leur écosystème », Valérie Schafer (dir.), *Information et communication scientifiques à l'heure du numérique*, Paris, CNRS Editions, 2014, p. 63.

<sup>2</sup> Bruno Bachimont, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques », *Interfaces numériques*, volume 4, n° 3, 2015, [En ligne], DOI : 10.25965/interfaces-numeriques.386.

<sup>3</sup> Valérie Schafer, « Renouveau de la science et science du renouveau à l'heure numérique », Valérie Schafer (dir.), *Information et communication scientifiques à l'heure du numérique*, Paris, CNRS Editions, 2014, p. 10.

<sup>4</sup> Odile Vilotte, François Houllier, Philippe Mauguin, Claude Ronceray, Christine Cherbut, *et al.*, « Document d'Orientation #Inra2025 », Inra, 2016, p.4.

<sup>5</sup> INRAE, « INRAE2030. Partageons la science et l'innovation pour un avenir durable », document d'orientation, décembre 2020, p. 32.

*administratives et scientifiques* »<sup>1</sup>. En effet, est élaborée à ce moment-là une réflexion poussée non seulement sur la gestion pérenne des documents d'activité administratifs et des archives des chercheurs, mais aussi sur les données de la recherche elles-mêmes dont Denis Poupardin, économiste de formation à l'Inra, interroge leur préservation<sup>2</sup>. Son appel restera cependant vain, conduisant à des « *lacunes archivistiques* » conséquentes selon l'historien Pierre Cornu, et notamment « *la perte de fonds archivistiques importants, le retard des versements aux Archives nationales ou départementales, et enfin l'extrême dispersion et l'absence d'inventaire des archives encore conservées par l'Inra* »<sup>3</sup>. Le management des documents d'activité, tant nécessaire à la réalisation des missions quotidiennes de l'Institut qu'à des fins réglementaires et patrimoniales, ne se réalise péniblement qu'à partir de la fin des années 2000. Il est finalement étroitement associé, au cours de ces dernières années, au déploiement de la gestion raisonnée des données de la recherche.

Ce déficit actuel de « traces » de l'histoire de l'Institut interroge les historiens des sciences qui font pourtant face à un organisme de recherche dont le poids de la mémoire est monumental, dans tous les sens du terme. En 2018, Philippe Mauguin, directeur de l'Institut, soulignait en préface de *L'histoire de l'Inra, entre science et politique* la force singulière de la mémoire collective dans l'Institut. C'est, écrit-il, « *une mémoire cristallisée dans de grandes réussites scientifiques, dans la conquête de nouveaux fronts de science, dans des crises affrontées et dépassées collectivement, jusqu'aux responsabilités qui sont les nôtres aujourd'hui dans un monde où les enjeux agricoles, alimentaires et environnementaux s'interpénètrent aussi bien à l'échelle des territoires de vie de nos concitoyens qu'à celle de la gouvernance planétaire.* »<sup>4</sup> La force d'une mémoire travaillée par le monde dans lequel elle se déploie et auquel elle se confronte et répond ; une mémoire travaillée par l'histoire.

Etudier la gestion des données de la recherche et des archives des sciences de du nouvel Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) né en 2020, c'est ainsi s'intéresser à l'avènement de nouvelles pratiques de recherche devant nécessairement composer avec des contraintes exogènes inévitables : comment le premier institut de recherche agronomique en Europe répond-il aux nouveaux enjeux scientifiques, politiques, économiques et sociétaux de cette modernité réflexive ? Surtout par quelles pratiques, quels acteurs, quelles transactions, et finalement à quelles fins ? Le rôle des mouvements d'ouverture et des outils numériques, décisifs à de multiples échelles (chercheur, laboratoire, bibliothèque, administration, Etat, *etc.*) en contexte

---

<sup>1</sup> Denis Poupardin, Communication, Colloque «Archorales» du 2 décembre 2008, Inra, [https://www6.inrae.fr/comitedehistoire/content/download/3002/30151/version/1/file/Expose\\_Denis\\_Poupardin.pdf](https://www6.inrae.fr/comitedehistoire/content/download/3002/30151/version/1/file/Expose_Denis_Poupardin.pdf).

<sup>2</sup> Mireille Jean, Denis Poupardin, « Les archives des unités de recherche : le point de vue des scientifiques de l'Institut national de la recherche agronomique (I.N.R.A.) », *La Gazette des archives*, n°176, 1997, p. 28 à 49.

<sup>3</sup> Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra, entre science et politique*, Versailles, Editions Quae, 2018, p. 12.

<sup>4</sup> *Ibid.*, p. 8.

de crises multiples, interroge le renouvellement d'une épistémologie des sciences agronomiques par le numérique à partir duquel penser et transmettre les savoirs<sup>1</sup>. Une transmission à la fois intracommunautaire et intertemporelle qui pose en outre la question de la pérennisation des données et documents produits par le numérique, qui se fait aujourd'hui plus que jamais équipement technique de la mémoire sociale : quelle ambition d'un archivage numérique de ces données et documents dont leur pérennisation questionne le « faire archive » ?

Cette étude, en termes bibliographiques et méthodologiques, s'inscrit ainsi dans le sillon de différents champs de recherche, des sciences de l'information et de la communication (SIC) à l'histoire des sciences, qui ont eu à cœur d'étudier numérique et sciences, sciences et société, société et mémoire ou encore mémoire et histoire. Il a ainsi semblé nécessaire, pour comprendre la complexité de la gestion des données de la recherche, de saisir à la fois ses mécanismes info-communicationnels, mais aussi son cadre historique et sociologique si singulier.

La problématique de la gestion des données de la recherche est une thématique récente et riche des travaux des SIC. Elle s'articule avec de nombreux autres sujets, particulièrement celui de l'*open science* et de l'*open access* afin notamment de caractériser les modalités de diffusion du savoir<sup>2</sup>. Ainsi, le chercheur en SIC Joachim Schöpfel a fait de l'ouverture des données de la recherche un de ses axes de recherche, afin d'interroger ces enjeux mêlant sciences et société par le truchement de la donnée<sup>3</sup>. En outre, les points d'intérêt des mouvements d'ouverture interpellent l'environnement de la publication scientifique en forte tension économique, mais aussi politique et, évidemment, scientifique<sup>4</sup>.

Par ailleurs, le cadre de la gestion des données de la recherche, dans notre perspective d'interaction avec l'archivage numérique, modifie le champ de l'étude des archives des sciences. Les travaux se multiplient à partir des années 2000 et dressent tous le constat d'un retard manifeste dans l'organisation d'une collecte systématique et l'avènement d'une politique globale d'archivage des documents des organismes de recherche valide à long terme<sup>5</sup>. Thérèse Charmasson, conservateur du patrimoine, s'est à plusieurs reprises penchée sur la question partageant notamment son désarroi dans cet intérêt relatif pour les archives des sciences.

Enfin, le numérique se fait l'environnement de toutes ces problématiques en déterminant les modalités de leur compréhension et de leur appréhension. Dans un axe d'étude de l'archivage des données de la recherche à des fins patrimoniales,

---

<sup>1</sup> Bernard Jacquemin, Joachim Schöpfel, Renaud Fabre, « Libre accès et données de recherche. De l'utopie à l'idéal réaliste », *Études de communication*, 52 | 2019, p. 14.

<sup>2</sup> Chérifa Boukacem-Zeghmouri, Hans Dillaerts, « Information scientifique et diffusion des savoirs : entre fragmentations et intermédiaires », *Revue française des sciences de l'information et de la communication*, n° 15, 2018, <http://journals.openedition.org/rfsic/5522>.

<sup>3</sup> Bernard Jacquemin, Joachim Schöpfel, Renaud Fabre, « Libre accès et données de recherche. De l'utopie à l'idéal réaliste », *Études de communication*, 52 | 2019, p. 11-26.

<sup>4</sup> Joachim Schöpfel (dir.), *La publication scientifique : analyses et perspectives*, Paris, Lavoisier : Hermes science publications, 2008, 367 p.

<sup>5</sup> Thérèse Charmasson, « Archives scientifiques ou archives des sciences : des sources pour l'histoire », *La revue pour l'histoire du CNRS*, n° 14, CNRS Éditions, 2006, DOI: 10.4000/histoire-cnrs.1790.

l'étude des relations entre mémoire et numérique révèle toutes les aspérités de leur articulation complexe et souvent impensée<sup>1</sup>. Bruno Bachimont, chercheur en SIC, ouvre depuis plusieurs années déjà le front de ces questionnements à la valeur heuristique certaine<sup>2</sup>.

Une fois le champ des SIC traversé, nous devons ensuite nous tourner vers l'histoire, la sociologie et la philosophie des sciences pour comprendre la gestion des données de la recherche agronomique dans toute sa complétude. Particulièrement prolifiques depuis les années 1990, les travaux dans la continuité desquels nous souhaitons nous inscrire cherchent à se dégager des approches strictement linéaires et dépolitisées de la production des connaissances pour restituer les sciences occidentales dans la profondeur du temps de l'histoire et le relief de la société. Le sociologue et philosophe Bruno Latour étudie depuis de nombreuses années cette *science en train de se faire*<sup>3</sup> qui travaille, dans la tradition des Lumières, de nombreux dualismes, tels sujet objet, nature culture, bien connus en anthropologie<sup>4</sup>.

Ces considérations permettent alors d'écrire une histoire de l'Institut dans toute sa complexité en délaissant, pour ce faire, l'illusion d'une rationalisation inéluctable du monde moderne au profit de la réintroduction de l'erreur, de l'égarement, du faux dans l'histoire des sciences. L'historien Pierre Cornu déploie cette approche, lui qui coécrit en 2018 une nouvelle histoire de l'Inra<sup>5</sup> et interroge frontalement le rapport de ses acteurs à leurs propres traces. En effet, cet environnement historique et sociologique complexe dépasse les seules considérations intellectuelles pour travailler l'entière des activités d'un organisme de recherche comme INRAE confronté à des enjeux scientifiques et éthiques forts. L'intelligibilité à l'échelle globale des traces et documents produits par *la science en action* dépend d'une gestion des données de la recherche raisonnée dont l'étude que nous en proposons doit ainsi permettre d'en restituer les conditions d'épanouissement.

Cette étude se propose, dans son plan, de retracer le cycle de vie des données de la recherche agronomique, de leurs nouvelles conditions d'avènement que sont les mouvements d'ouverture, et notamment l'open science, à leur utilisation comme archives pour nourrir une histoire des sciences agronomiques modernes. Le premier temps de ce travail sera ainsi consacré à la définition de la « donnée de la recherche » et de la science ouverte, que ce soit dans sa dynamique globale ou au sein de la recherche agronomique. Ces mises en perspective permettront dans un second temps d'étudier la gestion effective des données de la recherche et des documents d'activité au sein d'INRAE porté par des ambitions et des acteurs pluriels. Enfin, venant clore

---

<sup>1</sup> Pascal Robert, Mnémotechnologies. *Une théorie générale critique des technologies intellectuelles*, op. cit., p. 310.

<sup>2</sup> Bruno Bachimont, *Patrimoine et numérique : technique et politique de la mémoire*, Paris, INA, 2017, 246 p.

<sup>3</sup> Bruno Latour, *La science en action : introduction à la sociologie des sciences*, Paris, Gallimard, 1995, 663 p.

<sup>4</sup> Philippe Descola, *Par-delà nature et culture*, Paris, Gallimard, 2005, 792 p.

<sup>5</sup> Pierre Cornu, Egizio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra, entre science et politique*, Versailles, Editions Quae, 2018, 463 p.

le cycle de vie des données, nous approfondirons les sciences agronomiques mises en perspective par l'étude épistémologique de l'histoire des sciences, dont la matière première qu'est l'archive se voit transformée dans ses conditions d'existence par le numérique.

# LA SCIENCE OUVERTE ET LE LIBRE ACCES : DE NOUVELLES PRATIQUES POUR UNE NOUVELLE SCIENCE

---

## LA FABRIQUE DU SAVOIR

### Les données, matière première de la recherche

Qu'est-ce que la donnée de la recherche ? « *Si personne ne me le demande, je le sais ; si je cherche à l'expliquer à celui qui m'interroge, je ne le sais plus* ». Cette formule de Saint Augustin, originellement au sujet du temps<sup>1</sup>, souligne ici une des particularités du terme de « donnée de la recherche » : si tout le monde en a une compréhension immédiate, il semble difficile d'en cerner théoriquement la notion. Le plus souvent au pluriel, ce qui n'est pas anodin, le terme recouvre en effet une large gamme d'acceptations dont tout acteur de leur gestion – et moins de leur production – s'interroge sur leur caractérisation. L'*Australian National Data Service* explique en 2017 dans son guide de management qu'« *il est difficile de fournir une définition des données de recherche faisant autorité, car toute définition dépend du contexte dans lequel la question est posée.* »<sup>2</sup> La clef de la compréhension des données de la recherche serait donc le contexte ; or, quelle définition commune des données au moment de leur production, aux moments de leur utilisation, ou encore au moment de leur archivage ? Un modèle de définition posé demeure-t-il en outre valable pour toutes les communautés scientifiques ? Sachant que chaque temps et chaque lieu peuvent être marqués par l'implication de nouveaux acteurs, créant de nouvelles masses de données, connaître le contexte n'est qu'un point d'entrée pour éclairer notre lanterne.

### Quelques premières définitions

Cette difficulté de définition caractéristique s'observe dès les années 1990 aux Etats-Unis, dont l'*Office of Management and Budget* du gouvernement fédéral s'essaie à l'exercice, avec la particularité de prendre plus de temps à décrire ce que ne sont *pas* les données de la recherche que ce qu'elles sont : « *Les données de recherche sont définies comme le matériel factuel enregistré et communément accepté dans la communauté scientifique comme étant nécessaire pour valider les résultats de la recherche, à l'exception d'aucun des éléments suivants : analyses préliminaires, ébauches d'articles scientifiques, plans de recherches futures, examens par les pairs ou communications avec des collègues. Ce matériel "enregistré" exclut les objets physiques (par exemple, les échantillons de*

---

<sup>1</sup> Saint Augustin, *Confessions*, L. XI, XIV. 17.2, trad. Péronne et Écalle, Nathan, 1998.

<sup>2</sup> « *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.* », Australian National Data Service, "What is research data", *ANDS Guide*, 2017, p. 1, <https://www.ands.org.au/guides/what-is-research-data>.

laboratoire). Les données de recherche n'incluent pas non plus : [...]. »<sup>1</sup>. S'ensuit alors une nouvelle liste d'objets qui n'entrent pas dans la catégorie du « matériel factuel nécessaire à la validation des résultats », laissant le flou sur ce que cela recouvre. Ce modèle de définition est repris la décennie suivante par l'OCDE pour qui les données de la recherche sont « *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.* »<sup>2</sup>

Commun à ces deux définitions, l'engagement des données de la recherche se réalise sur les enjeux de validation de faits scientifiques au sein d'une communauté établie qui les reconnaît en tant que telles. « Fait », « communauté » ou encore « validation » sont des notions depuis plusieurs décennies travaillées par la sociologie des sciences et rediscutées à l'encontre de leur usage par les scientifiques eux-mêmes. Dès 1929, le médecin et sociologue Ludwik Fleck s'évertuait à appréhender la « genèse d'un fait scientifique » qu'il décrivait comme le résultat d'efforts continus d'un grand nombre d'individus dans une seule et même dynamique<sup>3</sup>. Cette dynamique plus ou moins uniforme rassemble alors une « *communauté de personnes [...] qui interagissent intellectuellement* », « *qui échangent des idées* » et, en l'occurrence, des données de la recherche ; cette communauté forme un « collectif de pensée » qui se fait alors « *le vecteur du développement historique d'un domaine de pensée, d'un état du savoir déterminé et d'un état de la culture, c'est-à-dire d'un style de pensée particulier* »<sup>4</sup>. Son approche sociologique et historique du fait scientifique invite en effet à penser le développement des sciences de manière contextuelle et systématiquement contingente : « *une autre succession des événements historiques, une autre évolution des innovations techniques, ou bien d'autres formes d'organisation sociale auraient certainement produit d'autres "styles de pensée" et d'autres "faits"*. »<sup>5</sup>

Or, si les données de la recherche sont ces « *enregistrement factuels [...] généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats* », alors elles sont les produits d'un style de pensée en propre qui est « *cette force contraignante spécifique s'exerçant sur la pensée* », qui est « *la disposition pour telle manière de voir ou d'appréhender et non pas telle autre* », qui

---

<sup>1</sup> « *Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: Preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples). Research data also do not include : [...].* » Office of Management and Budget, "2 Cfr 215 - Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and other Non-Profit Organizations", 1993, amendée le 30/09/99, p. 96, <https://www.govinfo.gov/content/pkg/CFR-2012-title2-vol1/pdf/CFR-2012-title2-vol1-part215.pdf>.

<sup>2</sup> OCDE, *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, 2007, p. 18.

<sup>3</sup> Ludwik Fleck, *Genèse et développement d'un fait scientifique*, Paris, Les Belles lettres, 2005 [1929], 280 p.

<sup>4</sup> *Ibid.*, p. 74.

<sup>5</sup> Nathalie Jas, « Préface », Ludwik Fleck, *Genèse et développement d'un fait scientifique*, op. cit., p. XXVIII.

est finalement cette « *totalité de ce qui est intellectuellement disponible* »<sup>1</sup> au scientifique. Un style de pensée peut évoluer, notamment face à l'agrégation progressive de connaissances nouvelles qu'il ne reconnaît pas et qu'il tente d'assimiler. Les données de la recherche sont ainsi les fruits de l'activité d'un collectif de pensée qui entretient son champ d'expertise jusqu'aux limites et apories de son style de pensée. La recherche agronomique de la seconde moitié du XX<sup>e</sup> siècle jusqu'à nos jours témoigne de cette rencontre de styles de pensée et de leur évolution, par un rapport de force longtemps très inégal entre biologie moléculaire et systémique agraire, entre approche réductionniste et perspective holiste de la recherche agronomique<sup>2</sup> ; tous deux défendaient leur style de pensée dont dépendait la production des idées et des données, et l'un ne reconnaissait nullement à l'autre la validité de celles-ci. Ce sont finalement les événements historiques, les innovations techniques et les crises successives, notamment environnementales, sanitaires et alimentaires, qui conduisirent à l'émergence d'un nouveau style de pensée proposant une approche totalisante de l'environnement, du vivant et de l'agriculture. Les données de la recherche ne sont donc jamais des absolus, mais toujours associées à un contexte et produites relativement à celui-ci.

### **Typologie des données**

Les données de la recherche n'existent pas à l'état naturel, elles ne sont pas capturées ; elles sont des construits résultant de l'enregistrement d'un événement, comme l'indique les définitions usant de ce terme. En effet, « *la science enregistre pour transporter ce que l'on appelle de l'information (et qui est à proprement parler une mise en forme) dans le temps et l'espace* »<sup>3</sup>. Cette mise en forme se réalise parfois à plusieurs niveaux, conduisant certains à opérer une distinction des données de la recherche par étape, des données « brutes » (*raw data*) aux données « élaborées » (*process data*). Les premières se voudraient « *neutres, objectives et largement indépendantes du contexte, de l'analyse ou de l'observateur* »<sup>4</sup>, quand les secondes renverraient au travail de recherche traitant, analysant, sélectionnant les résultats pour produire et valider les résultats de la recherche, les « faits scientifiques ». Les professionnels de l'information scientifique et technique (IST) de l'Inra procèdent à cette différenciation dans leur définition des données de la recherche<sup>5</sup>. C'est n'est toutefois pas nécessairement la vision des producteurs des données eux-mêmes, comme le révèlent les résultats de certaines enquêtes sociologique : la notion de données brutes, dans leur perspective, « *n'est pas revendiquée en tant que telle en interne par les producteurs et les manipulateurs de*

---

<sup>1</sup> Ludwik Fleck, *Genèse et développement d'un fait scientifique*, op. cit., p. 116.

<sup>2</sup> Pierre Cornu, « La recherche agronomique française dans la crise de la rationalité des années soixante-dix », *Histoire de la recherche contemporaine*, t. III, n° 2, 2014, p. 154 à 166.

<sup>3</sup> Pascal Robert, *Mnémotechnologies. Une théorie générale critique des technologies intellectuelles*, op. cit., p. 21.

<sup>4</sup> Anne Thessen, David Patterson, "Data issues in the life sciences", *ZooKeys*, (150), 2011, <https://doi.org/10.3897/zookeys.150.1766>.

<sup>5</sup> Sylvie Cocaud, Pascal Aventurier, « Les entrepôts de données de recherche », *Action Nationale de Formation*, CNRS, juillet 2017, Vandoeuvre-les-Nancy, France, p. 7, hal-01595599.

*données* »<sup>1</sup>. Toutefois, cette distinction transparait parfois dans leurs pratiques par le prisme du régime de véridicité : c'est ce que les résultats du projet de recherche *Ecrito* sur les documents scientifiques informels d'un laboratoire de mécanique des fluides mettent en relief<sup>2</sup>. Les données brutes produites sont en effet soumises à une première opération de traitement, fruit « *d'une méthodologie propre à la discipline* », distinguant dès lors les « *vraies données* » des « *fausses données* »<sup>3</sup>. Or le rapport au vrai et au faux est toujours une affaire délicate au sein des sciences moderne. En outre, en contexte d'ouverture (*open data* notamment) les données ont déjà « *vécues* » ; elles sont dépendantes de réseaux sociotechniques formateurs, d'un contexte de production *situé*. L'enjeu est alors de les détacher, de les « *dé-spécifier de leurs usages initiaux pour les préparer à un vaste horizon d'usages possibles* »<sup>4</sup>.

Ce retravail des données par les chercheurs n'est toutefois pas une découverte. Le sociologue des sciences Bruno Latour avait déjà mis au jour la manière dont la production et l'enregistrement de l'information au sein des laboratoires se réalisaient par la médiation de la technique, d'« instruments » dont les résultats finaux sauvegardés, et parfois à termes publiés, étaient le fruit « *d'un long processus de laboratoire* » ; la donnée de la recherche est « *extraite des instruments de cette salle, nettoyée, redessinée et mise en forme* »<sup>5</sup>. Sans revenir ici sur le débat interrogeant qui de la théorie ou de la donnée précède l'autre, la sociologie des sciences a ainsi su montrer depuis les années 1980 qu'il n'existe pas de telles choses que des données « neutres, objectives et indépendantes du contexte » dans lequel elles sont produites. Le rôle de l'instrumentation, particulièrement dans les sciences de la nature et du vivant, est en effet fondamental, elle qui permet l'interprétation du réel par le scientifique et dont les représentations qui en sont extraites se font « *vues de l'esprit* »<sup>6</sup>, à la dimension changeante non-négligeable – mais difficilement avouable et souvent impensée – précédant le processus de fixation des données au travers notamment de leur publication.

Or les définitions du gouvernement fédéral des Etats-Unis et de l'OCDE relèguent en second plan la dimension technique de l'enregistrement des événements et des faits scientifiques. La Commission européenne, cependant, fait de la donnée de la recherche une « *donnée de la recherche numérique* » dans son programme Horizon 2020. « *Les “données de recherche numériques” sont des informations sous forme numérique (en particulier des faits ou des chiffres), collectées pour être examinées et utilisées comme base de raisonnement, de discussion ou de calcul ; ceci comprend les statistiques, les résultats d'expériences, les mesures, les observations*

---

<sup>1</sup> Jérôme Denis, Samuel Goëta, « La fabrique des données brutes : Le travail en coulisses de l'open data », Clément Mabi, Laurence Monnoyer Smith (dir.), *Ouvrir, partager, réutiliser : Regards critiques sur les données numériques*, Paris, Éditions de la Maison des sciences de l'homme, 2017, p. 16, <https://doi.org/10.4000/books.editionsmslh.9050>.

<sup>2</sup> Muriel Lefebvre (coord.), *Les documents scientifiques informels : un patrimoine peu exploré, témoin de la construction des savoirs*, Rapport de recherche, Université de Toulouse, 2012, 56 p., [https://ecrito.hypotheses.org/Rapport\\_Final.pdf](https://ecrito.hypotheses.org/Rapport_Final.pdf).

<sup>3</sup> Urfistinfo, « “Données” de la recherche, les mal-nommées », <https://urfistinfo.hypotheses.org/2581>.

<sup>4</sup> *Ibid.*, p. 16.

<sup>5</sup> Bruno Latour, *La science en action. Introduction à la sociologie des sciences*, op. cit., p. 156.

<sup>6</sup> Bruno Latour, « Les “vues” de l'esprit », *Réseaux. Communication – Technologie – Société*, n° 27, 1987, p. 79 à 96.

*résultant de travaux de terrain, les résultats de l'enquête, les enregistrements d'interviews et images.* »<sup>1</sup> Si cette définition s'inscrit dans le cadre de la présentation des ambitions de la science ouverte à l'échelle européenne, elle n'en demeure pas moins lier étroitement données de la recherche et numérique qui se fait de plus en plus technologie intellectuelle première de la construction des savoirs scientifiques. Le chercheur en sciences de l'information et de la communication Bruno Bachimont invite ainsi à penser le numérique moins comme un outil que comme un environnement au sein duquel « *envisager des problèmes inédits, poser des problèmes qui n'existeraient pas dans un contexte où cette puissance technique serait absente* »<sup>2</sup>.

Lorsque la notion de donnée est abordée, sa définition peut restituer la plupart du temps l'impression d'une « *existence monolithique* »<sup>3</sup> ; or dans cet environnement totalisant, les données sont multiples et plurielles. Il importe donc d'étudier leur réalité pour les comprendre, et *in fine* leur appliquer les procédures de traitement adéquates. Lyne Da Sylva, en ce sens, distingue quatre catégories de données : les mégadonnées, les données de la recherche, les données ouvertes et les données liées<sup>4</sup>. Chacune de ces catégories peut recouvrir partiellement une autre : certaines mégadonnées peuvent être des données de la recherche, mais toutes les données de la recherche ne sont pas des mégadonnées (*big data*) ; les données de la recherche peuvent se faire données ouvertes, mais toutes les données ouvertes ne le sont pas, particulièrement dans la dynamique d'ouverture qui touche largement le secteur public et l'administration. Bruno Bachimont, quant à lui, propose une approche par la science des données, prenant pour objet d'étude celles-ci et particulièrement les mégadonnées qui impose un changement de paradigme. Avec l'avènement des mégadonnées, les chercheurs semblent en effet être passés d'une épistémologie de la mesure à une épistémologie de la donnée ; et les deux ne se confondent pas. Les premières sont la construction raisonnée d'unités fondées sur le réel à partir duquel, et en finalité duquel, sont élaborées les lois universelles qui le décrivent. Les données, elles, changent le rapport aux lois et à l'universalité : ce sont « *des informations, c'est-à-dire des expressions formatées selon une structure formelle ou symbolique de manière arbitraire, et collectées grâce à leur format* »<sup>5</sup>. Le caractère arbitraire n'est pas anodin, ni le formatage, nous le verrons par la suite. Mais ils permettent dès maintenant de saisir davantage les caractéristiques des

---

<sup>1</sup> « *Digital research data' are information in digital form (in particular facts or numbers), collected to be examined and used as a basis for reasoning, discussion or calculation; this includes statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.* » Commission européenne, *Horizon 2020*, H2020 Program, Version 5.2., Juin 2019, [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/amga/h2020-amga\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf), p. 252.

<sup>2</sup> Bruno Bachimont, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques. Principes pour une science des données », *Interfaces numériques*, Volume 4 – n° 3/2015, 10.25965/interfaces-numeriques.386.

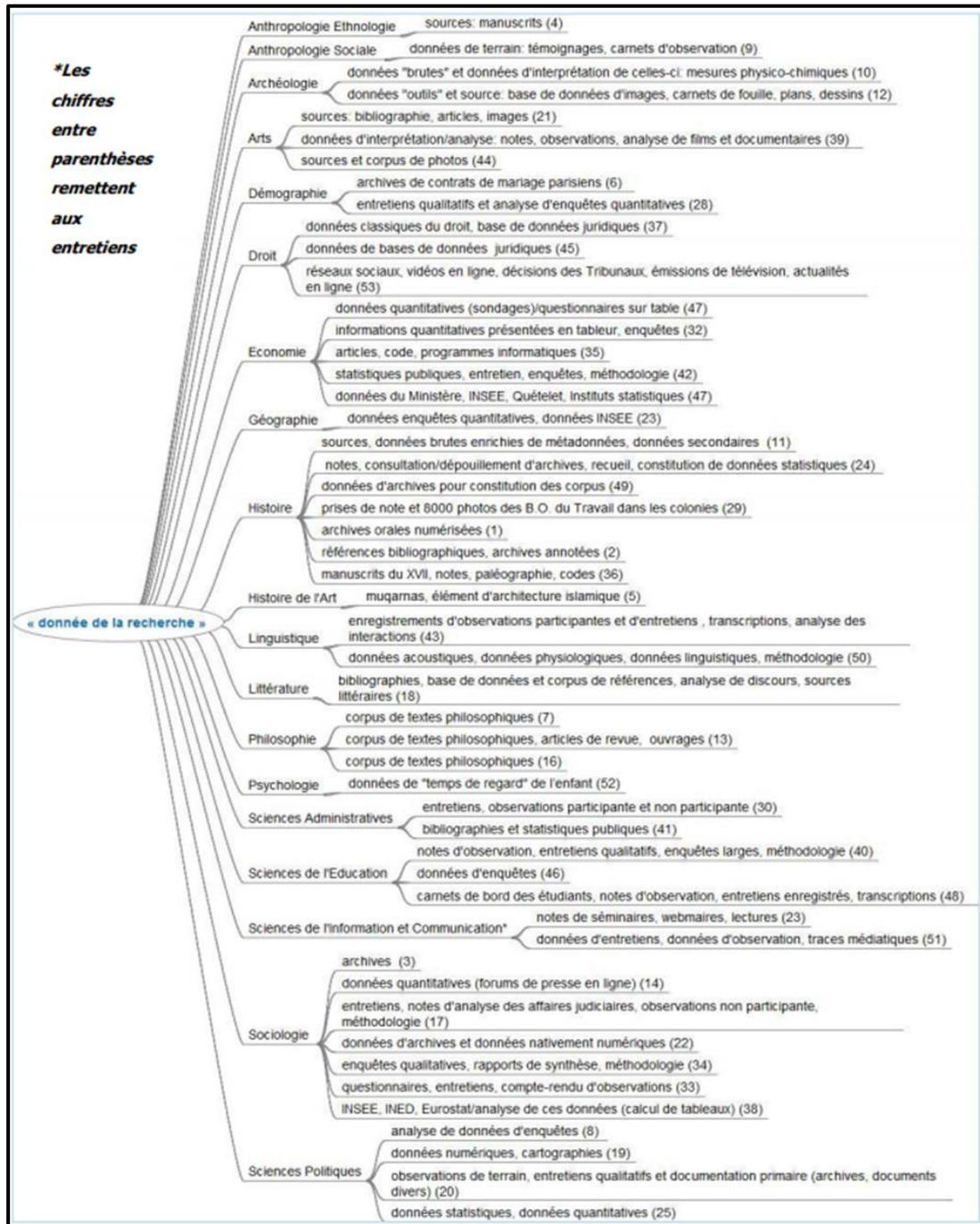
<sup>3</sup> Lyne Da Sylva, « Les données et leurs impacts théoriques et pratique sur les professionnels de l'information », *Documentation et bibliothèque*, VI. 63, n° 4, octobre-décembre 2017, <https://doi.org/10.7202/1042308ar>.

<sup>4</sup> *Ibid.*

<sup>5</sup> *Ibid.*, 2.1.

données de la recherche qui, particulièrement pour les sciences de la nature et du vivant, sont presque exclusivement enregistrées en environnement numérique.

Les sciences humaines et sociales, elles, n'échappent pas à l'environnement numérique totalisant. Elles peuvent toutefois développer un autre rapport à la donnée de la recherche, ce que Francisca Cabrera, dans son mémoire sur les données de la recherche en SHS<sup>1</sup>, souligne après avoir rencontré une cinquantaine de chercheurs et réalisé l'arborescence suivante.



Arborescence des données de la recherche en SHS – Francisca Cabrera, 2014.

<sup>1</sup> Francisca Cabrera, *Les données de la recherche en Sciences humaines et sociales : enjeux et pratiques. Enquête exploratoire*, Mémoire de fin d'étude, Conservatoire national des arts et métiers, 2014, 238 p.

Il apparaît ici d'autant plus ardu de donner une définition stricte des données de la recherche que les collectifs de pensée de chaque discipline vont tendre à caractériser à leur manière les matériaux de recherche qu'elles manipulent. A partir de cette visualisation, et afin de définir les caractéristiques des données de recherche propre à chaque champ, Francisca Cabrera distingue les disciplines à approche herméneutique et textuelle des sciences à approche expérimentale et de terrain. Les premières, telles l'histoire et la philosophie, travaillent des sources textuelles, photos, vidéo, des corpus bibliographiques, juridiques, des articles et ouvrages scientifiques, surtout un ensemble moins souvent pensé comme données de la recherche que comme « *données matériaux* » indissociables « *de la construction du texte* » et de l'élaboration « *aboutie du discours à travers la publication* »<sup>1</sup>. Les secondes, telles l'anthropologie, l'archéologie, l'économie, la géographie ou encore la sociologie ont une approche expérimentale et de terrain qui restitue la donnée dans sa forme « *d'informations brutes collectées suite à une problématique de départ* »<sup>2</sup>. Ces dernières se rapprochent sensiblement de la conception de la « donnée de la recherche » des sciences de la nature et du vivant, sinon par le prisme de l'instrumentation, du moins par leurs approches itératives, expérimentales, ou encore observationnelles.

C'est d'ailleurs une des manières de définir les données de la recherche que celle de les appréhender par leur méthodologie de collecte, avec toujours en perspective ce mouvement de traduction d'un événement en donnée par enregistrement. Violaine Rebouillat, au sein de sa thèse de doctorat en sciences de l'information et de la communication<sup>3</sup>, opère ce rapprochement en soulignant les travaux de la *National Science Foundation*<sup>4</sup>. Celle-ci distingue les données de la recherche issues de l'observation (*observational data*) qui sont idéalement conservées indéfiniment, celles produites par une approche computationnelle (*computational data*) dans le cadre de l'élaboration de modèle ou de simulation numérique, et dont la préservation doit se faire en relation avec les dispositifs qui ont permis leur production, enfin celles obtenues lors d'expérimentation (*experimental data*) et dont la sauvegarde doit impérativement être pensée en termes de coût et de condition de reproduction des résultats. S'il est systématiquement précisé le sort final de ces données, c'est que la *National Science Foundation* a fait le choix d'associer étroitement la méthodologie de collecte à la stratégie de préservation des données de la recherche dans l'identification de leur cycle de vie.

Les archivistes, dans cette perspective, n'ont pas tardé à se positionner sur le sujet de la gestion des données de la recherche. C'est d'ailleurs un point d'entrée parmi d'autres de leur définition : celui de leur pérennisation. Dès 1993 dans les murs de l'Inra, l'économiste Denis Poupardin invite à s'intéresser à la pérennisation

---

<sup>1</sup> *Ibid.*, p. 54.

<sup>2</sup> *Ibid.*, p. 55.

<sup>3</sup> Violaine Rebouillat, *Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs*, thèse de doctorat en sciences de l'information et de la communication, Conservatoire national des arts et métiers, 2019, p. 30.

<sup>4</sup> National Science Foundation, *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*, National Science Board, Septembre 2005, p. 19, <https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>.

de l'ensemble des documents et données produits par l'activité scientifique. Graduellement, il distingue les « *collections de matériel biologique vivant ou mort* », les « *documents intermédiaires* », ici comprenant de manière différenciée les cahiers de laboratoires, les répertoires de techniques et de protocoles expérimentaux, les recueils de « *données brutes* », ainsi que les photographies et films ; enfin les « *documents achevés* », soit la littérature scientifique, la littérature grise, les thèses et mémoires, les communications au congrès ou encore les posters, les rapports et les études.

Les première et dernière catégories semblent *de facto* exclues des définitions des données de la recherche précédemment parcourues. En outre, au sein des documents intermédiaires, les cahiers de laboratoire et répertoires de techniques sont des objets explicitement refusés par les définitions<sup>1</sup>. Demeurent alors les « données brutes », les photographies et vidéos qui seraient les seules aujourd'hui pensées comme données de la recherche. Cette approche archivistique déployée à l'Inra au début des années 1990 a pour mérite de maintenir l'activité scientifique dans une certaine cohérence, associant les données brutes aux répertoires de techniques et aux cahiers de laboratoire, liant donc informations collectées et intelligibilité de leur mise en pratique dans un processus de construction d'un fait scientifique. C'est une perspective que l'on peut retrouver sans surprise sous la plume d'un archiviste écrivant pour la section Aurore de l'Association des archivistes français (AAF) rassemblant les archivistes des universités, rectorats, organismes de recherche et mouvements étudiants. Les données de la recherche sont ainsi pour eux « *l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. A ce titre, elles constituent une partie des archives de la recherche.* »<sup>2</sup> Plusieurs éléments ici contrastent avec les définitions précédentes : les deux traits saillants sont d'une part les termes génériques d'« information » et de « matériaux », laissant une plus grande marge de manœuvre dans leur appréhension, d'autre part les actions de « collecte » et de « documentation » : les données brutes ou élaborées, pour être des données de la recherche, doivent être documentées, elles doivent posséder les attributs externes qui en permettent l'intelligibilité.

Cet aspect de documentation, pour terminer le tour d'horizon des définitions, renvoie à un engagement spécifique de la production des données de la recherche à l'ère du numérique et de la « science des données » : celui de nouveaux acteurs extérieurs aux communautés scientifiques au sens large qui, nous le présentions en introduction, s'emparent du sujet dans sa forme politique et financière. Cette caractéristique est loin d'être anecdotique car l'impulsion des acteurs politiques et des organismes de financement dans la prise de conscience des faiblesses en termes de gestion des données de la recherche est depuis les années 2000 grandissante. Si nous prendrons le temps de revenir sur la chronologie de ce déploiement, il s'agit

---

<sup>1</sup> Ainsi la définition de l'OCDE qui les exclue nommément : OCDE, *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, 2007, p. 18.

<sup>2</sup> Julien Pomart, « AAF / Section Aurore : Un groupe de travail sur les données de la recherche », *Archives de la Fondation Maison des sciences de l'homme*, 18 juillet 2014, <https://archivesfmsh.hypotheses.org/1209>.

toutefois ici de souligner que certains chercheurs, conscients du poids et du rôle croissants des organismes de financement à toutes les étapes de la recherche, se fondent sur cet aspect politique et financier pour élaborer de premières définitions adaptées à ces nouveaux cadres hors des champs de la recherche. C'est notamment le cas du conseil scientifique de l'Inra en 2012 qui, à la suite du rapport du groupe de travail « Gestion et partage des données », donne la définition suivante, reprenant les principaux termes croisés jusqu'à présent : « *En l'absence d'une définition juridique de ce qu'est une donnée, nous appellerons dans le contexte de ce rapport, donnée scientifique, ou donnée de la recherche, l'enregistrement d'une information qui représente le matériau de base d'une activité de recherche ayant bénéficié d'un financement sur fonds publics.* »<sup>1</sup>

Ainsi, qu'est-ce qu'une donnée de la recherche ? Les définitions sont nombreuses et plurielles ; parfois en creux, souvent énumératrices, elles se rejoignent toutefois sur quelques termes précis, dont celui d'enregistrement. Ce dernier se confronte, dans la tradition des sciences de l'information et de la communication, aux notions de document, de contenu ou encore d'inscription. Selon Bruno Bachimont, Le contenu est « *le fait premier du matériel faisant sens* »<sup>2</sup> et « *ce qui s'adresse à quelqu'un qui le reçoit comme tel* »<sup>3</sup>. Le contenu, lorsqu'il est enregistré, se fait *inscription*, qui est la condition d'existence dans le temps du contenu par l'enregistrement. Le document enfin est une inscription en contexte : « *De l'inscription qui, étant un contenu, repose sur une forme sémiotique qui propose un registre du sens, on passe au document qui concrétise sa signification dans son contexte.* »<sup>4</sup> Le document est alors le lieu de « *tension entre l'objet documentaire* » manipulable « *et le contexte documentaire qui lui donne son intelligibilité* »<sup>5</sup>, il est la « *permanence et la mémoire de l'événement* ». Dans cette perspective, les données de la recherche sont l'ensemble des informations d'un événement dont la traduction intelligible en document se réalise par leur enregistrement à partir d'une technologie intellectuelle désignée. Elles sont donc « *le vecteur de l'intelligence et de la compréhension* »<sup>6</sup> de l'événement ; leur cadre d'avènement est la technologie intellectuelle qui les restitue en contexte<sup>7</sup>.

## La production des données à l'ère numérique

Le terme de numérique, le plus souvent déterminé comme attribut, recouvre une large gamme d'acceptations mais demeure un « *impensé* »<sup>8</sup> à la fois éthique et

---

<sup>1</sup> Christine Gaspin, Dominique Pontier, Laurence Colinet, Frédéric Dardel, Alain Franc, *et al.*, *Rapport du groupe de travail sur la gestion et le partage des données*, Avis du conseil scientifique, Inra, juin 2012, p. 9.

<sup>2</sup> *Ibid.*, p. 18.

<sup>3</sup> *Ibid.*, p. 22.

<sup>4</sup> *Ibid.*, p. 35.

<sup>5</sup> *Ibid.*, p. 17.

<sup>6</sup> Bruno Bachimont, « L'archive et la massification des données : une nouvelle raison numérique », *La Gazette des archives*, n°245, 2017, p. 31.

<sup>7</sup> Pascal Robert, *Mnémotechnologies. Une théorie générale critique des technologies intellectuelles*, *op. cit.*, p. 53.

<sup>8</sup> *Ibid.*, p. 310.

politique. Il est avant tout mobilisé pour son pouvoir instrumental inégalé et tend en cela à reléguer en second plan sa capacité à mettre au jour « *des problèmes inédits qui n'existeraient pas dans un contexte où cette puissance technique serait absente* »<sup>1</sup>. C'est donc sous-jacent une évolution cognitive significative que le numérique induit, d'autant plus dans le champ de la production des savoirs sur le réel dont l'appréhension dépend étroitement de la technique. Ainsi dans le monde de la recherche, Valérie Schafer l'énonçait en 2014, « *le numérique a contribué à accompagner et modifier les pratiques scientifiques elles-mêmes et la chaîne de la communication et de l'information scientifiques s'est transformée, complexifiée* »<sup>2</sup>. Evidemment, dès lors que le numérique se fait non plus outil mais environnement, il « accompagne et modifie » à tout niveau de la recherche ; c'est cependant sur le sujet des données de la recherche que nous nous attarderons ici. En effet, les outils numériques ont modifié à la fois leur nature, leur mode de production, les conditions et le cadre de leur traitement et les modalités de leur préservation.

Le trait saillant des effets du numérique sur les données de la recherche s'observe dans leur prolifération. Dans le cadre de l'Inra et de la recherche agronomique, Denis Poupardin, dans son état des lieux des archives de l'Institut de 1993, identifiait d'ores et déjà cette problématique que lui partageaient les chercheurs. « *“On n'arrive même plus à faire le tri des données qui nous sont utiles”* » selon un chef de service de laboratoire de biologie végétale ; « *“Il y a eu une époque où on n'avait pas assez de données. Aujourd'hui, on en accumule plus qu'on ne parvient réellement à en traiter !”* »<sup>3</sup>, renchérit un autre. Cette prolifération conduit alors à distinguer les données passées des données présentes, les secondes nécessairement plus justes et véridiques par la seule raison de leur primeur. Denis Poupardin se souvient ainsi en 2008 : « *Croulant déjà sous la masse des données fournies par les appareils nouveaux qui fonctionnaient souvent en continu dans leurs laboratoires, ceux-ci ne ressentaient plus le besoin d'en extraire d'autres d'archives plus anciennes, souvent mal classées et au contenu difficile à interpréter.* »<sup>4</sup> Les outils numériques soulevaient ainsi d'ores et déjà la problématique de la massification des données et celle de leurs conditions d'existence dépendant de l'évolution technologique des instruments. Les données produites par des appareils rapidement obsolètes étaient elles-mêmes jugées, par le truchement de leur outil de production, imprécises, partielles, si ce n'est tout simplement fausses : « *“A mesure que les appareils se perfectionnent, explique un chercheur de l'Inra, les anciennes mesures deviennent de moins en moins crédibles. Aussi devient-il hasardeux de vouloir rapprocher des données obtenues à des dates différentes pour conclure à une évolution* »<sup>5</sup>. Un autre chercheur observe quant à lui qu'en éco-toxicologie,

---

<sup>1</sup> Bruno Bachimont, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques », *op. cit.*

<sup>2</sup> Valérie Schafer, « Renouveau de la science et science du renouveau à l'heure numérique », *op. cit.*, p. 21.

<sup>3</sup> Mireille Jean, Denis Poupardin, « Les archives des unités de recherche : le point de vue des scientifiques de l'Institut national de la recherche agronomique (I.N.R.A.) », *La Gazette des archives*, n°176, 1997, p. 45.

<sup>4</sup> Denis Poupardin, « Communication », Colloque « Archorales » du 2 décembre 2008, Inra, *op. cit.*

<sup>5</sup> Mireille Jean, Denis Poupardin, « Les archives des unités de recherche : le point de vue des scientifiques de l'Institut national de la recherche agronomique (I.N.R.A.) » dans *La Gazette des archives*, n°176, 1997, p. 46.

relate Denis Poupardin, « 80 % des données anciennes sont fausses, pour des raisons instrumentales et méthodologiques »<sup>1</sup>.

Le numérique « accélère tout un ensemble de phénomènes que l'on avait l'habitude de traiter sur le temps long : la fragilité des supports, l'obsolescence des techniques, la massification des contenus »<sup>2</sup>. Surtout, il modifie les termes de la collecte, du traitement, de la visualisation et de la préservation des données de la recherche, soit autant de modes d'actions qui, selon Bruno Bachimont, s'appliquent selon des principes arbitraires et indépendants les uns des autres. Dans ce basculement épistémologique de la mesure à la donnée, du poids des lois et de l'universalité à la dynamique de la masse et de l'équivocité, plusieurs ruptures s'opèrent : « une rupture des données par rapport à leur origine et leur nature, le mode de collecte les rassemblant en dépit de leur hétérogénéité ; une rupture du traitement par rapport aux données ; une rupture de ce qui est montré par rapport à ce qui est calculé »<sup>3</sup>. Ces ruptures offrent alors l'opportunité de travailler les grandes masses de données dont la cohérence n'est pas l'attribut premier. Elles valorisent en effet la rencontre de données hétérogènes indépendamment de leur structuration disciplinaire ; elles rapprochent ce qui auparavant était maintenu dans des univers différents<sup>4</sup>. Ce n'est donc pas sans hasard que l'on observe parallèlement au déploiement de ce renversement épistémologique le développement des approches systémiques accordant une large place à la rencontre des disciplines et de leurs résultats.

La recherche agronomique du XXI<sup>e</sup> siècle la défend, afin d'appréhender les enjeux environnementaux, sanitaires et alimentaires à venir qui ne peuvent plus se comprendre par la seule étude des cellules qui composent le vivant, mais bien par un regard large, interdisciplinaire, permis par les ressources inégalées du numérique. C'est la biologie des systèmes qui, « du gène à l'agroécosystème et dans tous les règnes du vivant, nécessite de croiser un ensemble de disciplines formelles (mathématiques et informatique notamment) et de disciplines d'observation (sciences du vivant et physiques) »<sup>5</sup>. Elle advient grâce à « la gestion, l'intégration et l'exploitation des données massives », comme l'Institut l'annonce dans son document d'orientation « #Inra2025 ». L'Inra acte dès lors du rôle des mégadonnées dans l'élaboration de nouvelles questions de recherche qui « vont émerger de jeux de données sans précédent » ; de nouvelles questions que les communautés de chercheurs reconnaissent ne pas avoir abordées jusqu'à présent, voire ni même imaginées<sup>6</sup>. Dans ce mouvement, « certaines communautés scientifiques naissent ; d'autres révisent leurs approches »<sup>7</sup> : autrement dit, le numérique joue un rôle déterminant dans la formation de collectifs de pensée, au sens de Ludwik Fleck, qui

---

<sup>1</sup> *Ibid.*, p. 46.

<sup>2</sup> Bruno Bachimont, *Patrimoine et numérique*, *op. cit.*, p. 13.

<sup>3</sup> Bruno Bachimont, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques », *op. cit.*

<sup>4</sup> *Ibid.*

<sup>5</sup> Odile Vilotte, Francois Houllier, *et al.*, « Document d'Orientation #Inra2025 », *op. cit.*, p. 31.

<sup>6</sup> *Ibid.*, p. 29.

<sup>7</sup> *Ibid.*, p. 29.

se créent et se réunissent autour d'un style de pensée porteur en mesure de concurrencer les communautés concurrentes. Qui de la nouvelle épistémologie de la donnée ou de l'approche systémique précède l'autre ? Il faudrait une étude bien davantage approfondie pour saisir les entremêlements de leur influence respective. Toutefois, si le numérique se fait définitivement technologie intellectuelle majeure et légitimée en environnement scientifique, ses liaisons avec l'avènement de ce nouveau style de pensée qu'est la systémique ne seraient pas illogiques.

Ainsi, enjeu commun aux données de la recherche, à l'épistémologie de la donnée, les nouveaux collectifs de pensée se confrontent aux mégadonnées, ou *big data*. Définies comme « *l'ensemble des données produites en temps réel et en continu, structurées ou non, et dont la croissance est exponentielle* »<sup>1</sup>, elles se distinguent dans leur méthodologie de production privilégiée qui se voudrait automatique. Si le terme « automatique », en environnement numérique, peut laisser à désirer, les données massives ont pour cadre de définition la règle des « trois V », volume, variété, vitesse, auxquels s'est ajouté au fur et à mesure des années un quatrième adapté aux enjeux de leur management, la véracité. Les données de la recherche et les données massives sont distinctes : certaines de l'une peuvent être de l'autre, mais elles ne se confondent pas. Pour Lyne Da Silva, la collecte, l'analyse et la fouille constituent les enjeux centraux des mégadonnées quand les données de la recherche, elles, concentrent les questions de la planification de leur préservation et de leur diffusion<sup>2</sup>. La problématique du volume est cependant essentielle, tant pour les unes que les autres. Si les questions de leur stockage peuvent en effet attirer l'attention, c'est surtout la dimension de leur traitement qui interroge ; d'ailleurs, « *personne ne consulte les données elles-mêmes. Des algorithmes fonctionnent selon une lecture globale [...] : les outils lisent les informations pour nous et les reconstruisent à travers des visualisations* » qui sont d'ailleurs parfois « *difficiles à interpréter du fait de la difficulté de comprendre ce qui est visualisé* »<sup>3</sup>.

Par ailleurs, si d'aucuns parlent de *data deluge*, il peut être toujours pertinent d'apporter un regard historique sur les questions de prolifération de données issues et à destination de l'activité scientifique. Ainsi, selon l'historien des sciences Bruno Strasser, les scientifiques de l'ère numérique ne sont pas les premiers à se confronter à des productions de données inédites par leur nature ou leur volume. Dès la Renaissance, des communautés d'individus produisaient et/ou recevaient des masses de données, provenant notamment des Amériques, des cabinets de curiosités ou encore des Muséums d'histoire naturelle<sup>4</sup>. Toutefois Bruno Strasser ne cherche pas à démontrer une croissance, ininterrompue depuis, du volume des données à traiter mais invite plutôt à regarder du côté de la relation entre production des données et

---

<sup>1</sup> Lyne Da Sylva, « Les données et leurs impacts théoriques et pratique sur les professionnels de l'information », *op. cit.*

<sup>2</sup> *Ibid.*

<sup>3</sup> Bruno Bachimont, « L'archive et la massification des données : une nouvelle raison numérique », *La Gazette des archives*, n°245, 2017, p. 40.

<sup>4</sup> Bruno Strasser, « Data-driven Sciences : From Wonder Cabinets to Electronic Databases », *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 43, n° 1, mars 2012, p. 85 à 87.

dispositifs de traitement de celles-ci pour comprendre les spécificités historiques et techniques de notre temps.

En effet, une des caractéristiques des mégadonnées est peut-être moins leur abondance que les dispositifs techniques et logiciels élaborés pour les traiter. Or, cette abondance, qui peut être perçue comme un progrès pour la recherche scientifique, ne peut l'être qu'à l'aide d'une gestion raisonnée de ces données car « *l'information n'est jamais neutre et le lien se complique, dès qu'il y a pléthore, entre information, connaissance, culture et communication, pour les sciences, comme pour tout autre domaine* »<sup>1</sup>. Surtout, cette information pléthore advient dans un « *capitalisme de plateforme* » au sein duquel le rôle des GAFAM, alerte l'Inra en 2019, est maintenant « *prépondérant dans la collecte, la gestion et le traitement des données qui deviennent centrales dans la production des connaissances, à l'intersection des relations de pouvoir, de la valeur économique et des connaissances validée* »<sup>2</sup>. Les données massives, en relation avec les données de la recherche, impliquent donc une éthique et une politique des *Big Data* auxquelles peuvent contribuer à façonner les professionnels de l'information scientifique et technique en charge de la gestion des données de la recherche et de leur pérennisation. Ceux-ci sont en effet en relation directe avec les communautés de scientifiques en capacité de donner l'intelligibilité et l'interprétation nécessaire à la compréhension de ces données massives.

## LES AMBITIONS DU PARTAGE DES DONNEES

### Les sciences en crise

« *Le progrès constitue véritablement pour nous à la fois une mesure de l'avancée du temps et la marque identificatoire qui autorise celui qui parle à juger. Qui autorise aussi à simplifier les récits, puisque le progrès permet de sélectionner dans une situation ceux qui sont dans l'illusion et ceux qui sont dans la vérité. Le progrès fait le tri entre ce qui est digne d'être conservé et amplifié et ce qui peut, avec quelques douleurs transitoires, être abandonné au passé.* »<sup>3</sup>

La philosophe des sciences Isabelle Stengers définit ici les traits caractéristiques du concept de « progrès » qui, depuis le XIX<sup>e</sup> siècle, structure l'histoire des sociétés occidentales. Ces traits toutefois rencontrent au cours de la seconde moitié du XX<sup>e</sup> siècle des résistances de plus en plus fortes, mettant au jour un récit sous-jacent contraire « au sens de l'histoire », du progrès et de la modernité. Les « Trente Glorieuses », pourtant, se fondaient sur un scientisme souverain actant les sciences « *à l'origine de tout développement technique, lui-même à l'origine du*

---

<sup>1</sup> Dominique Wolton : « Abondance et gratuité : pour quoi faire et jusqu'où ? Entretien avec Joëlle Farchy, Pascal Froissart et Cécile Méadel », Valérie Schafer (dir.), *op. cit.*, 2014, p. 44.

<sup>2</sup> Inra, *Transition numérique et pratiques de recherche et d'enseignement supérieur en agronomie, environnement, alimentation, sciences vétérinaires à l'horizon 2040*, Résumé de la prospective, 27 juin 2019, p. 22.

<sup>3</sup> Isabelle Stengers, *L'invention des sciences modernes*, Paris, Éditions La Découverte, 1993, p. 170-171.

*progrès industriel, économique et social* »<sup>1</sup>, et finalement du progrès humain. L'engagement de la France dans la course au nucléaire, à la fois civil et militaire, est alors caractéristique de cette volonté de rendre au pays la place que le pouvoir politique estimait lui être due dans le concert des nations<sup>2</sup>. L'agriculture, dans un pays en reconstruction, devient tout autant un enjeu de modernité et « *le monde rural français, ses habitants, ses paysages, ses infrastructures*, selon l'historien Jean-Luc Mayaud, *ne sont bientôt plus reconnus dans les discours de l'englobant que pour une seule et unique fonction : produire des biens agricoles, en quantités toujours croissantes* »<sup>3</sup>. La recherche agronomique se fait alors le bras armé du pouvoir politique et technique dans la modernisation de l'agriculture française par le biais de l'Institut national de la recherche agronomique créé au sortir de la guerre.

Outre la problématique de la mécanisation des activités de production, les enjeux au sein de l'Inra se portent sur les sciences du vivant qui entrent au cours des années 1960 dans une nouvelle ère marquée par les Prix Nobels de physiologie ou de médecine de trois biologistes français récompensant leurs avancées dans le champ de la biologie moléculaire. Si Jacques Monod, un de ses architectes, voit dans cette étude des *mécanismes* du vivant à de très petites échelles la levée de ses derniers secrets<sup>4</sup>, d'autres scientifiques restent circonspects. L'Inra, au début des années 1970 est ainsi « *à la croisée des chemins* »<sup>5</sup>, oscillant entre l'affermissement du génie génétique dans un axe de recherche fondamentale rivalisant à l'international, et la poursuite d'une science des champs qui se préoccupe des résultats et de leur développement. Originellement généraliste et proche du terrain, la recherche agronomique s'engage dans la voie du réductionnisme biologique et confie aux soins des instituts techniques les enjeux de la valorisation des résultats. Or tous les chercheurs de l'Institut ne partagent pas cet idéal de recherche fondamentale dans le champ de l'agronomie que certains préfèrent saisir dans son amplitude technique et sociale, et non seulement biologique, afin de résister, selon l'historien Pierre Cornu, « *à l'érosion, au délitement, à l'atomisation du social et du vivant* »<sup>6</sup>.

A plus large échelle, les critiques à l'encontre des avatars du progrès gagnent en visibilité, tandis que le compromis fordiste, toléré dans les premières années des « Trente Glorieuses » et promu par les élites politiques et scientifiques dans

---

<sup>1</sup> Dominique Pestre, « Le nouvel univers des sciences et des techniques : une proposition générale », Dominique Pestre, Amy Dahan (dir.), *Les sciences pour la guerre*, Paris, Éditions de l'EHESS, 2004, p. 13.

<sup>2</sup> Gabrielle Hecht, *Le rayonnement de la France : énergie nucléaire et identité nationale après la Seconde Guerre mondiale*, Paris, Éditions La Découverte, 2004, 346 p.

<sup>3</sup> Jean-Luc Mayaud, « La porte de l'étable. Approche sociohistorique des acteurs, techniques et représentations de la production du vivant (XIX<sup>e</sup> et XX<sup>e</sup> siècles) », Bertrand Hervieu, Bernard Hubert (dir.), *Sciences en campagne*, Paris, Éditions de l'Aube, 2009, p. 23.

<sup>4</sup> Jacques Monod, *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, Paris, Editions du Seuil, 1970, 243 p.

<sup>5</sup> Pierre Cornu, « La recherche agronomique française dans la crise de la rationalité des années soixante-dix », *Histoire de la recherche contemporaine*, t. III, n° 2, 2014, p. 154.

<sup>6</sup> Pierre Cornu, *La passion naturaliste. Trois études d'anthropologie historique de la « question agraire » à l'époque contemporaine*, Mémoire original pour l'habilitation à diriger des recherches en histoire contemporaine, Université Lyon 2, 2012, p. 340.

l'impératif de modernisation<sup>1</sup>, se morcèle. Les mouvements antinucléaires commencent à faire entendre leurs voix à l'encontre du pouvoir technocratique<sup>2</sup> qui ne peut plus sereinement discriminer « *ceux qui sont dans l'illusion et ceux qui sont dans la vérité* », selon les mots d'Isabelle Stengers. Dans un autre registre, les critiques s'amplifient à l'encontre du productivisme agricole dont on constate progressivement les effets socialement aliénants et préjudiciables pour l'environnement. En effet, des processus inattendus, sinon ignorés ou sous-estimés, émergent : de premiers avertissements marquants apparaissent dès les années 1960 sur les conséquences de l'agriculture intensive, notamment dans l'utilisation de pesticides<sup>3</sup>. Ces avertissements essaient les décennies suivantes en s'élargissant aux problématiques alimentaires et sanitaires. La crise de « la vache folle » des années 1990 marque alors un tournant décisif pour la recherche agronomique, et entame la confiance des consommateurs et des citoyens. L'Inra est ainsi sommé de « *devoir fournir une expertise rassurante sur une crise dont il n'est pas responsable, mais qui a révélé bien des failles dans un dispositif sanitaire dans lequel il joue un rôle central* »<sup>4</sup>.

Les sciences sont dès lors brutalement confrontées aux conséquences de leurs propres résultats et convoquées sur la place publique pour y répondre. Le biologiste Richard Lewontin souligne avec ironie cette quête éperdue de la recherche agronomique pour rattraper inlassablement les défaillances dont elle est elle-même la source : « *Ainsi, "les agriculteurs ont besoin d'un pesticide pour éliminer un insecte devenu un ravageur parce que les adventices sur lesquelles il vivait ont été détruites par les herbicides, lesquels ont été introduits pour éliminer le sarclage mécanique, lequel est rendu impossible par la haute densité de plantation, laquelle a été accrue parce que les plantes ont été sélectionnées pour leur productivité à haute densité, laquelle leur permet de tirer parti de l'utilisation massive d'engrais, ce qui rend les plantes encore plus appétissantes aux ravageurs et ainsi de suite"* ». »<sup>5</sup> Les sciences confrontées aux méfaits de leurs propres produits induisent une scientification réflexive, c'est-à-dire la mise en œuvre du « *doute scientifique jusqu'aux fondements intrinsèques et aux conséquences externes de la science elle-même* »<sup>6</sup>. C'est une évolution scientifique et technique qui « *devient à elle-même un problème* »<sup>7</sup>, une *modernité réflexive* selon le sociologue Ulrich Beck, qui est confrontée à des objets avec lesquels les sciences ne parviennent plus à composer.

---

<sup>1</sup> Céline Pessis, Sezin Topçu, Christophe Bonneuil (dir.), *Une autre histoire des « Trente Glorieuses »*, Paris, La Découverte, 2013, 309 p.

<sup>2</sup> Christophe Bonneuil, « Les transformations des rapports entre sciences et société en France depuis la Seconde Guerre mondiale : un essai de synthèse » dans Joëlle le Marec, Igor Babou, *Sciences, médias et société*, Actes du colloque des 15, 16, 17 juin 2004, Lyon, ENS-LSH, p. 22.

<sup>3</sup> Rachel Carson, *Silent Spring*, Greenwich, Fawcett, 1962, 304 p.

<sup>4</sup> Pierre Cornu, Egizio Valceschini, Odile Maeght-Bourmay, *L'histoire de l'Inra, entre science et politique*, Versailles, Editions Quae, 2018, p. 348.

<sup>5</sup> Richard Lewontin cité par Jean-Pierre Berlan, « Champs, contre-champ » dans Georges Fermé (dir.), *Science, pouvoir et argent. La recherche entre marché et politique*, Paris, Autrement, 1993, p. 87.

<sup>6</sup> Ulrich Beck, *La société du risque. Sur la voie d'une autre modernité*, op. cit., p. 342.

<sup>7</sup> *Ibid.*, p. 350.

Se profile ainsi la fin d'un régime de rationalité hégémonique depuis plusieurs siècles, un désenchantement de la scientification du monde qui travaille les dualismes recteurs de notre modernité, tels nature culture, sujet objet, tradition modernité, et dont les sciences modernes s'en faisaient jusqu'à présent les hérauts. Celles-ci se fondent en effet, selon l'anthropologue de la nature Philippe Descola, sur une idée de nature ontologiquement séparée de l'idée de culture, et qui « *comme champ d'enquête et d'expérimentation scientifique, comme objet à exploiter et à améliorer, cette nature-là accède à une existence que bien peu songent à remettre en doute* »<sup>1</sup>. Une nature sur laquelle l'homme, « *comme maître et possesseur* », y travaille l'universalité de ses caractéristiques et de ses lois, où il se fait seul *sujet* à traiter le reste des existants comme *objet*. Une *idée* de nature qui est une construction unique de notre rapport au monde et historiquement loin d'être majoritaire au sein des groupes humains.

Or l'un des troubles de la modernité réflexive résulte de la confrontation des sciences à des objets qu'elles n'arrivent plus à traiter, justement, comme de simples *objets*. Ce sont, selon le sociologue des sciences Bruno Latour, des objets hybrides de nature et de culture, des « *chimères* » qui ne se sentent bien installées « *ni du côté des objets, ni du côté des sujets, ni au milieu* »<sup>2</sup>, partagées entre la nature et la culture. C'est le trou dans la couche d'ozone, le réchauffement climatique, ou encore l'ère géologique de l'Anthropocène qu'est la nôtre et qui acte définitivement ce brouillage des frontières. Ce sont des pesticides, des plantes sélectionnées et des insectes ravageurs qui résistent, par l'imprévisibilité de leurs effets sur le monde, à leur objectivation comme simple matière soumise au déterminisme des lois physiques et sont dès lors à la source du développement d'approches totalisantes.

Ainsi l'Inra, à partir des années 1997-1998, multiplie les efforts de réorganisation stratégique, bousculant les départements scientifiques et les directions en faisant du terme « *environnement* » le mot clef gouvernant la recherche agronomique dans l'avènement des approches systémiques<sup>3</sup>. La notion de système agraire apparaît dès les années 1970, comme style de pensée concurrent et, dans un premier temps, abondamment délégitimé par la biologie moléculaire<sup>4</sup>. Il exprime alors « *l'interaction entre un système bioécologique représenté par le milieu naturel et un système socioculturel à travers des pratiques issues notamment de l'acquis technique* »<sup>5</sup>. Le terme finalement évolue, devenant aujourd'hui le style de pensée dominant par un changement de paradigme appréhendant les systèmes biologiques « *comme des systèmes dynamiques complexes pouvant et devant être étudiés en tant*

---

<sup>1</sup> Philippe Descola, *Par-delà nature et culture*, op. cit., p. 134.

<sup>2</sup> *Ibid.*, p. 72.

<sup>3</sup> Pierre Cornu, « Emergence et légitimation de la problématique environnementale dans la recherche agronomique publique française », Séminaire de recherche, 1er février 2019, Laboratoire d'études rurales, Lyon.

<sup>4</sup> Pierre Cornu, « La recherche agronomique française dans la crise de la rationalité des années soixante-dix », *Histoire de la recherche contemporaine*, t. III, n° 2, 2014, p. 154-166.

<sup>5</sup> Bertrand Vissac, André Hentgen, *Présentation du département SAD. Éléments pour une problématique de recherche sur les Systèmes agraires et le Développement*, Inra, 1979, p. 6, cité dans Pierre Cornu, « La recherche agronomique française dans la crise de la rationalité des années soixante-dix », op. cit.

que tels »<sup>1</sup>. L'approche réductionniste a échoué à rendre compte seule du réel, dès lors qu'il n'apparaît plus « possible de déduire simplement le comportement global des systèmes biologiques à partir des propriétés de leurs composantes »<sup>2</sup>. Ces approches par les systèmes induisent ainsi des transformations profondes dans les sciences du vivant résultant dans une certaine mesure de l'essor des sciences participatives, mais aussi et surtout de l'inflation des données en biologie, « permise par le développement sans précédent des technologies d'acquisition à toutes les échelles (données omiques, phénotypiques, écologiques, climatiques, ...) »<sup>3</sup>.

La recherche agronomique est dès lors « fortement attendue, selon l'Institut dans son document d'orientation INRAE2030, pour mieux comprendre les ressorts de ces transformations et pour contribuer à concevoir des solutions opérantes », et ce « pour les acteurs privés et publics »<sup>4</sup>. INRAE acte d'ailleurs à cette occasion la fin de la vision déterministe face à ses objets de recherche complexes et totaux qui sortent des murs du laboratoire : « La recherche n'est pas déterministe : il ne suffit pas d'allouer des moyens importants aux priorités identifiées pour apporter les réponses attendues. La part de sérendipité constitutive de la Science et le renouvellement rapide des enjeux invitent au contraire à maintenir une souplesse et une agilité de l'organisation, qui doit favoriser une certaine diversité des approches et une certaine autonomie des projets de recherche. »<sup>5</sup> Confrontées aux objets hybrides difficilement saisissables qui imposent cette « souplesse » et cette « agilité », les sciences ne sont donc plus légitimes pour produire seules un discours sur le réel en s'enfermant dans la boîte noire qu'est le laboratoire<sup>6</sup>. Or dans la modernité réflexive où les chercheurs perdent leur monopole sur la connaissance, « la science devient de plus en plus nécessaire, mais de moins en moins suffisante à l'élaboration d'une définition socialement établie de la vérité »<sup>7</sup>. Surtout, elle doit composer avec un nouveau régime de véridicité fondé sur l'ouverture, le partage et la transparence afin d'une part d'appréhender les objets complexes que sont les systèmes biologiques, d'autre part de répondre de ses actes auprès de la société dans laquelle elle se déploie, et de laquelle elle ne peut s'extraire. L'avènement du numérique à ce titre œuvre pour un décloisonnement dès lors qu'il « rend la fabrique de la science visible, et parfois des coulisses qu'elle aimerait cacher »<sup>8</sup> selon Valérie Schafer.

---

<sup>1</sup> Carole Caranta, Hervé Monod, *et al.*, « Réflexion prospective interdisciplinaire. Approches prédictives pour la biologie et l'écologie », Rapport de synthèse, Inra, 2019, p. 6.

<sup>2</sup> *Ibid.*, p. 6.

<sup>3</sup> *Ibid.*, p. 6.

<sup>4</sup> INRAE, « INRAE2030. Partageons la science et l'innovation pour un avenir durable », INRAE, document d'orientation, décembre 2020, p. 10.

<sup>5</sup> *Ibid.*, p. 37.

<sup>6</sup> Bruno Latour, *la science en action*, *op cit.*, 663 p.

<sup>7</sup> Ulrich Beck, *La société du risque. Sur la voie d'une autre modernité*, *op. cit.*, p. 343.

<sup>8</sup> Valérie Schafer, « Renouveau de la science et science du renouveau à l'heure numérique », Valérie Schafer (dir.), *Information et communication scientifiques à l'heure du numérique*, Paris, CNRS Editions, 2014, p. 20.

## Les enjeux scientifiques et éthiques

De l'*open access* à l'*open government*, l'idéal d'ouverture a mobilisé un ensemble de communautés au premier rang desquelles les scientifiques se sont positionnés en précurseurs en la matière depuis les années 1990. Faisant parfois remonter les héritages aux premiers modèles de publication scientifique du XVII<sup>e</sup> siècle<sup>1</sup>, c'est en premier lieu les enjeux de partage des données et de libre accès à la littérature scientifique qui interrogèrent certains pans de la recherche scientifique, en proie à l'essoufflement de tout un système de contrôle et d'évaluation des connaissances scientifiques.

Le constat récurrent au cours des deux dernières décennies de cas de fraudes et de rétractations dans la recherche scientifique est une des nombreuses manifestations de cet essoufflement. La représentation ci-dessous<sup>2</sup> explicite ainsi l'évolution parallèle des publications et des rétractations qui, en proportion, est en forte hausse ces dernières années.



Nombre de publications et de rétractations par année - Yaroslav Pigenet, 2014.

Le phénomène de l'édition prédatrice<sup>3</sup> couplé au modèle économique éreintant les organismes de recherche et les établissements d'enseignement supérieur et de recherche (ESR)<sup>4</sup>, en contexte de pression à la publication pour l'évaluation de la recherche – dont témoigne la sentence « *publish or perish* » –, sont autant de raisons qui entament la confiance au sein et en dehors des communautés scientifiques. Les enjeux scientifiques se cristallisent fortement sur la problématique de l'accès à la

<sup>1</sup> Bernard Rentier, *Science ouverte, le défi de la transparence*, Bruxelles, Académie royale de Belgique, 2018, 144 p.

<sup>2</sup> Yaroslav Pigenet, « Fraude : mais que fait la recherche ? », *CNRS Le journal*, 2014, Vol. 278.

<sup>3</sup> Arnaud Saint-Martin, « L'édition scientifique "piratée". Passage en revue et esquisse de problématisation », *Zilsel*, vol. 4, n° 2, 2018, p. 179 à 202.

<sup>4</sup> Richard Van Noorden, « Open Access: The True Cost of Science Publishing » dans *Nature*, 2013, n° 495, p. 426–29, [En ligne], <http://dx.doi.org/10.1038/495426a>.

connaissance qui était, jusqu'à ces dernières années, particulièrement restreint du fait du coût des abonnements aux revues scientifiques détenues à hauteur de 40% du marché de l'édition scientifique<sup>1</sup> par une poignée d'entreprises : « de 1975 à 1995, leurs prix ont augmenté de 200 à 300 %, dépassant largement l'indexation des budgets de bibliothèques »<sup>2</sup>. Une restructuration de ce système s'impose donc, lui qui est bien souvent jugé « absurde » et ce jusqu'aux plus hautes instances de la recherche : « nous faisons la recherche, observe Antoine Petit, président-directeur général du CNRS, nous référons la recherche, nous éditons la recherche puis nous payons pour lire tout ce que nous venons de faire »<sup>3</sup>.

La question du libre accès des publications scientifiques rejoint par ailleurs les enjeux de la gestion des données de la recherche qui a un impact évident sur la méthodologie de contrôle de la production du savoir qu'est l'évaluation par les pairs : sans possibilité d'accéder aux données qui ont conduit aux résultats présentés, la valeur des faits scientifiques chute drastiquement. L'incapacité à apporter la preuve d'un élément clé d'une publication – certains parlent de *lost science*<sup>4</sup> – est ainsi « un risque majeur dans le contexte actuel de la recherche »<sup>5</sup> que les chercheurs ne peuvent plus ignorer. C'est d'ailleurs une observation mesurée à de multiples reprises : dans le champ de la génétique au début des années 2000, du fait de l'impossibilité d'accéder aux données, 28% des généticiens signalaient ne pouvoir confirmer les résultats de recherche publiés<sup>6</sup>. Plus récemment, une étude révèle que l'accessibilité aux données, toute discipline confondue, chute de 17% par an à la suite de la publication dans une revue scientifique<sup>7</sup>. Par ailleurs, au-delà des enjeux de transparence et de traçabilité de la recherche, il s'agit aussi d'optimiser la production des savoirs scientifiques qui, par l'absence d'une gestion rationnelle et partagée des données de la recherche, entre dans une boucle de redondances parfois bien dommageable alors que la science nécessite plutôt « des efforts dupliqués, qui consistent à répliquer des résultats »<sup>8</sup>. La science ouverte et le libre accès sont donc des enjeux scientifiques majeurs à toutes les échelles et pour toutes les disciplines.

---

<sup>1</sup> Jean-Baptiste Veyrieras, « Science ouverte : la révolution nécessaire », *CNRS Le Journal*, 8 janvier 2020, [En ligne], <https://lejournel.cnrs.fr/articles/science-ouverte-la-revolution-necessaire>.

<sup>2</sup> Françoise Vandooren, « Evolution de l'accès aux publications scientifiques », Joachim Schöpfel (dir.), *La publication scientifique : analyses et perspectives*, Paris, Lavoisier, 2008, p. 130.

<sup>3</sup> Antoine Petit, « La science ouverte, une révolution nécessaire », *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>4</sup> Direction de l'Information Scientifique et Technique du CNRS, « Quels enjeux pour le travail de la science ? », *Une Science ouverte dans une République numérique — Guide stratégique : Études et propositions en vue de l'application de la loi*, Marseille, OpenEdition Press, 2017, doi :10.4000/books.oep.1717.

<sup>5</sup> Alain Rivet, Marie-Laure Bachèlerie, Auriane Denis-Meyere, Delphine Tisserand, *Traçabilité des activités de recherche et gestion des connaissances. Guide pratique de mise en place*, Mission pour les initiatives transverses et interdisciplinaires, CNRS, 2018, p. 9.

<sup>6</sup> Eric G. Campbell, Brian R. Clarridge, Manjusha Gokhale, Lauren Birenbaum, Stephen Hilgartner, Neil A. Holtzman, David Blumenthal, « Data Withholding in Academic Genetics », *Journal of the American Medical Association*, 23/30 Janvier 2002 — Vol 287, n°4, p. 473.

<sup>7</sup> Timothy Vines, *et al.*, « The Availability of Research Data Declines Rapidly with article Age », *Current Biology*, 2014, Vol. 24, p. 94 à 97.

<sup>8</sup> Antoine Petit, « La science ouverte, une révolution nécessaire », *op. cit.*

Certaines communautés se sont toutefois historiquement davantage engagées dans les mouvements d'ouverture : c'est le cas des mathématiques qui sont selon Antoine Petit « *moins obnubilées par ses histoires d'impact factor* »<sup>1</sup>, mais aussi de l'astronomie qui a d'ailleurs été pionnière, engagée dès la fin du XX<sup>e</sup> siècle dans l'établissement de standards de partage des données nécessitant « *un lourd travail* » toutefois à l'origine « *de nouvelles pratiques qui ont révolutionné les méthodes de travail des scientifiques* »<sup>2</sup>. Des archives ouvertes pour le dépôt des versions *pre-print* de documents à la diffusion du modèle de partage de logiciels en *open source*, l'astronomie marque le monde de la recherche par sa « *culture très collaborative basée sur le fait que tous les grands projets d'astronomie (au sol et dans l'espace) sont réalisés dans le cadre de grands consortiums internationaux* »<sup>3</sup>. La recherche agronomique quant à elle, par les approches systémiques notamment, se retrouve engagée dans des infrastructures de recherche semblables et au centre desquelles, pour toutes les disciplines, les dispositifs techniques coûteux et élaborés de collecte des mégadonnées mobilisent une large communauté de scientifiques. Dès lors, « *comme ces données s'acquièrent automatiquement, leur possession n'est aucunement une affaire de mérite. Il en découle une volonté de mise en commun* »<sup>4</sup>, que le document d'orientation #Inra2025 confirme dans sa forme prospective :

« *En 2025, la production de données est accélérée (robotique, capteurs connectés, imagerie, simulation) et leur échange est généralisé. Nos infrastructures de recherche s'y seront adaptées dans leur organisation et dans leurs technologies. Notre capacité à recenser et qualifier les données les aura rendues accessibles au partage et à la réutilisation par d'autres, décuplant la capacité d'investigation collective pour des systèmes alimentaires durables, de la production à la consommation, dans un contexte de transition climatique.* »<sup>5</sup>

La notion explicite de « réutilisation » est ainsi au centre de l'action de l'Institut qui, tout en faisant de la recherche fondamentale, demeure, soulignons-le, une maison d'ingénieurs. La directrice de la Direction pour la science ouverte (DipSO) d'INRAE Odile Hologne nous le confirme lors d'un entretien : la science ouverte améliore certes la circulation des connaissances mais aussi et surtout œuvre « *à l'exploitation des contenus rendus accessibles* »<sup>6</sup> par l'outil numérique.

La notion implicite d'interdisciplinarité, quant à elle, est une part constitutive d'INRAE. Si l'interdisciplinarité joue un rôle dans la science ouverte, elle dépasse toutefois dans les murs d'INRAE le seul cadre de la révolution numérique et

---

<sup>1</sup> *Ibid.*

<sup>2</sup> Françoise Genova, « Vision internationale autour des données de recherche », *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>3</sup> Françoise Genova, *et al.*, « Building a Disciplinary, World-Wide Data Infrastructure », *Data Science Journal*, 16, 2017, DOI: <http://doi.org/10.5334/dsj-2017-016>.

<sup>4</sup> Jean-Gabriel Ganascia, « Sciences, réseaux et cognition après le déluge », Valérie Schafer (dir.), *Information et communication scientifiques à l'heure du numérique*, *op. cit.*, p. 54.

<sup>5</sup> Odile Vilotte, François Houllier, Philippe Mauguin, Claude Ronceray, Christine Cherbut, *et al.*, « Document d'Orientation #Inra2025 », Inra, 2016, p. 29.

<sup>6</sup> Odile Hologne, entretien du 16 juin 2021.

épistémologique actuelle. En effet, la recherche agronomique entretient depuis sa création une relation privilégiée avec l'interdisciplinarité, en mobilisant dans un premier temps, dans un versant plutôt « pluridisciplinaire », les sciences économiques et sociales en soutien du développement et de la valorisation des résultats scientifiques et des politiques de modernisation agricole<sup>1</sup>. Puis dans un second temps, dans le contexte du travail réflexif de l'Inra sur ses objectifs de recherche en temps de crise, la notion se renverse et l'interdisciplinarité se fait, selon le sociologue de l'Inra Marcel Jollivet au début des années 1990, « *une tentative de réponse* » à la crise de la rationalité technoscientifique et aux problèmes qu'elle induit. L'interdisciplinarité à l'Inra se veut alors en 1992, en réaction aux problématiques qu'elle rencontre, « *une pratique de recherche qui les prenne en compte et en charge* »<sup>2</sup>, pavant de la sortant le chemin pour la science ouverte. Le document d'orientation INRAE2030 publié en 2020 fait en effet écho à cette pratique de recherche prenant en compte et en charge « *la gestion des systèmes complexes étudiés (systèmes évolutifs, contraints, incertains, et pouvant impliquer une diversité d'acteurs)* », dont leur compréhension et leur développement « *sont consubstantiels de pratiques de plus en plus interdisciplinaires et transdisciplinaires* »<sup>3</sup>.

Toutefois ces pratiques interdisciplinaires, qu'elles soient le fruit d'une recherche agronomique en proie à des objets hybrides complexes impliquant « *la mobilisation de connaissances et données diverses et hétérogènes* »<sup>4</sup> ou la résultante des mouvements d'ouverture en science, peuvent être délicates à mettre en œuvre, sinon souvent au prix d'un savoir ou d'une communauté sacrifiée. La problématique des données numériques, résolument hétérogènes dans la recherche agronomique et ailleurs, interroge en effet la capacité des acteurs engagés à les traiter, et pouvant « *engendrer des exclusions et favoriser une science à deux vitesses* »<sup>5</sup> Ainsi l'interdisciplinarité face l'hétérogénéité des savoirs et des données doit être selon le philosophe des sciences Léo Coutellec « *une démarche d'humilité, pas de maîtrise. [...] Il s'agit de prendre soin des objets pas de les réduire dans une illusion de maîtrise.* »<sup>6</sup> Car la science ouverte, telles les deux faces d'une même pièce, est bien une réponse à la fois scientifique et éthique à un ensemble de questionnements qui traversent depuis plusieurs décennies les sciences en crise. Moins que les sciences, c'est peut-être d'abord le concept de « science » lui-même qui est en crise, « *une*

---

<sup>1</sup> Jean-Luc Mayaud, « La porte de l'étable. Approche sociohistorique des acteurs, techniques et représentations de la production du vivant (XIX<sup>e</sup> et XX<sup>e</sup> siècles) » dans Bertrand Hervieu, Bernard Hubert (dir.), *Sciences en campagne*, Paris, Éditions de l'Aube, 2009, p. 13 à 32.

<sup>2</sup> Marcel Jollivet, « Un chapitre de l'histoire récente d'une vieille et grande question : les rapports homme-nature » dans Marcel Jollivet (dir.), *Sciences de la nature, sciences de la société. Les passeurs de frontières*, Paris, CNRS Éditions, 1992, p. 27.

<sup>3</sup> INRAE, *Stratégies d'INRAE à horizon 2030*, INRAE, document d'orientation, décembre 2020, p. 32.

<sup>4</sup> *Ibid.*, p. 32.

<sup>5</sup> Inra, *Transition numérique et pratiques de recherche et d'enseignement supérieur en agronomie, environnement, alimentation, sciences vétérinaires à l'horizon 2040, Résumé de la prospective*, 27 juin 2019, p. 42.

<sup>6</sup> Léo Coutellec, *La science au pluriel. Essai d'épistémologie pour des sciences impliquées*, Versailles, Editions Quae, 2015, p. 76.

*crise de sa représentation et des moyens de sa reconnaissance* »<sup>1</sup>, à laquelle peut alors répondre les mouvements d'ouverture. Bien sûr les enjeux d'échanges et de partage des données se construisent, dans un premier temps, pour les communautés de chercheurs elles-mêmes ; mais la science ouverte se fait dans un second temps une ouverture à la société offrant l'opportunité à la recherche scientifique d'assumer sa part « impliquée » et « plurielle » selon les termes de Léo Coutellec. Toutes les sciences sont en effet « *nécessairement prises dans un contexte, immergées dans un paysage de valeurs et d'intentions, portées par des collectifs d'humains* »<sup>2</sup>. C'est pourquoi le directeur d'INRAE François Houllier dit de la recherche agronomique qu'elle est « *une science belle, utile et partagée* »<sup>3</sup> ; outre les interrogations sur le caractère esthétique d'une science, ce slogan – répété et resassé dans les médias et les documents de promotion de l'Institut – sous-tend une dimension forte d'éthique de la recherche et d'éthique *en* recherche, légitimant les activités de recherche agronomique sur des objets complexes sortis des murs clos du laboratoire.

Les chercheurs admettent ainsi progressivement l'ouverture de leurs travaux au grand public<sup>4</sup>, car la science, nous le disions précédemment, devient « de plus en plus nécessaire, mais de moins en moins suffisante à l'élaboration d'une définition socialement établie de la vérité »<sup>5</sup>. INRAE, ayant acté cette nouvelle dimension de la production et de la diffusion des savoirs, se déclare ainsi en 2020 « *au service de l'intérêt général* » en s'engageant « *à respecter, dans la conduite de ses travaux de recherche et dans son management, les principes et les valeurs éthiques et déontologiques de responsabilité, impartialité, intégrité, dignité et probité* »<sup>6</sup>. La notion de responsabilité, première des valeurs éthiques citées, est centrale. Dès lors qu'advient un nouveau régime de rationalité et de véridicité par l'implication d'un large éventail d'acteurs qui prennent leur part dans la création des savoirs, mais aussi dans l'assomption de leurs effets sur le réel, une responsabilité épistémique émerge. Cette responsabilité désigne, selon Léo Coutellec, une « *responsabilité sociale et politique liée aux savoirs* »<sup>7</sup> qui concerne non pas une personne, ni même une communauté de personnes mais bien la société dans son ensemble engagée dans un processus commun de production et de partage des connaissances, et la seule à même d'assumer l'ampleur des nouveaux risques qui, parfois, s'étendent à une échelle inimaginable telle l'Anthropocène. La responsabilité à propos des savoirs est donc collective, au-dessus des responsabilités juridiques et morales : nous sommes tous co-responsables du devenir des savoirs, de leur implication, de leur conséquence. Dès lors, « *il ne s'agit plus d'attendre des savoirs de la science (la demande sociale), il ne s'agit plus seulement de se battre contre les savoirs de la science (la critique sociale), il s'agit d'exiger son partage pour assumer collectivement la*

---

<sup>1</sup> *Ibid.*, p. 8.

<sup>2</sup> *Ibid.*, p. 55.

<sup>3</sup> François Houllier, « Une science belle, utile et partagée », *Le Journal du Dimanche*, 5 juin 2016, [En ligne].

<sup>4</sup> Cherifa Boukacem, « Table-ronde » dans *La science ouverte : une révolution nécessaire*, *op. cit.*

<sup>5</sup> Ulrich Beck, *La société du risque : sur la voie d'une autre modernité*, *op. cit.*, p. 344.

<sup>6</sup> INRAE, *Stratégies d'INRAE à horizon 2030*, INRAE, document d'orientation, décembre 2020, p. 8.

<sup>7</sup> Léo Coutellec, *La science au pluriel. Essai d'épistémologie pour des sciences impliquées*, *op. cit.*, p. 49.

*responsabilité des savoirs produits ou à produire (l'implication sociale) »<sup>1</sup>, ce que la science ouverte œuvre à réaliser.*

## **Les enjeux politiques et économiques**

Les mouvements d'ouverture marquent depuis les années 1990 une nouvelle dynamique majeure dans le champ de la recherche. Toutefois, par les enjeux pluriels dont le libre accès et la science ouverte sont porteurs, d'autres pans de la société lui nient de plus en plus le monopole, sinon de la création, du moins de la gestion et de la diffusion des savoirs. « *La production de connaissances est aujourd'hui considérée comme un enjeu trop important pour laisser ce minimum d'autonomie aux chercheurs* »<sup>2</sup> explique Isabelle Stengers. Or la notion d'autonomie est un principe fondamental au sein des sciences modernes, qui toutefois oscille en tension entre idéal et réalité. En effet, la science intéresse l'économie et les pouvoirs politiques ; et si elle se *doit* d'être désintéressée, les chercheurs, eux, ne peuvent l'être eu égard des considérations historiques, sociales, politiques, économiques, financières ou encore environnementales qui touchent chacun d'entre nous. Au-delà de ces perspectives, il s'agit surtout d'acter une science qui, particulièrement depuis la Seconde Guerre mondiale et l'avènement de la *Big Science*, celle des grands moyens et des grands programmes<sup>3</sup>, compose avec des horizons de recherche de progrès et d'innovation particulièrement complexes. Bruno Latour s'est ainsi évertué à mettre au jour dès les années 1980-1990 ces transactions que l'historienne des sciences Bernadette Bensaude-Vincent restitue en ces termes :

*« Toute discipline scientifique un tant soit peu innovante doit :*

*a) réussir à prendre en compte des "alliances" avec de groupes puissants acceptant que leurs propres fins passent par les succès de cette discipline ;*

*b) "mobiliser le monde", se doter d'un ensemble de ressources, d'instruments, d'équipements sans lesquels les idées nouvelles resteraient des idées, sans prise sur ce qu'elles visent ;*

*c) mettre en scène son travail, construire une "représentation publique" de sa recherche, de telle sorte que le public en reconnaisse la légitimité ;*

*d) enfin assurer, au sein de la communauté scientifique, l'autonomie de cette recherche, faire reconnaître sa crédibilité, la formation qui lui correspond, ses critères de validité. »<sup>4</sup>*

Ainsi les scientifiques développent des stratégies complexes pour nouer des relations avec divers acteurs et instruments dont dépend la production des connaissances et, en l'occurrence, des données de la recherche. Si cette perspective

---

<sup>1</sup> *Ibid.*, p. 50.

<sup>2</sup> Isabelle Stengers, *Au temps des catastrophes. Résister à la barbarie qui vient*, Paris, La Découverte, 2013, p. □□□.

<sup>3</sup> Dominique Pestre, Amy Dahan (dir.), *Les sciences pour la guerre*, Paris, Éditions de l'EHESS, 2004, 401 p.

<sup>4</sup> Bernadette Bensaude Vincent, « L'histoire dans l'enseignement scientifique. Un vœu pieux toujours renouvelé » dans Anne-Lise Rey (dir.), *Méthode et histoire. Quelle histoire font les historiens des sciences et des techniques*, Paris, Classiques Garnier, 2014, p. 427

d'une science *impliquée* tendait à être ignorée dans cet idéal d'autonomie et de surplomb des chercheurs, elle est aujourd'hui confrontée à une modernité réflexive qui « *n'accepte plus comme acquise "l'infaillibilité" des scientifiques comme c'était le cas au XIX<sup>e</sup> siècle* »<sup>1</sup>.

Dans la dynamique d'ouverture que connaissent les sciences depuis les années 1990, cela se traduit par une évolution de la liberté scientifique d'ouvrir ses travaux à l'avènement de politiques publiques d'ouverture des données, ce que certains parfois regrettent, peut-être en idéalisant les années 1990 : « *Les motivations profondes étaient l'échange, la mise en visibilité décidée par les chercheurs. Aujourd'hui, les orientations sont celles des mandats obligatoires, de la visibilité des financements et des institutions, de l'ouverture systématique des données sans que soit réellement posée la question de la valeur de ces données.* »<sup>2</sup> C'est en effet un mouvement dont se saisit le pouvoir politique dans le monde entier à partir des années 2000, particulièrement sur le sujet des publications scientifiques dont le modèle économique n'est pas satisfaisant. Un principe recteur apparaît : les savoirs produits sur financement public doivent être portés à la connaissance des citoyens, et les servir.

En 2004, la Commission Sciences et Technologies de la Chambre des communes du Royaume-Uni rejette le modèle de la communication savante fondé seul sur l'édition scientifique aux mains de quelques entreprises ; un modèle qui ne peut plus tenir dans le contexte de « *l'accroissement du volume des activités de recherche, l'augmentation des prix de la documentation et la stagnation des budgets des bibliothèques* »<sup>3</sup>. La même année, les Etats-Unis ouvrent le débat sur l'oligopole des éditeurs scientifiques et la Commission européenne, par la Direction générale de la recherche, initie une étude sur l'évolution économique et technique des marchés des publications scientifiques. Première et principale recommandation des résultats de l'étude : le libre accès du public aux résultats de la recherche financée par des fonds publics. Progressivement, les enjeux d'ouverture font leur chemin dans les instances publiques jusqu'à concerner la problématique de la donnée, qu'elle soit issue de la recherche ou d'ailleurs. L'*open access* et l'*open science* se font *open data* caractérisant « *les pratiques proactives de publication de données produites dans le cadre d'une mission de service public et ne contenant pas d'informations personnelles* »<sup>4</sup>.

En France, toutes les données, dont celles issues de la recherche, doivent alors être vues comme des « *infrastructures essentielles et critiques du fonctionnement de*

---

<sup>1</sup> Dominique Wolton : « Abondance et gratuité : pour quoi faire et jusqu'où ? Entretien avec Joëlle Farchy, Pascal Froissart et Cécile Méadel » dans Valérie Schafer (dir.), *op. cit.*, 2014, p. 33.

<sup>2</sup> Ghislaine Charton, « L'Open science au prisme de la Commission européenne » dans *Education et sociétés*, vol. 1, n° 41, 2018, p. 183.

<sup>3</sup> Françoise Vandooren, « Evolution de l'accès aux publications scientifiques » dans Joachim Schöpfel (dir.), *La publication scientifique : analyses et perspectives*, Paris, Lavoisier, 2008, p. 138.

<sup>4</sup> Jérôme Denis, Samuel Goëta. « Les facettes de l'Open Data : émergence, fondements et travail en coulisses », Pierre-Michel Menger, Simon Paye (dir.), *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, Paris, Collège de France, 2017, p. 121.

*l'économie et de l'Etat* » selon le Premier Ministre ; « *La maîtrise de la production de la donnée, de son utilisation et de sa valorisation relève d'enjeux que l'on peut qualifier de souverains. [...] La construction d'une [...] infrastructure adaptée nécessite ainsi la mobilisation de plusieurs leviers – financiers, contractuels, juridiques et techniques – ainsi que l'adaptation des modèles de gouvernance* »<sup>1</sup>. Le numérique joue alors un rôle pilier déjà éprouvé dans le champ de la recherche, ce dont témoigne l'Inra se faisant depuis le début des années 2010 le fer de lance de la dynamique d'ouverture et de la transition numérique : « *Le numérique transforme les processus de recherche, la production, validation, valorisation et circulation des connaissances, ainsi que les collectifs et métiers de la recherche. L'élargissement des collectifs de recherche, facilité et transformé par les outils numériques, amène les chercheurs à tisser de nouveaux liens avec la société civile marchande et non marchande.* »<sup>2</sup> Ainsi, l'Institut devenu INRAE inscrit dans son document d'orientation pour la décennie à venir son engagement auprès des acteurs scientifiques et de la société : il s'agit en effet, en toute première orientation de politique générale, de « *placer l'innovation et l'expertise au cœur de nos relations avec la société pour renforcer notre culture de l'impact* »<sup>3</sup>.

Au-delà de l'objectif de transparence de la vie publique, l'ouverture de la production des savoirs par les acteurs politiques vise en effet à renouer sciences et société, dans une perspective peut-être « citoyenne », mais avant tout d'innovation et de dynamisme économique entre la recherche publique et la « société civile marchande ». La Commission européenne acte cette visée dès 2012 en associant « recherche et innovation » dans l'organigramme de sa direction dédiée<sup>4</sup>. Si le début des années 2010 étaient toutefois davantage marquées par les enjeux de transparence, de crédibilité, d'efficacité de la science en contexte ouvert, la Commission évolue dans son discours pour faire de l'innovation une part univoque de l'ouverture des données<sup>5</sup>. Dès 1996 l'OCDE pointait toutefois la restructuration des économies autour du savoir et de l'information, parallèle à l'étude croissante par les chercheurs de l'avènement d'une économie du savoir. « *Le terme "économie fondée sur le savoir" découle de la reconnaissance grandissante de l'importance de la connaissance et de la technologie dans les économies modernes* »<sup>6</sup>, affirme ainsi l'OCDE qui, dès lors, invite les économies à intégrer cette nouvelle réalité. Un espace européen de la recherche est alors créé au début des années 2000 pour engager l'Union Européenne « *dans la société de la connaissance* » : « *C'est*

---

<sup>1</sup> Réponse du Premier Ministre du 4 mars 2019 au référé de la Cour des Comptes du 11 décembre 2018, cité par Bernard Jacquemin, Joachim Schöpfel, Renaud Fabre, « Libre accès et données de recherche. De l'utopie à l'idéal réaliste », *Études de communication*, n° 52, 2019, p. 23.

<sup>2</sup> Inra, *Transition numérique et pratiques de recherche et d'enseignement supérieur en agronomie, environnement, alimentation, sciences vétérinaires à l'horizon 2040*, Résumé de la prospective, 27 juin 2019, p. 3.

<sup>3</sup> INRAE, *INRAE2030. Partageons la science et l'innovation pour un avenir durable*, INRAE, document d'orientation, décembre 2020, p. 6.

<sup>4</sup> « Organigramme de la direction Générale de la recherche et de l'innovation », [https://ec.europa.eu/info/sites/default/files/organisation\\_charts/organisation-chart\\_dg-rtd\\_fr.pdf](https://ec.europa.eu/info/sites/default/files/organisation_charts/organisation-chart_dg-rtd_fr.pdf).

<sup>5</sup> Ghislaine Charton, « L'Open science au prisme de la Commission européenne », *op. cit.*, p. 179.

<sup>6</sup> OCDE, *L'économie fondée sur le savoir*, Paris, OCDE/GD(96)102, 1996, p. 3, [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD\(96\)102&docLanguage=Fr](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD(96)102&docLanguage=Fr).

*essentiellement sur la connaissance sous ses différentes formes, sur sa production, son acquisition et son utilisation, que repose le développement économique et social.* »<sup>1</sup> En France, le Commissariat au Plan délivre un rapport sur « La France dans l'économie du savoir » admettant que « *la principale source de création de richesses réside désormais davantage dans les savoirs et les compétences, que dans les ressources matérielles. La compétitivité des entreprises et, au-delà, celle des nations, repose fondamentalement sur les capacités à créer et à utiliser les connaissances* »<sup>2</sup>

Ces visées seront complétées de la loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche inscrivant l'objectif de favoriser les interactions entre sciences et société, puis de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique fixant le cadre de la libre réutilisation des données publiques au profit de la société ; le Plan national pour la science ouverte, enrichi depuis juillet 2021 d'un Second, défend par ailleurs le partage des savoirs qui se fait alors un des leviers essentiels de la recherche et de l'innovation<sup>3</sup>. Les organismes de recherche, quant à eux, tentent de s'inscrire dans ces dynamiques de politique publique, notamment INRAE qui rédige dès 2020 une fiche « Données, partenariat et innovation dans un contexte de science ouverte » à destination de ses chercheurs et techniciens. Ces derniers sont ainsi informés sur une demi-douzaine de pages du contexte, de leurs devoirs ainsi que de leurs droits, notamment en matière de propriété intellectuelle<sup>4</sup> ; car c'est bien en effet sous un même principe directeur que le mouvement d'ouverture en relation avec la société se déploie : « aussi ouvert que possible, aussi fermé que nécessaire ».

Bien sûr, la science et les données ouvertes ne peuvent se résumer qu'au seul motif économique. La recherche scientifique se doit d'élargir ses perspectives en développant des stratégies d'inclusion des divers acteurs de la société qui ne veulent, ni ne peuvent, rester étrangers à la production des savoirs. Le « savoir populaire » par les sciences participatives accède ainsi à une nouvelle légitimité, qui n'est pas celle d'une « *science populaire* » dont parle en 1929 Ludwik Fleck, un passage nécessaire dans le processus de création d'un fait scientifique caractérisé par « *la suppression des détails et, essentiellement, des opinions conflictuelles, ce qui produit une simplification artificielle* » pour la société devant alors se contenter d'une « *image expressive créée au moyen de simplification et d'évaluation* »<sup>5</sup>. Non, c'est une science qui réhabilite dans leur complétude et leur complexité les savoirs de

---

<sup>1</sup> Communication de la Commission au Conseil, au Parlement européen, au Comité économique et social et au Comité des régions - Vers un espace européen de la recherche, 2000, /\* COM/2000/0006 final \*/.

<sup>2</sup> Commissariat général du Plan, *La France dans l'économie du savoir : pour une dynamique collective, La Documentation française*, 2002 p. 9, cité dans Pierre Cornu, Egizio Valceschini, Odile Maeght-Bourney, *L'histoire de l'Inra, entre science et politique*, op. cit., p. 405.

<sup>3</sup> « Deuxième plan national pour la science ouverte. Généraliser la science ouverte en France. 2021-2024 », Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 32 p, [https://www.ouvrirelascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte\\_2021-2024.pdf](https://www.ouvrirelascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte_2021-2024.pdf).

<sup>4</sup> INRAE, « Données, partenariat et innovation dans un contexte de science ouverte », 2020, 6 p, [https://datapartage.inrae.fr/content/download/3747/39730/version/1/file/Fiche\\_SO\\_PI\\_Partenerariats.pdf](https://datapartage.inrae.fr/content/download/3747/39730/version/1/file/Fiche_SO_PI_Partenerariats.pdf).

<sup>5</sup> Ludwik Fleck, *Genèse et développement d'un fait scientifique*, op. cit., p. 198.

communautés usuellement exclues du régime de véridicité des sciences modernes ; par ces objets hybrides qui s'affranchissent des barrières et des dualismes, tout à chacun peut être appelé à élaborer et partager son savoir. C'est ce que la systémique agraire au sein de la recherche agronomique pratique vigoureusement depuis les années 1980 en liant dans un même système environnement, animaux, hommes, techniques ou encore territoire. L'agriculteur a en effet « *de bonnes raisons de faire ce qu'il fait* » ; son savoir est lui-même source de connaissance pour le progrès scientifique et technique<sup>1</sup>. Les apports de l'ethnologie sont alors conséquents, vivifiant les revendications du sérieux des savoirs populaires, en particulier dans la recherche agronomique. Claude Lévi-Strauss ainsi ne s'y trompait pas dès 1963 : « *Pour transformer une herbe folle en plante cultivée, une bête sauvage en animal domestique, faire apparaître chez l'une ou chez l'autre des propriétés alimentaires ou technologiques qui, à l'origine, étaient complètement absentes ou pouvaient à peine être soupçonnées ; [...] pour élaborer les techniques, souvent longues et complexes, permettant de cultiver sans terre ou bien sans eau, de changer graines ou racines toxiques en aliments [...], il a fallu, n'en doutons pas, une attitude d'esprit véritablement scientifique, une curiosité assidue et toujours en éveil, un appétit de connaître pour le plaisir de connaître car une petite fraction seulement des observations et des expériences [...] pouvaient donner des résultats pratiques, et immédiatement utilisables.* »<sup>2</sup>

C'est ainsi sans surprise que le Président-directeur général de l'Inra François Houllier en 2015 se voit confier par les ministres en charge de l'éducation, de l'enseignement supérieur et de la recherche une mission de réflexion sur les sciences participatives, débouchant sur un livret de soixante-quatre pages<sup>3</sup>. Les sciences participatives désignent ainsi « *les formes de production de connaissances scientifiques auxquelles des acteurs non-scientifiques-professionnels, qu'il s'agisse d'individus ou de groupes, participent de façon active et délibérée* »<sup>4</sup>. Dans ce cadre-là, le numérique se fait l'instrument privilégié modifiant les pratiques de partage dépassant les communautés scientifiques, mais aussi et surtout la condition *sine qua non* pour « *garantir une approche scientifique rigoureuse* »<sup>5</sup>. Les sources de production des données de la recherche se diversifient, participant, dans le mouvement d'ouverture, à enrichir les sciences à la fois dans leur régime de véridicité et dans leur relation avec la société.

Bien sûr, si la science ouverte a pour idéal l'appréhension des savoirs scientifiques par tous les citoyens, cet accès se voulant universel est fait de conditions inévitables pouvant entraver les ambitions. En effet, il ne s'agit pas seulement d'*ouvrir* pour que les savoirs soient *reçus* ; il importe, comme le souligne

---

<sup>1</sup> Jean-Pierre Deffontaines, Claudine Friedberg, « Techniques et pratiques : à la jonction du naturel et du social » dans Marcel Jollivet (dir.), *Sciences de la nature, sciences de la société. Les passeurs de frontières*, op. cit., p. 351.

<sup>2</sup> Claude Lévi-Strauss, *La Pensée sauvage*, Paris, Plon, 1963, p. 27.

<sup>3</sup> François Houllier, Jean-Baptiste Merilhou-Goudard. *Les sciences participatives en France : Etats des lieux, bonnes pratiques et recommandations*, Inra, 2016, ffhal-02801940, 63 p.

<sup>4</sup> *Ibid.*, p. 12.

<sup>5</sup> *Ibid.*, p. 36.

Dominique Wolton, que « *les récepteurs aient les capacités cognitives pour les recevoir, les interpréter, les utiliser* ». Or le numérique, en dissociant la dimension du calcul nécessaire à la lecture d'un contenu de la dimension culturelle nécessaire à son intelligibilité, dédouble la problématique des récepteurs, la portant à la fois sur l'accessibilité technique et l'intelligibilité culturelle des savoirs partagés<sup>1</sup>. C'est donc un double effort que la science ouverte appelle, tant au niveau de l'outil que de l'objet culturel.

## UNE DOUBLE DYNAMIQUE INTERNATIONALE ET NATIONALE

### La science ouverte à l'international

L'histoire des mouvements d'ouverture, étroitement liés au monde de la recherche scientifique, doit être appréhendée à l'échelle internationale pour la saisir dans toute sa complétude. Toutefois ce court panorama que nous nous apprêtons à réaliser ne prétend pas à l'exhaustivité ; il s'agit ici de restituer certains des jalons décisifs dans l'avancée d'une discussion internationale sur le sujet de l'ouverture des données et du libre accès. Si c'est d'ailleurs ce dernier qui est sur le devant de la scène politique internationale, on pourra noter une incursion significative des enjeux de la science ouverte dans le champ de la recherche en génétique au début des années 2000. En effet, le projet de séquençage du génome humain, en plus d'attirer l'attention du public, engage les acteurs de son avènement dans la voie de l'ouverture dès lors que ceux-ci s'accordent sur la nécessité de rendre librement accessibles, dès que possible, tous les résultats obtenus. « *Cette initiative est basée sur la conviction, largement partagée, que le génome humain appartient à tous et, en conséquence, sur la volonté d'éviter que des laboratoires ou entreprises ne s'approprient des droits exclusifs (enclosures) sur certains gènes ou séquences d'ADN.* »<sup>2</sup> Une première forme d'*open data* motivée par des enjeux éthiques évidents. Toutefois, en dehors de ces projets qui interrogent en profondeur l'éthique de la science, c'est le terme d'*open access* qui mobilise l'attention : né au cours des années 1990 au sein des communautés scientifiques portées par le numérique, il est appréhendé durant les années 2000 par différentes instances internationales pour finalement ouvrir sur une réflexion globale de l'ouverture des données et des documents : de l'*open access* à l'*open government*.

La problématique de l'accès aux publications scientifiques constitue ainsi le premier jalon de la dynamique d'ouverture. Le dépôt des prépublications dans la base numérique *arXiv*, initié par le physicien américain Paul Ginsparg au début des années 1990, inaugure un autre rapport au partage de l'information scientifique<sup>3</sup> et suscite l'émergence de perspectives pour la publication que d'aucuns qualifient de

---

<sup>1</sup> Bruno Bachimont, *Patrimoine et numérique. Technique et politique de la mémoire*, Paris, 2017, INA, 246 p.

<sup>2</sup> Comité Consultatif Commun d'Éthique Pour La Recherche Agronomique, *Avis 8 sur les enjeux éthiques et déontologique du partage et de la gestion des données issues de la recherche*, Inra, 2016, fihal-02796585, p. 11.

<sup>3</sup> Bernard Rentier, *Science ouverte, le défi de la transparence*, Bruxelles, Académie royale de Belgique, 2018, p. 44.

subversives<sup>1</sup>. Le champ de la recherche en physique et en astronomie constitue un terrain fertile pour les idées d'ouverture : ces communautés scientifiques étaient en effet habituées depuis plusieurs années à la mise au point d'outils d'échange, à la notion d'interopérabilité des données, à celle de standardisation des techniques facilitant la production de résultats<sup>2</sup>. En Europe, des initiatives similaires ont été tentées, notamment le projet *E-BioSci*, sur le modèle de l'entrepôt américain *E-Biomed* dans le cadre de l'*European Molecular Biology Organization*. Ce projet obtient en septembre 2001 une subvention de l'Union européenne pour un développement sur trois ans qui toutefois n'aboutira pas à l'effet escompté<sup>3</sup>.

Les années suivantes, l'*open access* prend alors une envergure nouvelle sous l'impulsion de trois grandes déclarations publiques, la « *définition BBB* » pour Budapest, Bethesda et Berlin<sup>4</sup>. En effet, l'Initiative de Budapest du 14 février 2002, la Déclaration de Bethesda du 20 juin 2003 et la Déclaration de Berlin du 22 octobre 2003, sont autant de moments qui marquent un tournant de l'ouverture à l'échelle internationale. L'Initiative de Budapest propose le déploiement de l'accès ouvert pour la littérature scientifique, c'est-à-dire « *sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet* »<sup>5</sup>. Le projet est d'ores et déjà très ambitieux, et surtout installe le numérique et les réseaux comme garantie de la réussite de ces ambitions. Le texte présente toutefois un rapport du chercheur à la production des savoirs très épuré et traditionnel dans son appréhension par l'histoire des sciences, décrivant des scientifiques désireux de « *publier sans rétribution les fruits de leur recherche dans des revues savantes, pour l'amour de la recherche et de la connaissance* »<sup>6</sup>. Cette Initiative met au jour l'*Open Archives Initiative*, à l'origine du développement de la norme *Open Archives Initiatives Protocol for Metadata Harvesting* (OAI-PMH) en juillet 2001<sup>7</sup> et dont l'adoption, aujourd'hui généralisée, permet l'échange de métadonnées entre Archives ouvertes. Les Déclarations suivantes de Bethesda et de Berlin, auxquelles le CNRS, le CERN ou encore l'Inria ont affiché immédiatement leur adhésion, prolongent le texte de Budapest et précisent ce que le terme d'*open access* recouvre en insistant davantage

---

<sup>1</sup> Steven Harnad, « A Subversive Proposal » dans A. Okerson, J. O'Donnell (dir.), *Scholarly Journals at the Crossroads : a Subversive Proposal for Electronic Publishing*, Washington DC, Association of Research Libraries, 1995.

<sup>2</sup> Françoise Genova, et al., « Building a Disciplinary, World-Wide Data Infrastructure », *Data Science Journal*, 16, 2017, DOI: <http://doi.org/10.5334/dsj-2017-016>.

<sup>3</sup> Hélène Bosc, « Archives ouvertes : quinze ans d'histoire », C. Aubry, J. Janik, *Les Archives Ouvertes : enjeux et pratiques. Guide à l'usage des professionnels de l'information*, Paris, ADBS, 2005, p.

<sup>4</sup> Marin Dacos, « 1. Une définition », Peter Suber, *Qu'est-ce que l'accès ouvert ?*, Marseille, OpenEdition Press, 2016, [En ligne], 10.4000/books.oep.1605.

<sup>5</sup> « Initiative de Budapest pour l'Accès Ouvert », accessible sur [ouvrirlascience.fr](http://ouvrirlascience.fr) : [https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/Initiative-Budapest\\_fr.pdf](https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/Initiative-Budapest_fr.pdf), p. 1.

<sup>6</sup> « Déclaration de Berlin sur le libre accès à la connaissance en science exactes, sciences de la vie, sciences humaines et sociales », 22 octobre 2003, Open Access Max-Planck-Gesellschaft, <https://openaccess.mpg.de/Berliner-Erklaerung>.

<sup>7</sup> Hélène Bosc, « Archives ouvertes : quinze ans d'histoire », *op. cit.*, p. 26.

encore sur la suppression des barrières de partage des publications, que ce soit en termes légaux ou économiques, mais aussi sur toute autre forme de savoirs permettant l'accès à la « *connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales* »<sup>1</sup>. Ainsi, « *les contributions au libre accès se composent de résultats originaux de recherches scientifiques, de données brutes et de métadonnées, de documents sources, de représentations numériques de documents picturaux et graphiques, de documents scientifiques multimédia* »<sup>2</sup>. L'accès aux données se fait dès lors une problématique qui déborde des seules considérations scientifiques : des instances politiques internationales se font les relais des attentes de la société à ce sujet.

La même année, en 2004, l'OCDE s'empare du sujet par le Comité de la politique scientifique et technologique après que les ministres de la recherche des pays membres, ainsi que ceux de la Chine, de la Russie, ou encore d'Israël, réalisent une « Déclaration ministérielle sur l'accès aux données de la recherche financée par des fonds publics » qui invite l'organisation à penser un ensemble de lignes directrices « *fondées sur un ensemble convenu de principes destinés à faciliter l'accès dans des conditions optimales d'efficacité par rapport aux coûts des données numériques de la recherche* »<sup>3</sup>. En 2007, l'OCDE publie ainsi son document *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*<sup>4</sup> définissant un ensemble de principes dont l'ouverture, la transparence, l'interopérabilité, la qualité, la pérennité pour les données de la recherche, notamment financées sur fonds publics.

L'Union Européenne n'est évidemment pas en reste puisque le Conseil scientifique du Conseil Européen pour la recherche propose dès 2006 sa propre déclaration pour l'accès ouvert, insistant sur les enjeux d'attractivité des « *politiques imposant la mise à disposition des résultats de la recherche au public, au sein d'entrepôts en accès ouvert, dans un délai raisonnable après la publication* »<sup>5</sup>. Cette première initiative est alors suivie d'autres durant la fin de la décennie jusqu'à la recommandation de la Commission européenne du 17 juillet 2012 sur l'accès et la préservation de l'information scientifique, appelant à la définition d'une politique nationale pour chaque pays membre avec comme objectif un libre accès largement majoritaire à l'année 2020<sup>6</sup>. Ce calendrier, très rythmé, s'ajuste alors au grand programme européen Horizon 2020 concentrant ses financements sur les enjeux de

---

<sup>1</sup> *Ibid.*, p. 1.

<sup>2</sup> « *Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.* ». *Ibid.*, p. 1.

<sup>3</sup> « Déclaration ministérielle sur l'accès aux données de la recherche financée par des fonds publics », le 30 janvier 2004, consultable sur [ouvrirlascience.fr](https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/D%C3%A9claration-OCDE_fr.pdf) : [https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/D%C3%A9claration-OCDE\\_fr.pdf](https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/D%C3%A9claration-OCDE_fr.pdf), p. 3.

<sup>4</sup> OCDE, *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, 2007, 49 p.

<sup>5</sup> Conseil scientifique de l'European Research Council, « ERC Scientific Council Statement on Open Access », Décembre 2006, [https://erc.europa.eu/sites/default/files/press\\_release/files/erc\\_scc\\_statement\\_2006\\_open\\_access\\_0.pdf](https://erc.europa.eu/sites/default/files/press_release/files/erc_scc_statement_2006_open_access_0.pdf).

<sup>6</sup> « Recommandations de la Commission européenne du 17 juillet 2012 sur l'accès et la préservation de l'information scientifique », <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012H0417&rid=1>.

l'excellence scientifique, de l'innovation industrielle et des rapports sciences et société. Son démarrage au début de l'année 2014 est précédé, pour la problématique du partage des données de la recherche et du libre accès des publications scientifiques, d'un ensemble d'études, notamment celle du projet RECODE dont les conclusions publiées en octobre 2013 identifient les problèmes majeurs dans l'adoption de politiques de libre accès et présentent des recommandations quant à leur mise en place<sup>1</sup>. La publication des *Lignes directrices pour l'accès ouvert des publications scientifiques et des données de la recherche à l'horizon 2020*<sup>2</sup> en décembre 2013 inaugure le programme Horizon 2020 par le lancement des premières étapes des politiques pour le libre accès.

Parallèlement, les réflexions sur les processus mêmes de la recherche scientifique qu'inspire l'*open access* se multiplient, telle la *San Francisco Declaration on Research Assessment (DORA)* sur l'évaluation de la recherche, ses conditions d'avènement et sa méthodologie, à laquelle participe notamment l'Inra<sup>3</sup>. Cette Déclaration insiste sur les enjeux à la fois scientifiques, économiques, ainsi que sociétaux qu'implique l'élaboration d'une évaluation de la recherche de qualité. Une qualité qui dépend en outre de celle des données des résultats évalués conduisant un groupe de chercheurs à développer un ensemble de principes directeurs et adaptables à presque tous les contextes de recherche pour l'établissement des données de la recherche<sup>4</sup>. Ce sont les principes FAIR qui permettent, par ces ambitions de rendre les données trouvables (*Findable*), accessibles (*Accessible*), interopérables (*Interoperable*) et réutilisables (*Reusable*), d'établir des bases communes normées de création et de gestion des données.

Les années 2010 sont ainsi porteuses de nombreuses initiatives qui, plutôt de l'ordre de la recommandation sinon de l'incitation au départ, deviennent à la fin de la décennie de plus en plus coercitives. Les éléments notables de ces actions sont entre autres l'obligation de la rédaction d'un Plan de gestion des données (PGD) que la Commission européenne impose en 2013 pour toute recherche financée sur fonds européens. Ces exigences se répercutent à l'échelle des agences de financement nationales ; les effets de cette incitation majeure vont se faire de plus en plus visibles<sup>5</sup>. Par ailleurs se multiplient les infrastructures de recherche, outils majeurs des mouvements d'ouverture<sup>6</sup> dont l'organisation nécessite des moyens importants œuvrant au soutien des mouvements d'ouverture. Les infrastructures de recherche

---

<sup>1</sup> « Policy Recommendations for Open access to research Data in Europe », Projet FP7 RECODE, file:///C:/Users/Maison/Downloads/recode-d5\_1-policy-recommendations-\_final\_9958.pdf.

<sup>2</sup> « Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 », décembre 2013 pour la première version, [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf).

<sup>3</sup> San Francisco Declaration on Research Assessment (DORA), consultable sur [ouvrirlascience.fr](http://ouvrirlascience.fr), [ouvrirlascience.fr/wp-content/uploads/2018/12/DORA\\_French\\_V2.pdf](http://ouvrirlascience.fr/wp-content/uploads/2018/12/DORA_French_V2.pdf).

<sup>4</sup> Mark D. Wilkinson, Michel Dumontier, *et al.*, « The FAIR Guiding Principles for scientific data management and stewardship », *Scientific Data*, 2016, doi: 10.1038/sdata.2016.18.

<sup>5</sup> Françoise Genova, *et al.*, « Building a Disciplinary, World-Wide Data Infrastructure », *Data Science Journal*, 16, 2017, DOI: <http://doi.org/10.5334/dsj-2017-016>.

<sup>6</sup> Françoise Genova, « Vision internationale autour des données de recherche », *op. cit.*, p. 22.

sont « *des installations qui fournissent des ressources et des services aux communautés de recherche pour mener des recherches et favoriser l'innovation* »<sup>1</sup>. Ce sont donc des dispositifs complets comprenant notamment des « *équipements scientifiques majeurs ou ensembles d'instruments* », des « *collections, archives ou données scientifiques* » ou encore des « *systèmes informatiques et réseaux de communication* ». Ainsi citerons-nous l'*European Open Science Cloud* (EOSC), projet ambitieux d'un espace commun des données auquel participe INRAE, nous y reviendrons, ainsi que l'*European Strategic Forum for Research Infrastructure* (ESFRI) qui cadre des infrastructures des données porteuses pour de nombreuses disciplines scientifiques.

Enfin, à la fin de la décennie 2010, c'est le Plan S à l'échelle européenne, signé par onze agences de financements européennes dont l'Agence nationale de la recherche (ANR), qui impose un nouveau rythme avec l'objectif du libre accès de toutes les publications scientifiques résultant de recherches réalisées sur fonds publics nationaux ou européens. Le chemin parcouru en une décennie est ainsi gigantesque, ces dernières années étant marquées par une intériorisation croissante des enjeux de l'ouverture dans le champ de la recherche scientifique dont les communautés sont de plus en plus insérées dans le village mondial par les outils numériques et réceptifs à la dynamique d'ouverture.

## **Le mouvement en France**

La France n'est évidemment pas en reste dans les mouvements d'ouverture, s'associant aux dynamiques internationales plurielles parcourues précédemment tout en initiant ses propres programmes. Dès la fin des années 1990, certains chercheurs français enviaient les dépôts d'archives créés aux Etats-Unis, dont l'initiative *arXiv*. Le physicien Franck Laloë, à l'instar de Paul Ginsparg, est ainsi à l'origine du développement de la plateforme HAL dès 1999 dont le modèle proposait « *davantage de garantie dans le temps* »<sup>2</sup>. Le Centre pour la communication scientifique directe (CCSD) est alors créé afin notamment de prendre en charge le développement de la plateforme mais aussi, au cours des deux dernières décennies, de jouer sa part dans les mouvements d'ouverture par de nouveaux outils accompagnant l'Archive ouverte HAL. Des plateformes web sont ainsi mises à la disposition des communautés de chercheurs en soutien à l'activité scientifique, telles *SciencesConf.org*, pour l'organisation d'événements scientifiques par la mobilisation de réseaux, et *episciences.org*, pour la promotion d'un nouveau modèle de publication scientifique fondé sur les épi-revues renouvelant les approches en matière de diffusion et de contrôle des savoirs. L'archive ouverte HAL reste toutefois le projet phare, autour duquel les communautés scientifiques dans un

---

<sup>1</sup> « Research Infrastructures are facilities that provide resources and services for research communities to conduct research and foster innovation. », "What are Research infrastructures", European Research Infrastructures, [https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures_en)

<sup>2</sup> Jean-Baptiste Veyrieras, « Science ouverte : la révolution nécessaire », *CNRS Le Journal*, 8 janvier 2020, [En ligne], <https://lejournel.cnrs.fr/articles/science-ouverte-la-revolution-necessaire>.

premier temps se mobilisèrent peu, jusqu'à la « Convention de partenariat en faveur des archives ouvertes et de la plateforme mutualisée HAL », signée en 2013 à l'Académie des Sciences par un grand nombre d'institutions de recherche et d'organisme d'ESR actant la dynamique politique de l'*open access* engagée par la recherche scientifique française. HAL se fait ainsi une des plateformes de référence en s'inscrivant désormais comme une des « *infrastructures de recherche indispensables à la pratique scientifique* »<sup>1</sup>, au sens européen du terme, aux côtés de Collex-Persée, ou encore d'OpenEdition. En 2018, HAL est placé sous l'égide du Comité pour la Science Ouverte (CoSO) prenant le relai de la Bibliothèque Scientifique Numérique (BSN).

La BSN, créée en 2009, était une infrastructure de recherche au service de la recherche et des chercheurs comme outil d'appui pour les professionnels de l'IST. « *Dès le départ, raconte Stéphanie Groudiev alors coordinatrice du projet en 2013, le levier d'action de la BSN est posé comme la mise en place d'un pilotage politique partagé par les grands acteurs de l'enseignement supérieur et de la recherche, pour offrir des services à l'ensemble des chercheurs quels que soient leur communauté et leur statut, et créer de nouveaux modèles et équilibres économiques entre acteurs publics et privés dans l'édition scientifique.* »<sup>2</sup> Le projet BSN se compose de neuf segments de pilotage de cet appui à la recherche dont l'édition scientifique principalement, mais aussi l'Archive ouverte HAL, l'archivage pérenne par le CINES, et, ajouté ultérieurement comme dixième segment, les données de la recherche. La BSN réunissait alors de nombreux acteurs de la scène de la recherche, dont le CNRS, mais aussi l'Inra et l'Irstea. En 2018, la BSN devient le Comité pour la Science Ouverte explicitant clairement les orientations globales de la recherche scientifique et de l'IST qui ne peuvent plus se penser sans l'outil numérique en environnement ouvert. Le CoSO a dès lors pour missions de « *proposer des orientations et d'instruire des sujets sur les questions de Science ouverte, d'impulser et d'accompagner les actions associées* »<sup>3</sup> grâce notamment à l'organisation de quatre collèges (Publications, Données de la recherche, Compétences et formation, Europe et international). Cette action s'inscrit dans le sillon de la Déclaration *Open Government Partnership* de 2011 – que la France signe en 2014 – visant à « *rendre les gouvernements plus transparents, plus réactifs, plus responsables et plus efficace* »<sup>4</sup>. Surtout, si le CoSO ambitionne de développer les compétences de la science ouverte dans les communautés scientifiques, d'en coordonner l'action aux échelles nationale, européenne et internationale, de participer à la définition des principes et des orientations de la publication scientifique avec les principaux éditeurs, il entend aussi « *assurer la mise en œuvre coordonnée, avec les opérateurs*

---

<sup>1</sup> « Stratégie nationale des infrastructures de recherche », Ministère de l'enseignement supérieur, de la recherche et de l'innovation, Edition 2018, n°2, p. 150.

<sup>2</sup> Stéphanie Groudiev, « La bibliothèque scientifique numérique : un cadre politique de coordination des actions en faveur de l'IST », *Bulletin des bibliothèques de France*, t. 58, n° 1, 2013, p. 63.

<sup>3</sup> « Evolution de la BSN vers le Comité pour la Science Ouverte », [ouvrirlascience.fr](http://ouvrirlascience.fr), le 26 janvier 2018, [ouvrirlascience.fr/evolution-de-la-bsn-vers-le-comite-pour-la-science-ouverte-coso/](http://ouvrirlascience.fr/evolution-de-la-bsn-vers-le-comite-pour-la-science-ouverte-coso/).

<sup>4</sup> « Déclaration du gouvernement ouvert », *Open Government Partnership*, <https://www.opengovpartnership.org/fr/process/joining-ogp/open-government-declaration/>.

*d'enseignement supérieur et de recherche, d'un plan national visant à ouvrir à tous les publications et les données de la recherche »<sup>1</sup>.*

Ce plan est alors initié en juillet 2018 par la ministre de l'Enseignement supérieur, de la Recherche et de l'Innovation Frédérique Vidal. S'inscrivant ouvertement dans la suite de la Déclaration *Open Government Partnership* et de l'appel d'Amsterdam *Call for Action on Open Science* mobilisant les Etats européens, le Plan national pour la science ouverte doté d'un budget de 5,4 millions d'euros pour la première année est conçu en trois axes majeurs : généraliser l'accès ouvert aux publications, structurer et ouvrir les données de la recherche, s'inscrire dans une dynamique durable, européenne et internationale. Concernant le second axe, l'ambition était de rendre les données de la recherche produites conformes aux principes FAIR en participant à la formation des acteurs de la recherche et la diffusion des bonnes pratiques de structuration des données<sup>2</sup>. Ce plan national est d'ailleurs une opportunité pour la France de s'inscrire institutionnellement dans les programmes internationaux pour une harmonisation large des enjeux de gestion et d'ouverture des données de la recherche. La France participe en effet à la *Research Data Alliance*, réseau international promouvant ces bonnes pratiques, justifiant alors le troisième axe d'inscription dans une dynamique internationale.

De premiers jalons ont été posés lors de ces trois dernières années dans le sillon du Plan national pour la science ouverte, telles la création ou l'affermissement de politiques de science ouverte au sein des établissements d'ESR et des organismes de recherche, l'élaboration d'une documentation large et institutionnelle de guides et de recommandations à destination des acteurs de la recherche publique, ou encore la création d'un Fonds national pour la science ouverte en juillet 2019 doté de 3,1 millions d'euros, dont un million d'euros provenant des économies réalisées sur l'éditeur Elsevier selon Marin Dacos, conseiller scientifique pour la science ouverte auprès du Directeur général de la recherche et de l'innovation au Ministère de l'enseignement supérieur, de la recherche et de l'innovation<sup>3</sup>. Le reste est d'ailleurs dédié à l'accroissement de la biodiversité en environnement scientifique et numérique.

Ce premier plan est finalement suivi d'un second annoncé en juillet 2021<sup>4</sup>, invitant à renforcer le mouvement d'ouverture qui s'inscrit progressivement dans les pratiques des chercheurs. Plus complet et ambitieux que le premier, ce Deuxième Plan national pour la science ouverte se structure en quatre axes couvrant les enjeux habituels de la publication scientifique et des données de la recherche en science ouverte, mais aussi la question des codes sources produits par la recherche, et donc

---

<sup>1</sup> « Le comité pour la science ouverte », [ouvrirelascience.fr](https://www.ouvrirelascience.fr/presentation-du-comite/), <https://www.ouvrirelascience.fr/presentation-du-comite/>.

<sup>2</sup> « Premier Plan national pour la science ouverte », Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 4 juillet 2018, 12 p.

<sup>3</sup> Marin Dacos, « Le plan national pour la science ouverte. Où en sommes-nous ? », *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>4</sup> « Deuxième plan national pour la science ouverte. Généraliser la science ouverte en France. 2021-2024 », Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, [https://www.ouvrirelascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte\\_2021-2024.pdf](https://www.ouvrirelascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte_2021-2024.pdf), 32 p.

la question de l'*open source*, ainsi que la problématique des compétences des acteurs de la recherche en science ouverte appelée à devenir un paradigme épistémologique majeur. Ce second Plan conduit en outre à la création d'une plateforme nationale des données de la recherche, *Recherche Data Gouv*, qui engage clairement la France sur la dynamique d'une harmonisation nationale de ses ambitions en science ouverte. Enfin il s'inscrit dans le sillon des efforts internationaux, et notamment européens, dans les engagements d'ouverture auxquels répond l'action publique française.

Ces deux plans nationaux reposent sur les efforts de nombreux acteurs, et notamment des agences de financement qui à partir du milieu des années 2000 jouent un rôle prépondérant dans la recherche scientifique française. Leur pouvoir incitatif puis, dans un second temps, coercitif participe du mouvement d'ouverture depuis le début des années 2010. L'Agence nationale de la recherche (ANR), créée en février 2005 en contexte de restructuration politique de la recherche française, joue en effet un rôle dans le développement d'une culture de l'ouverture au sein des projets de recherche. Dans le cadre de ces Plans nationaux, ses appels *Flash* questionnent les besoins des chercheurs « *face aux enjeux de la structuration, de l'accessibilité, de la réutilisation, de l'interopérabilité, de la citation, du partage et de l'ouverture des données de la recherche* »<sup>1</sup>. Signataire en 2013 de la « Convention de partenariat en faveur des archives ouvertes et de la plateforme mutualisée HAL », elle s'appuie sur la loi « Pour une République numérique » de 2016 pour demander aux chercheurs que toutes les publications issues de recherches financées en majorité sur fonds publics ou européens soient déposées en texte intégral dans une archive ouverte, HAL ou institutionnelle ; l'embargo de dépôt est alors de six mois pour les sciences de la nature et du vivant et de douze pour les sciences humaines et sociales. D'ailleurs, l'ANR recommande « *de privilégier la publication au sein de revues ou d'ouvrages nativement en accès ouvert* »<sup>2</sup>.

Si l'*open access* pour les publications est déjà activement travaillé depuis deux décennies, la question de l'ouverture des données de la recherche se pose toujours. L'ANR a ainsi fait le choix, dans la continuité des recommandations européennes et internationales, de poser comme condition *sine qua non* au financement d'un projet de recherche l'élaboration de plans de gestion de données dont le modèle peut être défini en relation avec le Comité pour la science ouverte sur la base des recommandations de la structure *Science Europe*<sup>3</sup>. Si les publications scientifiques en environnement ouvert interrogent leurs modalités juridiques de diffusion relatives à la propriété intellectuelle, les données de la recherche, elles, élargissent les questionnements en lien avec le secret industriel, la protection du patrimoine scientifique et technologique, le Règlement général pour la protection des données (RGPD) ainsi que la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. La problématique de la gestion des données de la recherche questionne

---

<sup>1</sup> Thierry Damerval, « Les engagements pour la science ouverte de l'ANR », *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>2</sup> *Ibid.*

<sup>3</sup> Science Europe, *Practical Guide to the International Alignment of Research Data Management*, 2021, 29 p.

donc le cadre légal de leur ouverture qui se pose le plus souvent en ces termes : aussi ouvert que possible, aussi fermé que nécessaire.

## Le cadre juridique

Les dynamiques d'ouverture internationales, européennes et françaises parcourues précédemment sont des cadres particulièrement porteurs pour l'avènement de nouvelles pratiques dans la recherche publique à même de répondre aux enjeux pluriels auxquels elle est confrontée. Ces dynamiques appellent toutefois à la construction d'un cadre juridique structuré en mesure d'apporter des réponses d'ordre légal aux communautés scientifiques sur l'arête entre ancien et nouveau monde de la recherche. Si le sujet des publications scientifiques fut rapidement cerné en termes juridiques dans l'environnement de l'*open access*, particulièrement en France où la réglementation sur le droit d'auteur est forte, les données de la recherche, elles, pâtiennent d'une part de leur difficile définition, d'autre part de cette distinction juridique avec l'écrit scientifique qui les fait objet public : les données de la recherche « *n'appartiennent pas aux chercheurs mais aux financeurs* »<sup>1</sup>. Leur gestion s'inscrit donc dans une réflexion à la fois particulière et globale, de l'*open science* à l'*open government* et donc par le prisme de l'*open data*.

Débordant bien largement du champ de la recherche, l'*open data* prend de l'envergure depuis les années 2000, sous l'impulsion de nombreuses initiatives plurielles, économiques, sociales, citoyennes investissant les outils du numérique et le web<sup>2</sup> mais aussi en termes politiques. Le Président des Etats-Unis Barack Obama marque les esprits, lui qui, au lendemain de son investiture le 20 janvier 2009, publie un *memorandum* sur la transparence de l'administration et l'*open government*<sup>3</sup>. En France, le sujet émerge en tant que tel au début des années 2010 faisant de l'*open data* « *une exigence et une opportunité* »<sup>4</sup>. La loi n° 78-753 du 17 juillet 1978, dite loi CADA, organisait d'ores et déjà l'accès aux documents administratifs, mais ces derniers recouvraient « *les documents achevés produits ou reçus, dans le cadre de leur mission de service public* »<sup>5</sup> et, par ailleurs, les établissements d'ESR se plaçaient sous régime dérogatoire relativement aux règles de communicabilité des données produites. Et c'est là l'une des difficultés : la définition des modalités juridiques de gestion des données de la recherche. Si aujourd'hui, certains affirment que leur statut juridique « *est relativement clair* »<sup>6</sup> à l'instar de Lionel Maurel, juriste et directeur adjoint scientifique à l'Institut des sciences humaines et sociales

---

<sup>1</sup> Marin Dacos, « Le plan national pour la science ouverte. Où en sommes-nous ? », *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>2</sup> *Legicom*, n° 58, vol. 1, « Open data : une révolution en marche », 2016, 128 p.

<sup>3</sup> « Memorandum For The Heads Of Executive Departments and Agencies », 21 janvier 2009, <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>.

<sup>4</sup> Gaëtan Gorce, François Pillot, *La protection des données personnelles dans l'open data : une exigence et une opportunité, Rapport d'information fait au nom de la commission des lois*, n° 469 (2013-2014) - 16 avril 2014.

<sup>5</sup> Céline Castets-Renard, Nathalie Gandon, « Open data des données de la recherche publique : entre réformes législatives et retour d'expérience sur un guide pratique à destination des chercheurs », *LEGICOM*, vol. 56, n° 1, 2016, p. 68.

<sup>6</sup> Lionel Maurin, « Introduction aux aspects juridiques de la production scientifique », *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019, p. 32.

du CNRS, il demeure pourtant pour bon nombre d'acteurs de la recherche et même de la vie publique difficilement cernable.

Plusieurs jalons au cours de deux dernières décennies ont permis de travailler la question de la donnée de la recherche en termes juridiques : la directive 2003/98/CE du Parlement européen aborde la question de la réutilisation des informations du secteur public et propose une approche finalement transposée dans le droit français en 2005 : les enjeux de l'*open data* sont alors posés en se fondant à la fois sur l'ouverture des données au public, mais aussi et surtout leur réutilisation pour une interaction économiquement – et démocratiquement – dynamique entre différents pans de la société. Il faudra toutefois attendre les années 2010 pour voir une réflexion juridique aboutir sur l'ouverture des données en France avec, dans un premier temps, la loi Valter en 2015 modifiant la loi CADA dans l'objectif de valoriser la mise à disposition des documents publics achevés et surtout leur utilisation libre et gratuite. Certains organismes désignés (mais aucun établissement d'ESR) échappaient toutefois par décret à ces dispositions si plus de 25% de leurs ressources provenaient de la vente de données<sup>1</sup>. La Cour des comptes adressa d'ailleurs un référé au Premier Ministre en décembre 2018 après avoir constaté « *des difficultés récurrentes et multiples pour se conformer au droit* »<sup>2</sup> de la part de certains organismes tels Météo France et IGN. Elle recommande alors de « *clarifier la doctrine et les conditions d'application des règles relatives à l'ouverture des données et des codes sources des logiciels, ainsi que celles afférant à la gestion des licences* » et de « *redéfinir les modèles économiques des opérateurs en tirant les conséquences de l'ouverture des données publiques et de l'attrition des ressources propres correspondantes* »<sup>3</sup> Il s'agit de le faire en relation avec la réglementation adéquate, à savoir la loi n° 2016-1321 du 7 octobre 2016, dite pour une République numérique, qui depuis 2016 engage des modifications significatives pour la gestion des données publiques dont, dans le monde de la recherche, les données correspondantes.

Dès 2013, le Code de la recherche, établi en 2004, avait acté la mutation plurielle du paradigme de la recherche publique du XXI<sup>e</sup> siècle confrontée aux nouvelles problématiques scientifiques, environnementales, politiques ou encore démocratiques ; la loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche modifiait en effet le Code de la recherche afin de redéfinir en ces termes les missions de la recherche publique :

« *La recherche publique a pour objectifs :*

a) *Le développement et le progrès de la recherche dans tous les domaines de la connaissance ;*

---

<sup>1</sup> Lionel Maurin, « La réutilisation des données de la recherche après la loi pour une République numérique », décembre 2017, [En ligne], <https://hal.archives-ouvertes.fr/hal-01908766/document>.

<sup>2</sup> Bernard Jacquemin, Joachim Schöpfel et Renaud Fabre, « Libre accès et données de recherche. De l'utopie à l'idéal réaliste », *Études de communication* [En ligne], 52 | 2019, p. 22

<sup>3</sup> Didier Migaud, « Référé S2018-3287 », Référé de la Cour des comptes du 11 mars 2019 sur la valorisation des données de l'IGN, de Météo France et du CEREMA, mars 2019, p. 5.

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 contraires, rendus publics et accessibles. »<sup>1</sup>*

Dans cet état d'esprit, la loi n° 2013-660 du 22 juillet 2013 ajoutait en outre au Code de la recherche un Livre V sur « *la valorisation des résultats de la recherche et le transfert de technologie en direction du monde économique et de la société civile, ainsi que des associations et fondations reconnues d'utilité publique* »<sup>2</sup>.

En 2016, la loi pour une République numérique vient modifier certaines dispositions du Code de la recherche dont l'article L533-4 de ce Livre V<sup>3</sup> (article 30 de la loi pour une République numérique) qui, en permettant le droit pour l'auteur d'un « *écrit scientifique* » de mettre à disposition en version numérique de manière gratuite et libre, la version finale et acceptée pour publication de son « *manuscrit* », s'inscrit clairement dans une logique d'*open access*. C'est une « *consécration légale* »<sup>4</sup> légitimant, par le droit, les efforts d'ouverture des communautés scientifiques au-delà de l'édition scientifique traditionnelle éreintante. L'article mentionne également les données de la recherche qui, caractérisées par leur mode de financement pour moitié public et déliées d'« *un droit spécifique ou d'une réglementation particulière* », sont de fait librement réutilisables dès lors qu'elles

---

<sup>1</sup> Article L112-1 du Code de la recherche, <https://www.legifrance.gouv.fr/codes/id/LEGIARTI000027747800/2013-07-24>.

<sup>2</sup> Loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000027735009/>.

<sup>3</sup> Article L533-4 du Code de la recherche, création loi n°2016-1321 du 7 octobre 2016 - art. 30, [https://www.legifrance.gouv.fr/codes/article\\_lc/LEGIARTI000033205794/](https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033205794/).

<sup>4</sup> Direction de l'Information Scientifique et Technique du CNRS, « La consécration légale d'un droit à l'*open access* » dans *Une Science ouverte dans une République numérique — Guide stratégique : Études et propositions en vue de l'application de la loi*, Marseille, OpenEdition Press. 2017, doi :10.4000/books.oep.1713.

ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche. Ainsi, à l'instar de toute donnée produite par une administration publique, les données de la recherche sont soumises à un principe d'ouverture par défaut et de libre réutilisation, signifiant « *que les administrations qui produisent des données doivent spontanément les mettre en ligne et les rendre librement et gratuitement réutilisables, à toutes fins, y compris les fins commerciales. Si un tiers connaît l'existence de données de recherche et exprime une demande de réutilisation, il n'est pas possible de lui opposer un refus, à moins de se trouver où s'applique une exception prévue par la loi.* »<sup>1</sup> INRAE à ce titre présente clairement le cadre juridique des données de la recherche<sup>2</sup> qu'il associe à des « documents administratifs », relevant dès lors du Code des relations entre le public et l'administration entré en vigueur le 1<sup>er</sup> janvier 2016. Ainsi l'ouverture des données, telles qu'elles sont définies par l'article L 311-2 CRPA, advient si celles-ci proviennent d'une mission de service public et sont dites « achevées ». Une donnée est achevée pour INRAE lorsque son traitement « *est finalisé et que l'on s'est assuré de la complétude des métadonnées et des documents liés à sa traçabilité* »<sup>3</sup> ; elle peut alors être partagée sous réserve de conditions.

En effet, si le droit d'auteur permet de donner un cadre strict à l'écrit scientifique, la libre réutilisation des données de la recherche se caractérise toutefois par un certain nombre d'exceptions nécessaires que spécifie la formule « aussi ouvert que possible, aussi fermé que nécessaire ». Aux côtés de contraintes techniques dans la gestion des mégadonnées et de contraintes économiques dans leur mise à disposition pour des acteurs économiques à même de créer de la valeur supplémentaire, existent donc des contraintes juridiques dans les modèles devant être repensées pour nouvellement associer respect du droit d'auteur, principe de créativité, rémunération juste des créateur et notion de bien commun des objets scientifiques<sup>4</sup>. La gestion réglementée des données de la recherche impose une démarche « *de documentation et de diffusion des données* »<sup>5</sup> qui n'est pas toujours envisageable dans tous les contextes relativement à la loi française. Ainsi peut-il exister un droit de propriété intellectuelle d'un tiers, et notamment un droit d'auteur. Des données de la recherche produites lors d'un partenariat avec une structure privée par cette dernière ne peuvent être librement communicables selon les modalités de l'article 30 de la loi pour une République numérique. Le droit d'auteur d'ailleurs s'applique déjà pour la publication scientifique qui n'est donc pas soumis à un régime d'ouverture par défaut ; elle est en réalité liée à son mode de financement et soumise à un embargo différencié selon les disciplines, à savoir six mois pour les

---

<sup>1</sup> Lionel Maurin, « Introduction aux aspects juridiques de la production scientifique », *op. cit.*, p. 31.

<sup>2</sup> INRAE, « Données de la recherche et science ouverte : le principe d'un accès "aussi ouvert que possible, aussi fermé que nécessaire" », Fiche technique, DipSO, INRAE, [En ligne], 7 p., [https://datapartage.inrae.fr/content/download/3746/39727/version/1/file/Fiche\\_SO\\_Juridique\\_ouverture\\_vf.pdf](https://datapartage.inrae.fr/content/download/3746/39727/version/1/file/Fiche_SO_Juridique_ouverture_vf.pdf).

<sup>3</sup> *Ibid.*, p. 2.

<sup>4</sup> Inra, *Transition numérique et pratiques de recherche et d'enseignement supérieur en agronomie, environnement, alimentation, sciences vétérinaires à l'horizon 2040*, Résumé de la prospective, 27 juin 2019, p. 45.

<sup>5</sup> Marin Dacos, « Le plan national pour la science ouverte. Où en sommes-nous ? », *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

publications dans des périodiques paraissant au moins une fois par an dans le champ des sciences, de la technique et de la médecine, et de douze mois pour les sciences humaines et sociales.

Par ailleurs, les données de la recherche sont librement communicables sous réserve de leur conformité au Règlement général de protection des données (RGPD) entré en vigueur en 2018. Ainsi, si le RGPD comprend des dérogations et le partage de données en environnement sécurisé, il n'est finalement possible d'ouvrir des données de recherche qu'à condition de les avoir caractérisées relativement aux données personnelles et, le cas échéant, de les avoir rendues anonymes ou ouvertes avec le consentement des personnes concernées. L'Inra, dès 2016, avait anticipé les enjeux de la protection des données personnelles en soulignant dans sa « Charte pour le libre accès aux publications et aux données » que « *la diffusion des données, obtenues par l'Inra ou transmises par des tiers, doit respecter les règles juridiques et éthiques qui prévalent, notamment dans l'utilisation des données personnelles ou soumises au secret (statistique, protection du patrimoine scientifique et technologique) ou dans l'utilisation de données acquises auprès de tiers* »<sup>1</sup>.

L'ouverture des données de la recherche suppose donc d'articuler des enjeux pluriels, de la protection des données personnelles aux préoccupations liées « *à la confidentialité, à la sécurité nationales, aux intérêts commerciaux légitimes, tels que les secrets d'affaires, et aux droits de propriété intellectuelle de tiers, conformément au principe "aussi ouvert que possible, mais aussi fermé que nécessaire"* »<sup>2</sup>. Cette caractérisation des exceptions est issue de la Directive 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public ; celle-ci prolonge les dispositifs juridiques français, notamment le CRPA dont l'article L321-2, en travaillant les données de la recherche à l'aune d'une richesse supplémentaire : les principes FAIR. Cette directive, selon Lionel Maurel<sup>3</sup>, peut dès lors avoir des effets importants et complémentaires à la loi pour une République numérique en transposant dans le droit les principes FAIR qui acquièrent dès lors une légitimité nouvelle et formelle. La gestion des données de la recherche se dote ainsi progressivement d'un appareil juridique cohérent sur lequel peuvent s'appuyer les communautés scientifiques, bien que leur définition reste toujours, semblerait-il, en suspens.

En effet, selon la directive européenne, les données de la recherche « *comprennent des statistiques, des résultats d'expériences, des mesures, des observations faites sur le terrain, des résultats d'enquêtes, des enregistrements d'entretiens et des images. Elles comprennent également des métadonnées, des*

---

<sup>1</sup> Inra, « Charte pour le libre accès aux publications et aux données », 2016, Inra, p. 6, [https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016\\_Charte-libre-access-Inra-Def\\_Inra-fran%C3%A7ais.pdf](https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016_Charte-libre-access-Inra-Def_Inra-fran%C3%A7ais.pdf).

<sup>2</sup> Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public [https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv:OJ.L\\_.2019.172.01.0056.01.FRA](https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.FRA)

<sup>3</sup> Lionel Maurin, « Introduction aux aspects juridiques de la production scientifique » dans *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019, p. 31.

*spécifications et d'autres objets numériques. Les données de la recherche diffèrent des articles scientifiques présentant et commentant des résultats de recherche scientifique effectuée par les auteurs.* »<sup>1</sup> Cette caractérisation large des données de la recherche, qui laisse aux communautés scientifique la place de manœuvrer selon leur contexte de production, est à l'image de leur définition dans le droit français. La loi pour une République numérique acte les données de la recherche comme des informations publiques à l'instar de toute autre donnée produite par une administration. Ce choix du législateur, raconte Lionel Maurin, fut l'objet d'âpres discussions à la fois sur la définition des données de la recherche et sur le caractère privé ou public des archives des chercheurs<sup>2</sup> ; des discussions qui ne sont toujours pas closes à l'heure actuelle. Certains, à l'instar de Céline Castets-Renard et de Nathalie Gandon, préfèrent sensibiliser les chercheurs au sujet des données de la recherche dans les termes d'origine de la loi CADA, apparentant celles-ci à des « *documents administratifs* »<sup>3</sup>. Cette vision, précédant la loi pour une République numérique et les recommandations des Plans nationaux pour la science ouverte, est ainsi reprise en 2016 par les professionnels de l'IST, telle Esther Dzalé-Yeumo à l'Inra, faisant des données des « *documents administratifs* », soit des informations « *de libre parcours* »<sup>4</sup>.

D'autres, à la manière de la section Aurore de l'AAF, font des données de la recherche « *l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. A ce titre, elles constituent une partie des archives de la recherche* »<sup>5</sup>. La section Aurore a à ce titre élaboré un référentiel de gestion des archives de la recherche, aux spécificités notables car « *il est parfois très difficile d'imputer une durée d'utilité administrative (DUA) correspondant à la durée de conservation des documents, étant donné leur usage scientifique plutôt qu'administratif* »<sup>6</sup>. Les archivistes ne se sont toutefois pas refusés à les intégrer comme archives de la recherche dont la gestion raisonnée doit être le lot de tout organisme de recherche.

Le cas d'INRAE est à ce titre intéressant, travaillant la problématique des données de la recherche depuis maintenant plusieurs décennies, s'associant à de nombreuses initiatives nationales et internationales, et se positionnant aujourd'hui, aux côtés du CNRS et d'autres instances, comme fer de lance de la science ouverte française.

---

<sup>1</sup> Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public, *op. cit.*, paragraphe 27.

<sup>2</sup> Lionel Maurin, « Introduction aux aspects juridiques de la production scientifique », *op. cit.*, p. 31.

<sup>3</sup> Céline Castets-Renard, Nathalie Gandon, « Open data des données de la recherche publique : entre réformes législatives et retour d'expérience sur un guide pratique à destination des chercheurs » dans *LEGICOM*, vol. 56, n° 1, 2016, p. 71.

<sup>4</sup> Esther Dzalé Yeumo, « Données de la recherche et recherche agronomique », stage ENSSIB, 07 juin 2016, p. 28

<sup>5</sup> Julien Pomart, « AAF / Section Aurore: Un groupe de travail sur les données de la recherche », *op. cit.*

<sup>6</sup> Section Aurore, *Référentiel de gestion des archives de la recherche*, Association des archives français, publié le 8 juin 2016, <https://www.archivistes.org/Referentiel-de-gestion-des-archives-de-la-recherche>, p. 2.

# INRAE ET LA GESTION DES DONNEES DE LA RECHERCHE

---

## L'ENGAGEMENT PRECURSEUR DES INSTITUTS FRANÇAIS DE RECHERCHE AGRONOMIQUE

### La recherche agronomique française à l'heure de l'ouverture

L'Institut national de la recherche agronomique, né au lendemain de la Seconde Guerre mondiale dans le but de se faire le fer de lance de la modernisation agricole, entre à partir des années 1990 dans une nouvelle phase d'introspection sur ses axes de recherche. Les nouveaux enjeux scientifiques et les outils du numérique bouleversent les ambitions de l'Inra qui est confronté, nous l'avons présenté plus tôt, à des problématiques plurielles et globales ; l'Institut est « à la croisée des chemins du changement global »<sup>1</sup> selon les auteurs de *L'histoire de l'Inra, entre science et politique*. Ainsi la directrice de l'Institut Marion Guillou, nommée en 2000, fait « de la révolution numérique et des outils de la modélisation le moyen d'une réunification des critères de scientificité de la recherche agronomique et d'une collaboration indirecte mais efficace entre sciences du vivant et sciences sociales pour la production de réponses pertinentes à la demande sociale »<sup>2</sup>. Le document d'orientation pour le contrat quadriennal 2001-2004 de l'Institut le présente d'ailleurs ainsi : « Les sciences biologiques connaissent une véritable révolution, qui conduit à leur appliquer le terme d'industrialisation ; elles sont au cœur d'enjeux économiques qui bouleversent les conditions de la concurrence et de la coopération entre la recherche publique et la recherche privée, est-il souligné dans le projet quadriennal de 2001. Le progrès de la modélisation et des outils informatiques ouvre des perspectives nouvelles pour l'étude des systèmes complexes mettant en jeu des processus qui associent différents niveaux d'organisation et se déroulent sur des échelles de temps et d'espace diversifiées. »<sup>3</sup> Les germes de la gestion raisonnée et nécessaire des données de la recherche en environnement ouvert et numérique commencent à être saisis ici par la direction de l'Inra.

Ce dernier se déploie d'ailleurs au cours des années 2000 en pleine restructuration politique de la recherche française marquée notamment par la création de l'ANR et de l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES). Or l'Inra, sous la présidence de Marion Guillou (2000-2012), parvient à conserver la cohérence de ses orientations scientifiques globales et à se positionner auprès de l'ANR, tout juste créé en 2005, pour piloter des programmes de recherche tant d'un point de vue scientifique que financier et administratif, tout

---

<sup>1</sup> Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra entre science et politique*, op. cit., p. 455.

<sup>2</sup> *Ibid.*, p. 392.

<sup>3</sup> Inra, *Contrat quadriennal. Un document pour l'action. Mise en œuvre des orientations 2001-2004*, Version provisoire, 24 septembre 2001, p. 7, cité dans Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra, entre science et politique*, op. cit., p. 392.

en prenant « *en charge la réflexion prospective relative à l'élaboration d'appels à projets. Le temps est loin où les universités et le CNRS pouvaient regarder de haut la recherche agronomique et la tenir à l'écart des grands programmes nationaux. L'Inra de Marion Guillou est installé au cœur du pouvoir scientifique français, avec des alliés solides* »<sup>1</sup>. Si le CNRS, parallèlement, se positionne dès les années 2000 par son envergure et sa force de frappe comme promoteur français de la science ouverte<sup>2</sup>, l'Inra, dans ces repositionnements scientifiques et politiques, n'est pas en reste. Comme le rappelle en 2019 Odile Hologne<sup>3</sup>, actuellement directrice de la Direction pour la Science ouverte (DiPSO) d'INRAE, l'Institut mène une politique de science ouverte depuis longtemps, mais celle-ci n'est véritablement affichée dans son document d'orientation que depuis 2016 avec le programme #Inra2025. La fusion en 2020 de l'Inra avec l'Irstea conduit à un nouveau document d'orientation actant définitivement les orientations du nouvel institut, INRAE, alors « *attentif à l'orientation et à l'élaboration de ses recherches ainsi qu'à la transmission et l'utilisation des résultats et enseignements de celles-ci, pour l'appui aux politiques publiques et pour l'innovation sous toutes ses formes* »<sup>4</sup>.

L'Irstea, quant à lui, a tout autant pris conscience dès les années 2000 du poids futur des données de la recherche et des mouvements d'ouverture en environnement numérique. Fondé en 1981 en tant que Centre national du machinisme agricole, du génie rural, des eaux et des forêts (CEMAGREF), il devient en 2011 l'Irstea dans une perspective de recherche finalisée renforçant la main de l'action publique sur les territoires. Pour Raymond Schiano, en charge de l'archivage numérique de l'Irstea à INRAE, l'Institut était avant la fusion « *centré sur une recherche scientifique ayant l'objectif de produire des solutions concrètes au bénéfice de la décision publique et débouchant sur l'action* »<sup>5</sup>. Il reposait alors sur quatre piliers :

- *Une recherche partenariale fondée sur une démarche de questionnements scientifiques avec de nombreux acteurs économiques ;*

- *Une recherche en appui aux politiques publiques à travers une importante coopération avec les ministères en charge de l'écologie et de l'agriculture, les collectivités territoriales et les services déconcentrés de l'État [...];*

- *Une coopération scientifique avec les organismes de recherche et les Universités ;*

- *Une dimension européenne et internationale avec notamment une implication dans les réseaux européens de recherche environnementale, l'évaluation des*

---

<sup>1</sup> Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra, entre science et politique*, op. cit., p. 422.

<sup>2</sup> Antoine Petit, « La science ouverte, une révolution nécessaire » dans *La science ouverte : une révolution nécessaire, 1ère journée pour la science ouverte au CNRS*, Paris, 8 octobre 2019.

<sup>3</sup> Odile Hologne, « Science Ouverte et publications à l'INRAE » dans *La science ouverte : une révolution nécessaire, 1ère journée pour la science ouverte au CNRS*, Paris, 8 octobre 2019.

<sup>4</sup> INRAE, « INRAE2030. Partageons la science et l'innovation pour un avenir durable », op. cit., p. 12.

<sup>5</sup> Raymond Schiano, « L'archivage des documents engageants à l'Irstea d'Aix-en-Provence : propositions pour le traitement homogène des dossiers papier et numériques au sein d'une équipe d'experts », *Théorie de l'information [IT]*, 2012, p. 4.

*politiques publiques européennes et des coopérations avec des universités et des établissements de recherche dans le monde.* »<sup>1</sup>

L’Inra et l’Irstea sont ainsi deux instituts confrontés à ce repositionnement global de la politique et de la société sur les enjeux environnementaux, alimentaires et sanitaires, et dont les objets d’étude en environnement numérique engendrent de nouveaux questionnements. Ces interrogations émergent dès les années 1990 à l’Inra, nous l’avons vu, sous la plume de Denis Poupardin qui étudie les « archives de la recherche » recouvrant un large spectre de la production documentaire, des données aux articles scientifiques. Son travail toutefois n’aboutit pas sous la politique escomptée, expliquant en 2008 : « *Les responsables à qui j’ai remis mes conclusions un an plus tard m’ont fait savoir, en effet, qu’ils n’avaient pas les moyens de les mettre en œuvre et ne souhaitaient en aucun cas se sentir engagé par elles. Retour donc à la case départ.* »<sup>2</sup>. Il faut donc attendre les années 2000 pour que la direction de l’Institut appréhende les enjeux sous-jacents à la production des données de la recherche en environnement numérique et ouvert. En effet, comme nous le raconte Odile Hologne, l’Inra est confronté d’une part au phénomène de *data deluge*, notamment en génomique où l’outil numérique est essentiel, et commençant à poser des problèmes de stockage ; la direction est alors interpellée sous l’angle des moyens mis à la disposition des équipes de recherche pour gérer ces données<sup>3</sup>. D’autre part, Marion Guillou engage l’Inra auprès de l’ANR dans une « *coopération logistique et scientifique* » afin de « *mettre à disposition ses compétences scientifiques et son savoir-faire organisationnel, ce qui soulage une agence novice, bien en peine, dans cette phase de démarrage, d’interagir avec la multiplicité des parties prenantes de la recherche* »<sup>4</sup> ; elle remarque alors, comme le rapporte Odile Hologne<sup>5</sup>, que les agences de financement à l’étranger mettent au point des *Data Sharing Policy*, notamment au Royaume-Uni et la *Biotechnology and Biological Sciences Research Council* dans le champs de la recherche en biologie, pour toutes les données produites sur fonds publics. Marion Guillou organise alors un groupe de travail sur le partage des données en 2009 qui délivre un rapport trois années plus tard.

Le contexte des années 2000 est évidemment porteur pour ces réflexions sur le partage des données de la recherche en aval de leur gestion raisonnée. L’Inra est d’ailleurs signataire de la Déclaration de Berlin dès juillet 2004, quand l’Irstea rejoint l’appel en mars 2006<sup>6</sup>. Dans le sillon des engagements qui devinrent les siens lors de la signature de la Déclaration, l’Inra se dote en 2005-2006 d’une archive ouverte institutionnelle, Prodinra, à destination de ses chercheurs et ingénieurs ;

---

<sup>1</sup> *Ibid.*, p. 4.

<sup>2</sup> Denis Poupardin, Communication au Colloque «Archorales» du 2 décembre 2008, Inra, [https://www6.inrae.fr/comitedhistoire/content/download/3002/30151/version/1/file/Expose\\_Denis\\_Poupardin.pdf](https://www6.inrae.fr/comitedhistoire/content/download/3002/30151/version/1/file/Expose_Denis_Poupardin.pdf).

<sup>3</sup> Odile Hologne, entretien du 16 juin 2021.

<sup>4</sup> Pierre Cornu, Egidio Valceschini, Odile Maeght-Bourney, *L’histoire de l’Inra, entre science et politique*, op. cit., p. 422.

<sup>5</sup> Odile Hologne, entretien du 16 juin 2021.

<sup>6</sup> « Déclaration de Berlin », Signataires, *Open Access Max-Planck-Gesellschaft*, <https://openaccess.mpg.de/319790/Signatories>.

l'Irstea quant à lui crée son dépôt institutionnel en 2009, CemOA, finalement intégré à Irstea Publications, et alimenté à partir des publications de ses agents. Les deux dépôts sont alors interconnectés à HAL jusqu'à leur bascule définitive quelques années plus tard.

Mais c'est bien toutefois l'organisation de ce groupe de travail sur la gestion et le partage des données qui marque un tournant au sein de l'Inra. Initié par la direction en septembre 2009 à la suite d'un audit du cabinet Ernst & Young sur les systèmes d'information de l'Institut, ce groupe de travail entreprend une « *réflexion prospective [...] dans un contexte d'évolution extrêmement rapide des technologies d'acquisition de données, dont la production massive qui en résulte a conduit au mouvement du Big Data* »<sup>1</sup>. L'Institut, en effet, s'interroge « *sur ses capacités à gérer, traiter et interpréter les très gros volumes qui sont/seront produits dans le cadre de ses programmes de recherche* », en particulier dans les champs de la biologie intégrative et systémique, qui se fait « *biologie numérique* »<sup>2</sup>, bien que les conclusions du groupe de travail s'ouvrent finalement « *à l'ensemble des domaines scientifiques de l'Inra pour lesquels le haut débit est (ou sera très vite) une réalité* »<sup>3</sup>. Composé de chercheurs, ingénieurs et professionnels de l'IST issus de l'Inra, mais aussi d'autres établissements d'ESR<sup>4</sup>, le groupe de travail rend ses conclusions en juin 2012 au sein d'un rapport qui devient le document d'orientation de l'Institut, en mettant en lumière « *les enjeux scientifiques et politiques de ce que l'on appelle désormais "la science ouverte" et "les mégadonnées" (big data) et qui concernent de nombreux domaines d'intérêt pour la recherche agronomique* »<sup>5</sup>. Onze grands principes sont alors définis, ainsi que neuf recommandations :

R1 - Définir la politique de l'établissement et la communiquer

R2 - Mettre en place un dispositif d'évaluation des données produites par l'Inra

R3 - S'impliquer dans les comités internationaux de standardisation

R4 - Développer un portail d'accès à un ensemble de ressources distribuées

R5 - Prendre en compte le cycle de vie des données dès l'élaboration des projets de recherche

R6 - Définir un cahier des charges pour les plateformes

R7 - Doter l'Inra d'infrastructures dimensionnées pour les stockages et calculs hautes performances

---

<sup>1</sup> Christine Gaspin, Dominique Pontier, Laurence Colinet, Frédéric Dardel, Alain Franc, *et al.*, *Rapport du groupe de travail sur la gestion et le partage des données*, Avis du conseil scientifique, Inra, juin 2012, p. 6.

<sup>2</sup> INRAE, *INRAE2030. Partageons la science et l'innovation pour un avenir durable*, INRAE, document d'orientation, décembre 2020, p. 21.

<sup>3</sup> *Ibid.*, p. 6.

<sup>4</sup> Christine Gaspin (Inra), Dominique Pontier (Université Lyon 1), Laurence Colinet (Inra), Frédéric Dardel (Université Paris Descartes), Alain Franc (Inra), Odile Hologne (Inra), Olivier Le Gall (Inra), Nicolas Maurin (Inra), Guy Perrière (CNRS), Christian Pichot (Inra), François Rodolphe (Inra).

<sup>5</sup> INRAE, « Un avis bienvenu à l'heure du numérique et de la "science ouverte" », Inrae.fr, [En ligne], <https://www.inrae.fr/actualites/avis-bienvenu-lheure-du-numerique-science-ouverte>.

R8 - S'engager dans une politique de gestion des compétences répondant aux besoins en émergence

R9 - Conduire une réflexion inter-organismes

Des groupes de travail succédant ce rapport sont alors organisés dans le but de travailler la mise en application des principes définis et de ces recommandations, afin, ensuite, « *de les traduire en termes stratégiques et opérationnels, incluant différents volets, humains (compétences), organisationnels, juridiques, matériels, techniques ou méthodologiques* »<sup>1</sup>. La direction organise en effet un chantier « gestion, partage et réutilisation des données » en 2013 mobilisant 80 acteurs concernés par les enjeux, scientifiques, informaticiens, documentalistes ou encore juristes<sup>2</sup>, qui œuvrent à la définition des données, des méthodes et des outils pour les appréhender, des stratégies à adopter, en termes scientifiques, techniques, mais aussi juridiques et déontologiques<sup>3</sup>.

L'expertise des chercheurs et des professionnels de l'IST de l'Inra se développe ainsi activement sur le champ des données de la recherche ; certains sont alors appelés à transmettre leurs savoirs et savoir-faire en dehors des murs de l'Institut. Odile Hologne, actrice centrale de la réflexion au sein du groupe de travail, participe au projet de la Bibliothèque numérique scientifique (BSN) créée en 2009. Elle rejoint ainsi certains de neuf segments initiaux de la BSN, notamment BSN4 consacré aux Archives ouvertes aux côtés d'Emmanuelle Jannès-Ober de l'Irstea, aujourd'hui directrice-adjointe de la DipSO d'INRAE. Lors de la création en 2014 du dixième segment sur les données de la recherche, Odile Hologne et Emmanuelle Jannès-Ober s'en font alors les responsables. Il est à noter que dans le segment 6, sur le sujet de l'archivage pérenne, on retrouve notamment Raymond Schiano, en charge de l'archivage numérique depuis l'Irstea jusqu'à aujourd'hui INRAE.

L'Inra, et l'Irstea parallèlement, se positionnent ainsi clairement sur les enjeux de la gestion et du partage des données de la recherche, mais aussi dans les mouvements d'ouverture qui marquent le paysage de la recherche scientifique française. L'article 2 du décret n° 2015-1517 du 23 novembre 2015 relatif à l'Institut national de la recherche agronomique marque clairement cette nouvelle dynamique. Il y est en effet écrit que :

*« Dans son domaine de compétences, l'institut :*

*1° Produit, publie et diffuse les connaissances scientifiques résultant de ses travaux de recherche et d'expertise ;*

*2° Organise, en l'absence de dispositions ou clauses contraires, l'accès libre aux données scientifiques et aux publications ;*

---

<sup>1</sup> *Ibid.*

<sup>2</sup> Comité consultatif commun d'éthique pour la recherche agronomique, *Avis 8 sur les enjeux éthiques et déontologique du partage et de la gestion des données issues de la recherche*, Inra, 2016, ffhal-02796585, p. 15.

<sup>3</sup> Odile Hologne. « Données de la recherche : rôle des professionnels IST à l'Inra », *Congrès de l'ADBU*, septembre 2013, Le Havre, France, p. 16, hal-02807047.

3° Contribue à l'élaboration de la politique et de la stratégie nationale et européenne de recherche ;

4° Apporte son concours à la formation à la recherche et par la recherche ;

5° 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 ;

6° Contribue au développement de la capacité d'expertise scientifique et d'appui aux fondations reconnues d'utilité publique et aux associations ; il apporte son soutien aux politiques publiques menées pour répondre aux défis de société, notamment aux besoins économiques, sociaux et environnementaux liés au développement durable ;

7° Contribue aux débats qui portent sur la place de la science et de la technologie dans la société. »<sup>1</sup>

Bref l'Inra se fait un organisme de recherche impliqué dans la société et pluriel par ses objets d'études, hybrides de nature et de culture, qui œuvrent à nourrir les réflexions sur les relations entre sciences et société, par l'ouverture de ses données, de ses résultats, de ses publications.

La même année, l'Inra met en ligne son site d'information Datapartage, conçu comme une boîte à outils accompagnant les communautés scientifiques dans la gestion et le partage des données de la recherche. En 2017, Irstea déploie de même son site *donnees-recherche-irstea.fr* qui se fait également boîte à outils dédiée à la gestion et la diffusion des données. L'Institut n'est en effet pas en reste sur ce sujet, lui qui est d'ailleurs engagé depuis 2015 dans le management de qualité par la mise au norme ISO 9001 de l'établissement. C'est à cette occasion, nous raconte Colette Cadiou, actuellement responsable du pôle Accès, capitalisation et diffusion des connaissances au sein de la DipSO et anciennement à l'Irstea, qu'est interrogée en profondeur la gestion des données de la recherche<sup>2</sup>. Le pôle IST de l'Institut publie alors le *Guide Irstea pour la gestion des données de la recherche* à destination des chercheurs et techniciens.

L'année suivante, l'Inra publie quant à lui sa « Charte pour le libre accès aux publications et aux données ». Etroitement liée au document d'orientation #Inra2025 qui fait la part belle à la science ouverte, la charte présente sept axes de travail, concrets et pragmatiques, apportant des modalités de réponses aux acteurs de la recherche agronomique pour cadrer les enjeux de science ouverte et de gestion des données de la recherche sur lesquels ils sont de plus en plus sollicités. Les sept axes sont les suivants :

- *Le libre accès aux informations issues de travaux de recherche doit être systématiquement recherché ;*

---

<sup>1</sup> Article 2 du décret n° 2015-1517 du 23 novembre 2015 relatif à l'Institut national de la recherche agronomique, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000031521114>.

<sup>2</sup> Colette Cadiou, entretien le 1<sup>er</sup> juillet 2021.

- *L'Inra organise la gestion et la valorisation de son information scientifique ;*
- *Rendre les informations interopérables pour garantir et faciliter leur réutilisation ;*
- *Privilégier le libre accès aux publications et aux données dans le respect du droit ;*
- *Pour l'évaluation du volet publication de l'activité des chercheurs et des ingénieurs, les commissions s'appuieront exclusivement sur les listes de publications et le texte intégral issus de l'archive institutionnelle Prodinra ;*
- *Garantir le stockage des informations numérisées et leur préservation sur le long terme ;*
- *Assurer une contribution des projets aux coûts de stockage et de valorisation des données ou aux frais de publication ;*
- *Assurer une contribution des projets aux coûts de stockage et de valorisation des données ou aux frais de publication<sup>1</sup>.*

Surtout, cette charte s'inscrit dans le sillon du travail du Comité consultatif commun d'éthique pour la recherche agronomique qui la même année, en 2016, publie son huitième avis sur les enjeux éthiques et déontologiques du partage et de la gestion des données issues de la recherche<sup>2</sup>. Ce livret complet de 32 pages marque l'entrée de la science ouverte dans les grandes orientations scientifiques, politiques et éthiques de l'Institut, que restitue le document d'orientation #Inra2025. Celui-ci inaugure en effet un renversement dans la prise en compte de la dynamique d'ouverture et de la gestion raisonnée des données de la recherche acté par le déploiement d'un entrepôt de données fonctionnel en 2018 : InraData. L'Irstea, la même année, rend accessible son catalogue des données de recherche IrsteaData. Les données se font dès lors une des briques au fondement d'une recherche agronomique de qualité.

En janvier 2020, la fusion de l'Inra et de l'Irstea marque le renforcement décisif de deux décennies d'efforts. INRAE, alliant les axes de recherche complémentaires de deux instituts, se dote d'une Direction pour la Science ouverte à la tête de laquelle se retrouvent Odile Hologne, ex-Inra, et Emmanuelle Jannès-Ober, ex-Irstea. La science ouverte se fait axe d'appui à la recherche décisif pour la recherche agronomique dont l'excellence nationale et internationale est directement liée aux efforts déployés par tous les acteurs engagés dans sa mise en œuvre. Ce n'est donc pas sans hasard que l'annonce du Deuxième Plan pour la Science ouverte a été réalisée par Frédérique Vidal au siège d'INRAE à Paris. Aujourd'hui sollicité pour installer la future plateforme nationale des données de la recherche engagée par ce Second Plan grâce à son expertise combinée sur les entrepôts de données DataInra

<sup>1</sup> Inra, « Charte pour le libre accès aux publications et aux données », 2016, Inra, 8 p., [https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016\\_Charte-libre-access-Inra-Def\\_Inra-fran%C3%A7ais.pdf](https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016_Charte-libre-access-Inra-Def_Inra-fran%C3%A7ais.pdf).

<sup>2</sup> Comité Consultatif Commun d'Éthique Pour La Recherche Agronomique, *Avis 8 sur les enjeux éthiques et déontologique du partage et de la gestion des données issues de la recherche*, Inra, 2016, fihal-02796585, 32 p.

et DataIrsa, le nouvel Institut s'est ainsi positionné comme acteur incontournable de la science ouverte en France.

## **La gestion des données de la recherche agronomique à l'ère du village mondial**

En 2019, la direction de l'Inra et les membres de l'alliance Agreenium, réunissant des établissements d'ESR dans le champ de la recherche agronomique, commandent une prospective sur le sujet de la transition numérique. Cette étude se projette à l'horizon 2040 dans l'objectif de saisir les enjeux futurs autour du numérique d'autant que, « *dans les domaines agronomique et vétérinaire, l'évolution de la recherche comme celle de la formation dépend fortement de la capacité à s'adapter aux changements en matière de données* »<sup>1</sup>. Les auteurs de la prospective consacrent ainsi un pan de leur étude à la question des données de la recherche en identifiant plusieurs enjeux : sur le sujet de la production et du stockage des données, doit-on et peut-on toutes les conserver ? Sur le traitement des données : quelles méthodes et quels outils, pour appréhender la diversité et le volume des données ? Sur la qualité des données, problématique centrale, complexe et globale qui se pose aux sciences appelant à mobiliser des jeux de données très hétérogènes : quelles exigences de qualité et d'interopérabilité ? Enfin, corollaire des problématiques de qualité, quelles modalités de circulation de la connaissance et des données ?

Cette prospective restitue alors les enjeux de la gestion des données de la recherche et de la science ouverte dans un environnement scientifique et numérique complexe. En effet, les nouvelles problématiques environnementales, alimentaires et sanitaires s'affranchissent des frontières et mobilisent des communautés à l'échelle de la planète, engagée dans l'Anthropocène. Par ailleurs, dominé et modifié par les GAFAM<sup>2</sup>, cet environnement peut rapidement devenir « *clivant dans les relations ESR société* ». Les auteurs se font alors alertes : « *Le monde académique, qui n'a pas pris en main de façon active son acculturation aux nouveaux outils numériques, est distancé par ces nouveaux relais d'opinion et sa crédibilité baisse encore. Il ouvre ses données mais elles sont rares du fait de la faiblesse de l'État et ses capacités financières à capter d'autres données sont faibles. Il cohabite avec une donnée ouverte maîtrisée par des GAFAM très puissants, non contraints, qui préconisent l'ouverture des données de la part de l'ensemble des acteurs, jouissent des données publiques, du travail des bénévoles et des traces des consommateurs, mais ne communiquent pas leurs algorithmes et captent les métadonnées ouvertes pour en faire un usage commercial.* » La recherche scientifique doit donc être en mesure de construire des espaces d'interactions fructueux par l'usage du numérique tant à destination de la société civile que d'elle-même. INRAE, et les deux instituts

---

<sup>1</sup> Inra, *Transition numérique et pratiques de recherche et d'enseignement supérieur en agronomie, environnement, alimentation, sciences vétérinaires à l'horizon 2040*, Résumé de la prospective, 27 juin 2019, p. 39.

<sup>2</sup> Acronyme désignant les grandes entreprises du numérique et leur modèle politico-économique : Google, Amazon, Facebook, Apple, Microsoft.

avant leur fusion, ont clairement compris la nécessité de s'inscrire, dans le cadre de la gestion des données de la recherche et du déploiement de la science ouverte, au sein de groupes nationaux ou internationaux en mesure de replacer les enjeux autour du numérique dans le giron des sciences agronomiques. Surtout, selon les mots d'Odile Hologne, il faut que les organismes de recherche, et les professionnels de l'IST en leur sein, travaillent « *avec les autres partenaires pour que chacun ne réinvente pas l'eau chaude dans son coin* »<sup>1</sup>.

Dès les années 2000, l'Inra sous la direction de Marion Guillou avait saisi l'importance pour l'Institut de se positionner dans un espace riche en opportunités scientifiques et politiques. Ainsi, lors de la valorisation de l'Espace européen de la recherche au début des années 2000, l'Inra, selon l'historien Pierre Cornu, trouve un moyen à long terme d'« *améliorer sa position* ». En effet, « *l'Inra adhère totalement à la déclaration contenue dans l'introduction de la seconde communication de la Commission relative à "l'Espace européen de la recherche" [...], selon laquelle : "la mondialisation de l'économie et de la communication, l'accélération des [découvertes] scientifiques et la multiplication de [leurs] conséquences sociales, font de [cet Espace] un objectif dont on ne peut différer la réalisation" »*<sup>2</sup>. L'Europe, dès lors, se fait contexte porteur pour la recherche agronomique française, conduisant l'Inra à expliquer en ces mots devant la Commission européenne que « *la politique de recherche agronomique européenne doit être considérée en lien avec les autres politiques communautaires dans la mesure où cette recherche répond à des objectifs [...] tels que : appui aux politiques internes et externes de l'Union, renforcement de la compétitivité des produits européens sur le marché mondial, contribution à l'indépendance européenne, connaissance et gestion des biens non marchands, etc.* »<sup>3</sup>. De manière plus globale encore, l'Inra s'engage dans la restructuration des dispositifs d'envergure que sont les infrastructures de recherche, « *faits d'équipements coûteux en moyens financiers et humains, qui réclament des investissements lourds et de long terme, mais dont dépend pour partie la production de données indispensables pour mener des recherches nouvelles et pour participer aux grands consortiums internationaux en plein développement* »<sup>4</sup>.

L'Inra, et maintenant INRAE, s'inscrit ainsi dans plusieurs dispositifs, à différentes échelles, qu'elles soient régionale, nationale, européenne si ce n'est mondiale, pour tenter, selon les mots du document d'orientation #Inra2025, « *d'harmoniser les orientations dans la recherche agronomique qui se réalise dans un cadre complexe d'enjeux interconnectés, qu'ils soient climatiques, environnementaux, scientifiques, éthiques et ou encore bioéthiques* »<sup>5</sup>. L'Institut fait

---

<sup>1</sup> Odile Hologne, entretien du 16 juin 2021.

<sup>2</sup> Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra, entre science et politique, op. cit.*, p. 422.

<sup>3</sup> Inra-DARESE, « Contribution de l'Inra à la préparation du 6<sup>ème</sup> PCRD dans le cadre de la construction de l'espace européen de la recherche », Note préparatoire, octobre 2000, cité dans Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra, entre science et politique, op. cit.*, p. 408.

<sup>4</sup> Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra entre science et politique, op. cit.*, p. 404.

<sup>5</sup> Odile Vilotte, Francois Houllier, Philippe Mauguin, Claude Ronceray, Christine Cherbut, *et al.*, « Document d'Orientation #Inra2025 », Inra, 2016, p. 32.

ainsi partie d'alliances nationales, telles Agreenium, Allenvi (Alliance nationale de recherche pour l'environnement), Aviesan (Alliance pour les sciences de la vie et de la santé) et d'autres, ainsi qu'internationales, telle la *Global Research Alliance* portée sur la réduction des émissions des gaz à effet de serre. Toutefois, si chacune de ces structures abordent, directement ou non, les enjeux de la gestion des données de la recherche et de la science ouverte, c'est au sein de dispositifs spécifiquement dédiés qu'INRAE s'inscrit dans ces questionnements. L'Institut fait ainsi partie depuis 2013 de la *Research Data Alliance* (RDA), dont l'intégration n'est pas étrangère aux conclusions du groupe de travail sur la gestion et le partage des données de la recherche. La *Research Data Alliance*, qui se fait action internationale, réunit des compétences diverses dans des réseaux de communautés scientifiques plurielles ce qui est, pour Françoise Genova, ancienne directrice du Centre de données astronomiques de Strasbourg et participante régulière à la *Research Data Alliance*, « essentiel pour travailler à des recherches transversales et interdisciplinaires et pour évaluer l'utilisation d'éléments génériques de l'infrastructure de données pour les besoins disciplinaires »<sup>1</sup>.

Ainsi, les professionnels de l'IST de l'Inra, dirigés par Odile Hologne, ont activement participé à la mise en place du groupe d'intérêt *Wheat Data Interoperability*, co-présidé par Esther Dzalé-Yeumo<sup>2</sup>. Ce groupe d'intérêt est créé en 2012 dans le cadre du G20 sous présidence française qui initie le projet *Wheat Initiative* dont l'objectif est d'accroître les savoirs sur la génétique du blé afin d'améliorer la résistance aux maladies et les rendements en contexte de crises alimentaires récurrentes. Cette coordination des recherches agronomiques à l'échelle mondiale appelait une réflexion sur la gestion des données de la recherche, d'autant que, selon Odile Hologne, « la communauté plante science était très sensible à la question de la data alors liée aux questions de séquençage, génotypage, phénotypage des céréales, notamment du blé »<sup>3</sup>; il y avait donc « toute une dynamique communautaire dans laquelle on a pu s'insérer et travailler au niveau international ». Le groupe de travail *Wheat Data Interoperability* émit finalement plusieurs recommandations dites RDA sur les données issues des recherches sur le blé et devant être, à terme, conformes aux principes FAIR. Cette initiative, fructueuse, est suivie de l'organisation d'autres groupes de travail, notamment sur l'interopérabilité des données sur le riz, recoupant le projet *Semantics for the Interoperability of Agricultural Data* (Agrisemantics) auquel participe INRAE.

En dehors de la *Research Data Alliance*, l'Institut, membre de la *Global Open Data for Agriculture and Nutrition* (GODAN), participe au mouvement de la *Coherence in Information for Agriculture Research* (CIARD) dédié à l'ouverture des connaissances en agriculture. L'Institut s'associe aussi au travail de *Science*

---

<sup>1</sup> Françoise Genova, *et al.*, « Building a Disciplinary, World-Wide Data Infrastructure », *Data Science Journal*, 16, 2017, DOI: <http://doi.org/10.5334/dsj-2017-016>.

<sup>2</sup> Esther Dzale Yeumo, Richard Fulss, Rosemary Shrestha, « Wheat Data Interoperability WG », <https://www.rd-alliance.org/group/working-and-interest-group-chairs-wheat-data-interoperability-wg/outcomes/wheat-data>.

<sup>3</sup> Odile Hologne, entretien du 16 juin 2021.

*Europe* travaillant notamment les enjeux de management de la donnée<sup>1</sup>. En dehors du champ international dédié à l'agriculture et la recherche agronomique, INRAE participe également à différents programmes et projets à l'échelle européenne au niveau de la donnée, notamment au projet *European Open Science Cloud (EOSC) Pillar*, initié en 2019, coordonnant les efforts nationaux de plusieurs pays européens (Autriche, Belgique, Allemagne, Italie et France) pour la mise en place d'un espace virtuel européen comme offre de service *cloud* sécurisé et transparent pour les chercheurs et professionnels des sciences, pouvant y déposer et y gérer leurs données de la recherche.

Ainsi INRAE, revendiquant son excellence dans le classement des organismes de recherche agronomique mondiale, joue un rôle moteur dans ces réseaux internationaux d'experts tant sur le versant de la recherche scientifique que celui des questions de gestion et de partage des données de la recherche. C'est un savoir-faire acquis au cours des deux dernières décennies mais initié, durant les années 2000, par le développement de l'*open access* et de la critique du modèle traditionnel de l'édition scientifique.

## Les premières pierres : réinventer l'édition scientifique

La sociologie des sciences l'étudie depuis plusieurs décennies, la « *science des périodiques* »<sup>2</sup>, soumise à la controverse et loin du caractère définitif et impersonnel d'un manuel, est un jalon fondamental et recteur du processus de création d'un fait scientifique auquel aucune communauté scientifique ne peut échapper. « *Un document devient scientifique, explique Bruno Latour, lorsque ce qu'il dit cesse d'être isolé et lorsque ceux qui sont engagés dans sa publication sont nombreux et explicitement indiqués dans le texte.* »<sup>3</sup> Or le processus d'édition scientifique connaît depuis la fin de la seconde moitié du XX<sup>e</sup> siècle une transformation majeure induite par les outils du numérique qui, en pleine crise de la modernité, participent de la restructuration épistémologique et éthique du « faire science ». Face au numérique, son modèle économique évolue vers le *Big Deal*<sup>4</sup> – soit un accès tarifé à un bouquet de revues –, et conduit à une concentration des bénéfices aux mains d'une poignée d'éditeurs. Ceux-ci détiennent en effet des ensembles de revues au sein desquelles les communautés scientifiques *doivent* publier pour légitimer la qualité et la pérennité de leurs recherches, profitant en cela de cet environnement engageant pour les chercheurs régulièrement caractérisé par la formule « *publish or perish* ». Le pacte originel liant chercheur et éditeur est alors rompu<sup>5</sup> : initialement caractérisé par la renonciation du premier sur tout rémunération de ses articles publiés pour la seule « *reconnaissance scientifique et sociale* » et les frais engagés du second dans

---

<sup>1</sup> Science Europe, *Practical Guide to the International Alignment of Research Data Management*, 2021, 29 p.

<sup>2</sup> Ludwik Fleck, *Genèse et développement d'un fait scientifique*, *op. cit.*, p. 206.

<sup>3</sup> Bruno Latour, *La science en action. Introduction à la sociologie des sciences*, Paris, La Découverte, 1989, p. 86.

<sup>4</sup> Pierre Carbone, « La relation consortiums-éditeurs », Joachim Schöpfel, *La publication scientifique. Analyses et perspectives*, Paris, Lavoisier, 2008, p. 229.

<sup>5</sup> Alain Jacquesson, Jean-Philippe Schmitt, « Les grands éditeurs face au mouvement open access », Joachim Schöpfel, *La publication scientifique. Analyses et perspectives*, Paris, Lavoisier, 2008, p. 107.

l'édition et la distribution des revues, il se confronte à l'*open access* promettant le partage facilité des connaissances scientifiques alors que les éditeurs, profitant de leur situation d'oligopole, augmentent les tarifs des abonnements dans le cadre du *Big Deal*. Ce sont alors plusieurs dizaines de millions d'euros dépensés annuellement par les bibliothèques des établissements d'ESR français pour pourvoir leurs collections des périodiques nécessaires à la consultation des résultats de la recherche et, *in fine*, à la validation des faits scientifiques.

Le mouvement d'*open access* comme il se déploie à partir des années 2000 se fait de la sorte une des réponses possibles à ce système jugé absurde, sinon pernicieux, et dont les conséquences plurielles se répercutent à diverses échelles. La définition BBB (pour Budapest, Bethesda et Berlin) redéfinit ainsi au début des années 2000 les modalités du partage des savoirs scientifiques caractérisé par l'accès libre impliquant la « *mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet* »<sup>1</sup>. Or si cette mise à disposition peut advenir sous différentes formes, les archives ouvertes se font un des instruments idoines de cette « *révolution nécessaire* »<sup>2</sup>. Initialement développées sur le modèle d'arXiv dans une perspective disciplinaire dédiée aux communautés scientifiques, les archives ouvertes sont finalement appréhendées comme infrastructures de recherche pour les établissements d'ESR à qui est dévolue leur gestion. En 2003 aux Etats-Unis, Cliff Lynch, directeur de la *Coalition for Networked Information* engagée dans le développement des technologie numériques de l'information pour la recherche et l'éducation, définit les archives ouvertes comme l'offre de service qu'une institution est en mesure de proposer à ses chercheurs et l'engagement de celle-ci dans la gestion de leurs ressources numériques, notamment leur préservation à long terme lorsque cela est approprié, ainsi que leur organisation, leur accès et leur distribution<sup>3</sup>. Les archives ouvertes institutionnelles se multiplient durant les années 2000, des Etats-Unis à la Norvège<sup>4</sup>, en passant évidemment par la France et le projet HAL du CCSD. Si HAL est un projet phare en France, plusieurs organismes de recherche déploient au cours de la décennie des archives ouvertes qui peuvent être thématiques, propres à une discipline ou à un établissement d'ESR. En 2012, ce sont 82 archives

---

<sup>1</sup> « Initiative de Budapest sur le libre accès », 14 février 2002, <https://www.budapestopenaccessinitiative.org/boai-10-translations/french>.

<sup>2</sup> Antoine Petit, « La science ouverte, une révolution nécessaire » dans *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>3</sup> Clifford A Lynch, « Institutional Repositories : Essential Infrastructure for Scholarship in the Digital Age », *Association of Research Libraries*, 226, February 2003, p. 2.

<sup>4</sup> Françoise Vandooren, « Evolution de l'accès aux publications scientifiques », Joachim Schöpfel, *La publication scientifique. Analyses et perspectives*, Paris, Lavoisier, 2008, p. 134-135.

institutionnelles qui sont recensées, dont 65 établissements ayant créé un espace HAL pour 17 équipés de leur propre dépôt local<sup>1</sup>.

C'est le cas de l'Inra qui, après avoir rejoint la Déclaration de Berlin, développe son archive ouverte institutionnelle ProdInra, dans la continuité de sa base patrimoniale PubInra créée en 1985 et disposant d'ores et déjà d'environ 100 000 notices<sup>2</sup>. Créée en 2005-2006, ProdInra répond aux besoins de cet institut combinant recherche fondamentale et approche finalisée, nécessitant en conséquence l'enregistrement d'une gamme diversifiée de produits. Ce dépôt institutionnel est envisagé comme une action décisive en faveur de l'*open access*, mais aussi comme une base patrimoniale numérique fiable « *pour la bibliométrie et le suivi des productions* »<sup>3</sup>. Selon Odile Hologne, entrée à l'Inra en septembre 2004<sup>4</sup>, l'Institut avait originellement une approche très documentaire de l'*open access*, « *compte tenu de sa tradition d'insertion de bases de publications dans un réseau international lié à la base bibliographique AGRIS gérée par la FAO [Organisation des Nations unies pour l'alimentation et l'agriculture].* » Or ses engagements internationaux en faveur du libre accès conduisent à renouveler la vision documentaliste de l'archive institutionnelle qui, dès lors, « *s'est retrouvée dans cette mouvance* ». Il ne s'agissait cependant pas de renier tous les savoir-faire des professionnels de l'IST, dont leur approche documentaire leur permit « *de disposer de services très poussés de back-office pour la gestion de la qualité des métadonnées* »<sup>5</sup> indispensable au déploiement d'un libre accès fiable.

ProdInra entendait recenser toutes les productions des chercheurs de l'Institut et en permettre la diffusion facilitée intracommunautaire. Toutefois pleinement insérée dans le contexte réflexif de la recherche agronomique en temps de crise, l'archive ouverte, couvrant tous les axes disciplinaires de l'Institut, s'ouvre aussi vers « *les professionnels, les pouvoirs publics, les étudiants, le grand public...* »<sup>6</sup> dans cet objectif d'*open access* au profit de la société dans son ensemble. Les scientifiques sont dès lors « *fortement encouragés à déposer leurs écrits en texte intégral dans cette archive institutionnelle et à les diffuser librement et sans restriction sous réserve de disposer des droits de diffusion* »<sup>7</sup>. En 2015, ce sont ainsi 215 000 références bibliographiques renseignées dont 25 800 avec un document en libre accès. Depuis 2012, ProdInra se caractérise par un flux de 15 000 dépôts par an dont 25% avec un document en libre accès. Ce sont alors 80 professionnels de l'IST ou référents d'unités formés qui accompagnent les chercheurs dans leurs

---

<sup>1</sup> « Rapport de l'étude n° 4 : l'articulation des archives des établissements et de l'archive nationale pluridisciplinaire HAL », Etudes COPIST, ISTEEX, p2018, p. 7.

<sup>2</sup> Fanny Dedet. « HAL et les archives ouvertes institutionnelles : l'exemple de ProdInra », *L'Open Access, questions économiques, éthiques et juridiques*, Université Clermont Auvergne, octobre 2015, Clermont-Ferrand, p. 17, hal-02801873.

<sup>3</sup> *Ibid.*, p. 17.

<sup>4</sup> Odile Hologne, entretien du 16 juin 2021.

<sup>5</sup> *Ibid.*

<sup>6</sup> INRAE, « ProdInra. L'archive ouverte de l'Inra », *ist.inrae.fr*, [En ligne], consulté le 24 juin 2021, <https://ist.inrae.fr/le-libre-acces-a-inrae/hal-inrae/prodInra/>.

<sup>7</sup> Inra, « Charte pour le libre accès aux publications et aux données », 2016, Inra, p. 5, [https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016\\_Charte-libre-access-Inra-Def\\_Inra-fran%C3%A7ais.pdf](https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016_Charte-libre-access-Inra-Def_Inra-fran%C3%A7ais.pdf).

dépôts, contrôlent les produits entrés, s'assurent de la qualité des éléments renseignés<sup>1</sup>. Cinq ans plus tard, en mai 2020, lorsque ProdInra ferme pour une migration du contenu dans HAL, ce sont ainsi 285 000 notices, 55 500 fichiers en libre accès et 27 000 en accès conditionné qui sont basculés<sup>2</sup>.

Les liens entre ProdInra et HAL sont étroits depuis le milieu des années 2000 ; en 2006, plusieurs organismes de recherche, dont l'Inra et l'Irstea aux côtés du CNRS, de l'INSERM et d'autres, signent un protocole d'accord en vue d'une approche coordonnée au niveau national pour l'archivage ouvert de la production scientifique<sup>3</sup>. L'Irstea n'est en effet pas en reste, déployant parallèlement à l'Inra et d'autres organismes son archive institutionnelle CemOA intégrée à la base documentaire Irstea Publications à une époque où, en 2005, les performances HAL sont jugées par les professionnels de l'établissement limitées<sup>4</sup>. A partir de 2010, l'Inra, quant à lui, décide de reverser régulièrement dans HAL à des fins de pérennisation certaines de ses productions entrées dans ProdInra, bien que les enjeux d'interopérabilité des contenus entre les deux plateformes aient été particulièrement délicates à gérer<sup>5</sup>. Le dépôt institutionnel de l'Institut se concevait ainsi comme la base active de l'environnement documentaire scientifique. En effet, la seule garantie pour « *le stockage des informations numérisées et leur préservation sur le long terme* » était de réaliser l'export des contenus vers HAL profitant des infrastructures serveurs du CINES à même d'assurer la pérennisation des données et documents ; ceux-ci basculaient dès lors dans une logique archivistique de patrimonialisation. Toutefois, constatant au fur et à mesure des années 2010 l'obsolescence du socle technologique de ProdInra, et développant une vision davantage politique de la gestion des publications et des données<sup>6</sup>, l'Inra fait finalement en 2019, après deux années de réflexions, le choix de HAL comme plateforme de dépôt institutionnel et devient cotutelle de l'Unité mixte de service du CCSD aux côtés du CNRS, de l'Inria et de l'Université de Lyon. L'Institut contribue au développement de HAL comme « *infrastructure nationale* » afin de permettre la « *réappropriation du processus de publication scientifique par les communautés* »<sup>7</sup>.

C'est en effet l'enjeu principal du mouvement *open access* : celui de permettre aux communautés de chercheurs de contester aux éditeurs scientifiques traditionnels le monopole de la publication scientifique, et en corollaire, du contrôle et de la validation des faits scientifiques. L'Inra œuvre depuis plusieurs décennies dans ce champ-là, en publiant ses propres revues académiques dont il est propriétaire ou co-

---

<sup>1</sup> Fanny Dedet. « HAL et les archives ouvertes institutionnelles : l'exemple de ProdInra », *op. cit.*, p. 7.

<sup>2</sup> INRAE, « ProdInra. L'archive ouverte de l'Inra », *ist.inrae.fr*, [En ligne], consulté le 24 juin 2021, <https://ist.inrae.fr/libre-acces-a-inrae/hal-inrae/prodInra/>.

<sup>3</sup> « Convention de partenariat en faveur des archives ouvertes et de la plateforme mutualisée HAL », Protocole d'accord V2, CCSD, 2016, p. 1, [https://www.ccsd.cnrs.fr/wp-content/uploads/2016/03/Convention\\_partenariat\\_HAL.pdf](https://www.ccsd.cnrs.fr/wp-content/uploads/2016/03/Convention_partenariat_HAL.pdf).

<sup>4</sup> « Rapport de l'étude n° 4 : l'articulation des archives des établissements et de l'archive nationale pluridisciplinaire HAL », Etudes COPIST, ISTEEX, 2018, p. 4.

<sup>5</sup> Fanny Dedet. « HAL et les archives ouvertes institutionnelles : l'exemple de ProdInra », *op. cit.*, p. 19 à 24.

<sup>6</sup> Odile Hologne, « Science Ouverte et publications à l'INRAE », *op. cit.*, p. 49.

<sup>7</sup> Agnès Magron, « De ProdInra et Irstea Publications au portail HAL-INRAE », CCSD, le 25 mars 2020, [En ligne], <https://www.ccsd.cnrs.fr/2020/03/de-prodInra-au-portail-hal-inrae/>.

propriétaire – telles *Le Lait*, *Veterinary Research*, *Annals of Forest Science*, *Inrae Sciences sociales*, etc.<sup>1</sup> –, avec le choix différencié selon les revues des voies dites verte ou dorée selon le modèle de partage des contenus retenu. La voie verte est ainsi choisie pour plusieurs de ses revues diffusées sur le modèle de l’abonnement et dont les archives sont librement accessibles via HAL après une période d’embargo. En 2016, ce sont 25 000 articles accessibles selon ce modèle comprenant pour certaines revues telles *Le Lait* et *Annals of Forest Science* les articles publiés depuis leur création (respectivement en 1921 et 1923 pour ces exemples)<sup>2</sup>. La voie dorée, elle, suppose que les coûts de libre accès immédiat des articles au sein des revues détenues par des éditeurs soient supportés sinon par le chercheur, du moins par les établissements d’ESR. Ce modèle, retenu pour quelques revues appartenant à l’Inra, sinon pour les articles des chercheurs de l’Institut publié dans d’autres périodiques, est largement décrié et ne peut semblerait-il subsister en l’état<sup>3</sup>. Preuve en est les abandons successifs de l’adhésion de l’Inra aux deux grands éditeurs promouvant la voie dorée, *Biomed Central* et *Public Library of Sciences (PLOS)*. En effet, l’Institut adhère en 2012 à ces offres permettant de couvrir cette même année 65% des publications en *open access*. Toutefois, l’adhésion à PLOS n’est pas renouvelée dès l’année suivante à la suite du changement du modèle économique de l’éditeur, et en 2019 pour *Biomed Central* en raison de la diminution du nombre d’articles soumis et de l’augmentation du coût de l’adhésion<sup>4</sup>. Pour reprendre les mots du directeur général du CNRS Antoine Petit, « *le fait que cela soit une science ouverte ne veut pas dire que c’est gratuit. Cela a un coût, il faut en être conscient* »<sup>5</sup>, et dès lors peut-être repenser les enjeux de la publication scientifique dans la dynamique d’ouverture originelle qu’était celle de l’*open access*.

Odile Hologne, en 2019, définit trois axes de recherche principaux pour l’avenir des publications en environnement ouvert :

« - adapter le modèle éditorial et économique des revues académiques dont l’INRAE est propriétaire au principe d’une science ouverte ;

- explorer et encourager des modèles alternatifs de publications ;

- envisager la sortie du système des abonnements aux revues à l’horizon 2020 »<sup>6</sup>.

Une des réponses peut alors provenir du développement des épisciences auxquelles l’Inra participe, notamment par la contribution à l’épi-revue *Journal of data mining and digital humanities*, dont l’objectif est de renouveler les modalités

---

<sup>1</sup> « Archives ouvertes des revues INRAE », <https://hal.archives-ouvertes.fr/ARINRAE/>.

<sup>2</sup> Inra, « Charte pour le libre accès aux publications et aux données », 2016, Inra, p. 7, [https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016\\_Charte-libre-access-Inra-Def\\_Inra-fran%C3%A7ais.pdf](https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016_Charte-libre-access-Inra-Def_Inra-fran%C3%A7ais.pdf).

<sup>3</sup> Minard, Philippe. « Les revues à l’âge numérique : au péril de l’idéologie », *Revue d’histoire moderne & contemporaine*, vol. 62-4, no. 5, 2015, p. 8 à 21, DOI : 10.3917/rhmc.625.0008.

<sup>4</sup> INRAE, « Le libre accès à INRAE », [ist.inrae.fr](http://ist.inrae.fr), [En ligne], <https://ist.inrae.fr/le-libre-acces-a-inrae/>.

<sup>5</sup> Antoine Petit, « La science ouverte, une révolution nécessaire » dans *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>6</sup> Odile Hologne, « Science Ouverte et publications à l’INRAE », *op. cit.*, p. 49.

de contrôle des résultats de la recherche par les pairs, le *peer reviewing*, en environnement d'*open access*. En effet, réinventer l'édition scientifique dans les mouvements d'ouverture, c'est aussi interroger ce contrôle des résultats aux fondements de la construction du fait scientifique et, en corollaire, l'évaluation des chercheurs dépendant de multiples éléments comme les taux de citation, les facteurs d'impacts, *etc.* C'est d'ailleurs un enjeu d'autant plus fort pour l'Inra que celui-ci ambitionnait au cours des années 2000 de rivaliser avec les grands organismes de recherche mondiaux « *qui le concurrencent de plus en plus sur ses propres champs* » et imposent, pour rester compétitif, « *une production scientifique calibrée selon les critères internationaux de l'excellence académique, fondés sur l'impact des publications* »<sup>1</sup>. L'Institut opère alors cet alignement pour finalement être classé, sur la période 2015-2019, « *au 3ème rang mondial en nombre de citations et au 4ème en nombre de publications* »<sup>2</sup>.

Ces mots, extraits du document d'orientation INRAE2030, témoignent de la situation contradictoire dans laquelle est actuellement l'Institut – et plus largement les chercheurs engagés dans la science ouverte –, promouvant sa place à l'international selon des critères dont il en dénonce par ailleurs ouvertement les limites. En effet, cette double dynamique de la science ouverte et de la course à l'excellence selon les modalités traditionnelles d'évaluation des faits scientifiques révèle un paradoxe pour le moment irréductible : « *le facteur d'impact, selon Cherifa Boukacem, professeur en sciences de l'information et de la communication, l'article, la revue indexée dans les bases de données sont pour le moment les seuls éléments de langage communs d'un pays à un autre.* »<sup>3</sup> Pourtant les chercheurs sont de plus en plus incités à « publier autrement » en accord avec les nouveaux dispositifs de partage des savoirs en environnement numérique, d'autant que des études observaient dès le début des années 2000 un accroissement des citations de publications en *open access* par rapport aux autres, parfois jusqu'à 157 %<sup>4</sup>.

Réinventer l'édition scientifique suppose donc de lutter contre l'inertie des communautés de pensées et les modalités actuelles du « faire science », ce que la science ouverte peut permettre en partie par la réorganisation de l'évaluation de la recherche et des chercheurs. Progressivement, celle-ci ne s'effectue plus par l'étude du facteur d'impact, dont la légitimité dans le système actuel a drastiquement chuté. Le Deuxième plan pour la Science ouverte l'inscrit clairement dans ses ambitions en préconisant de « *supprimer toutes les références à cet indicateur et au H-index dans les textes d'appels à projets et les formulaires de candidature* ». Il s'agit de faire des archives ouvertes un moment clé de l'évaluation de l'activité du chercheur et, *in fine*, une part de la pérennisation de ses traces. La DipSO, de concert avec la Direction à l'évaluation (DEV) et la Déléguée à l'intégrité-déontologie-éthique,

---

<sup>1</sup> Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra entre science et politique*, *op. cit.*, p. 404.

<sup>2</sup> INRAE, « INRAE2030. Partageons la science et l'innovation pour un avenir durable », *op. cit.*, p. 12.

<sup>3</sup> Cherifa Boukacem, « Table ronde » dans *La science ouverte : une révolution nécessaire*, *op. cit.*, p. 17.

<sup>4</sup> Frédérique Bordignon, Mathieu Andro. « Impact de l'Open Access sur les citations : une étude de cas », *I2D - Information, données & documents*, vol. 53, no. 3, 2016, p. 71.

œuvre ainsi à faire de HAL INRAE la plateforme de dépôt à partir de laquelle réaliser l'évaluation de la recherche agronomique<sup>1</sup>.

## LA GOUVERNANCE DES DONNEES A INRAE

### Les principes de la gouvernance

Le 1<sup>er</sup> janvier 2020, INRAE naît officiellement de la fusion de l'Inra et de l'Irstea ouvrant de nouvelles opportunités pour prolonger les efforts fournis sur les enjeux de science ouverte par les deux Instituts les deux décennies précédentes. C'est la naissance d'une réflexion sur la gouvernance des données en tant que telle, délimitant un ensemble de problématiques, de principes et de perspectives pour appréhender sur le long terme la question de la gestion des données de la recherche. Bien sûr, l'Inra et l'Irstea avaient déjà engagé des mesures de gestion des données de la recherche en environnement numérique et ouvert, nous l'avons vu, le premier mobilisant dès le début des années 2010 une cinquantaine de personnes dans l'élaboration d'une politique de partage des données<sup>2</sup>. La charte de déontologie de l'Inra inclut d'ailleurs depuis 2013 la question des données de la recherche qui, avec « *les savoir-faire et collections, la propriété industrielle* », sont « *un patrimoine de qualité à préserver et valoriser* »<sup>3</sup>. En effet, dès lors qu'elles sont « *l'une des composantes identifiables des productions de l'Institut, [...] 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.* »<sup>4</sup> Les esquisses d'une gouvernance des données sont déjà bien dessinées.

C'est toutefois avec la naissance d'INRAE que la question de la gestion des données de la recherche et de la science ouverte acquiert une dimension politique nouvelle. Le document d'orientation « INRAE2030 » réalisant une prospective des ambitions du nouvel institut tranche avec le programme précédent « #Inra2025 » réalisé en 2016. Ce dernier faisait de la science ouverte et du numérique les piliers stratégiques sur le devant de la scène de la recherche agronomique au sein de l'Inra. Or les perspectives ont évolué pour la prospective INRAE2030 : le numérique n'est plus mis en avant comme l'outil aux fondements de la mise au point d'une politique de science ouverte et de gestion des données de la recherche. Il s'agit certes toujours de « *placer la science, l'innovation et l'expertise au cœur de nos relations avec la*

---

<sup>1</sup> DipSO, « Une première année riche d'ouverture. Bilan 2020 », INRAE, juin 2021, p. 2, <https://www6.inrae.fr/dipso/content/download/3649/35674/file/Bilan-DipSO-2020.pdf>.

<sup>2</sup> Odile Hologne. « Données de la recherche : rôle des professionnels IST à l'Inra », *op. cit.*, p. 4.

<sup>3</sup> Inra, « Charte de déontologie de l'Inra », 2018, p. 3, <https://www6.inrae.fr/label-hr-excellence/content/download/3512/33926/version/2/file/charte%2Bde%2Bd%C3%A9ontologie.pdf>.

<sup>4</sup> *Ibid.*, p. 3.

*société pour renforcer notre culture de l'impact* »<sup>1</sup>, et en cela de s'engager « *résolument dans des démarches de science ouverte et participative* » contribuant « *au dynamisme de l'écosystème de recherche et d'enseignement supérieur national, en contribuant aux politiques de site et aux alliances de recherche* »<sup>2</sup>. Mais la question des données de la recherche bascule du côté des grands axes d'orientation scientifique, se faisant « *science des données* » associée aux « *technologies du numérique au service des transitions* »<sup>3</sup>, actant en cela le poids épistémologique du numérique inséré au sein des quatre autres grandes orientations scientifiques se concentrant sur les problématiques environnementales, agroécologiques et alimentaires, la gestion des risques, la bioéconomie et la transition climatique. Le numérique n'est plus seulement un nouvel outil de mise en œuvre de la recherche scientifique, mais bien le milieu<sup>4</sup> dans lequel elle se déploie et qui appelle une scientification de la production des données, une rationalisation à la hauteur des enjeux qui sont les leurs ; une gouvernance des données est donc indispensable.

Qu'est-ce que la gouvernance des données ? Pour Hadi Quesneville, administrateur des données scientifiques à INRAE, il s'agit de « *prendre les bonnes décisions tout au long du cycle de vie des données, ce qui impose de poser les questions sur la base de critères scientifiques, mais aussi juridiques, économiques, techniques et de politique scientifique* »<sup>5</sup>, et à ce titre, de définir les responsabilités de chacun dans le processus décisionnel. Pour ce faire, la Direction pour la Science ouverte, créée lors de la fusion des deux instituts, a organisé par une approche participative une consultation, pilotée par Odile Hologne, des acteurs engagés dans les différentes étapes de la production des données et concernés par leur gouvernance. Après plus d'une vingtaine d'enquêtes, une quinzaine d'entretiens, la réalisation d'un séminaire de travail en février 2020, et à l'appui des travaux du Comité pour la Science Ouverte, la DipSO a rendu public en décembre 2020 ses « *Principes pour la gouvernance des données* »<sup>6</sup> se décomposant en quatre axes : science, gestion des données, réglementation, innovation.

Le premier principe se veut favoriser le partage et la réutilisation des données « *en respectant les valeurs de la science* »<sup>7</sup> ; les maîtres-mots sont déontologie, intégrité scientifique, éthique du chercheur, qui doivent ainsi orienter tous les efforts de production, de gestion, de pérennisation et de réutilisation des données. Ces caractéristiques sont les seules en mesure de faire perdurer les « *valeurs de rationalité et de rigueur collectivement définies* » par un collectif de pensée en environnement de crise des sciences et des relations avec la société. Celles-ci d'ailleurs se fondent notamment sur les données dont la gouvernance en renforce la

---

<sup>1</sup> INRAE, « INRAE2030. Partageons la science et l'innovation pour un avenir durable », *op. cit.*, p. 6.

<sup>2</sup> *Ibid.*, p. 7.

<sup>3</sup> *Ibid.*, p. 20.

<sup>4</sup> Bruno Bachimont, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques. Principes pour une science des données », *op. cit.*

<sup>5</sup> Hadi Quesneville, Esther Dzalé, « Gouvernance des données », Café numérique 2, le 1<sup>er</sup> mars 2021.

<sup>6</sup> Direction pour la Science Ouverte, « Principes pour la gouvernance des données », INRAE, 10 décembre 2020, 8 p.

<sup>7</sup> *Ibid.*, p. 2.

valeur qui dépasse le seul champ scientifique, en devant être considérée « *dans les contextes de l'expertise, de l'appui aux politiques publiques et de la formation* ». Une gouvernance des données advient donc à l'appui d'une valorisation de la recherche, que ce soit pour le chercheur recevant « *la reconnaissance nécessaire à la poursuite de ses travaux* »<sup>1</sup>, ou pour d'autres communautés scientifiques au sein desquelles « *la réutilisation des jeux de données produites par d'autres est encouragée* »<sup>2</sup>, garantissant, par le contrôle des pairs, les conditions épistémologiques de la construction du fait scientifique.

Pour ce faire, la gouvernance des données pose en second principe la nécessité pour les données d'être « *gérées en vue de les rendre F.A.I.R.* »<sup>3</sup>. Les principes FAIR, élaborés par un groupe de chercheurs au milieu des années 2010<sup>4</sup>, sont aujourd'hui des principes recteurs d'une gouvernance aboutie travaillant à rendre les données trouvables (*findables*), accessibles (*accessible*), interopérables (*interoperable*) et réutilisables (*reusable*). En effet, dès lors que ces données sont utiles et nécessaires au travail de la recherche, et sans savoir si elles pourront l'être à nouveau dans le futur, elles doivent être appréhendées comme « *un patrimoine scientifique commun* »<sup>5</sup> dont la « *fairisation* » garantit la pérennisation des données numériques sur le long terme. Il s'agit donc, par ces principes, de mettre en œuvre « *une bonne gestion des données* » encouragée, voire peu à peu obligée, par les financeurs et les politiques publiques.

Les données de la recherche, et par extension tout type de donnée publique, adviennent par une gestion raisonnée peu à peu réglementée et dont le principe « *aussi ouvertes que possible, aussi fermées que nécessaire* » en régit le cadre. A INRAE, sa mise en application se réalise selon un processus établi par la DipSO (annexe 2). Les données sont alors considérées « *par défaut ouvertes et réutilisables gratuitement* »<sup>6</sup> selon les modalités juridiques que nous avons parcourues précédemment, sous réserve toutefois de nombreuses conditions et exceptions limitantes propres à la nature des données (issues d'un partenariat avec le privé, données à caractère personnel, sécurité nationale, *etc.*). La gouvernance des données à INRAE introduit dès lors une graduation de l'ouverture, du vert – les données sont librement accessibles – au rouge – le partage des données est interdit, ou limité au regard d'une réglementation spécifique –, en passant par le segment orange – l'ouverture des données doit être instruite afin d'évaluer le degré de fermeture adéquat au regard de leurs caractéristiques et de la réglementation dont elles dépendent.

---

<sup>1</sup> Hadi Quesneville, Esther Dzalé, « Gouvernance des données », Café numérique 2, le 1<sup>er</sup> mars 2021.

<sup>2</sup> Direction pour la Science Ouverte, « Principes pour la gouvernance des données », *op. cit.*, p. 3.

<sup>3</sup> *Ibid.*, p. 3.

<sup>4</sup> Mark D. Wilkinson, Michel Dumontier, *et al.*, « The FAIR Guiding Principles for scientific data management and stewardship », *Scientific Data*, 2016, doi: 10.1038/sdata.2016.18.

<sup>5</sup> Direction pour la Science Ouverte, « Principes pour la gouvernance des données », *op. cit.*, p. 3.

<sup>6</sup> *Ibid.*, p. 4.



Illustration "Principes pour la gouvernance des données" - INRAE, 2018.

Enfin, le quatrième et dernier principe de la gouvernance des données à INRAE acte leur ouverture comme contribution décisive « à l'innovation et à la création de valeur pour la société »<sup>1</sup>. Cette fin, qu'est « l'innovation », justifiant les moyens, dont se fait la gestion des données, s'inscrit dans le cadre national et européen présenté plus tôt modelant les mouvements d'ouverture comme creuset de l'innovation technologique et économique. Pour le Deuxième Plan pour la science ouverte, abordant notamment les enjeux de l'*open source*, la « *mutualisation des efforts de développement à l'échelle de la communauté scientifique internationale au sens large : académique, industrielle, citoyenne [...] est aujourd'hui un levier essentiel de la recherche et de l'innovation* »<sup>2</sup>. De même INRAE, à la fois inscrit dans le sillon de ces ambitions et à l'initiative même de nombre d'entre elles, présente-t-il sa prospective pour la décennie à venir sous le précepte suivant : « *Partageons la science et l'innovation pour un avenir durable* »<sup>3</sup>. Pour ce faire, l'Institut déploie une gouvernance des données inscrivant l'innovation comme quatrième et dernier principe actant clairement l'orientation finalisée de la science ouverte et du partage des données.

INRAE met ainsi en œuvre une politique proactive sur les enjeux de création, de gestion et de partage des données de la recherche au sein des communautés scientifiques et en dehors. Elle définit à la fois les modalités d'ouverture des données que l'Institut produit mais aussi que celui-ci reçoit, comme le rappelle Hadi Quesneville<sup>4</sup>. Ces principes s'inscrivent donc dans une politique globale de science ouverte que la DipSO, sous la direction d'Odile Hologne, est en train de finaliser par la réalisation d'un plan d'action devant être communiqué en septembre 2021<sup>5</sup>. Ce plan d'action mobilisera un ensemble d'acteurs dont les rôles et les responsabilités sont rappelés dans le document de gouvernance par le schéma suivant.

<sup>1</sup> *Ibid.*, p. 4

<sup>2</sup> « Deuxième plan national pour la science ouverte. Généraliser la science ouverte en France. 2021-2024 », *op. cit.*, p. 17.

<sup>3</sup> INRAE, « INRAE2030. Partageons la science et l'innovation pour un avenir durable », INRAE, document d'orientation, décembre 2020, 52 p.

<sup>4</sup> Hadi Quesneville, Esther Dzalé, « Gouvernance des données », Café numérique 2, le 1<sup>er</sup> mars 2021.

<sup>5</sup> Odile Hologne, entretien du 16 juin 2021.

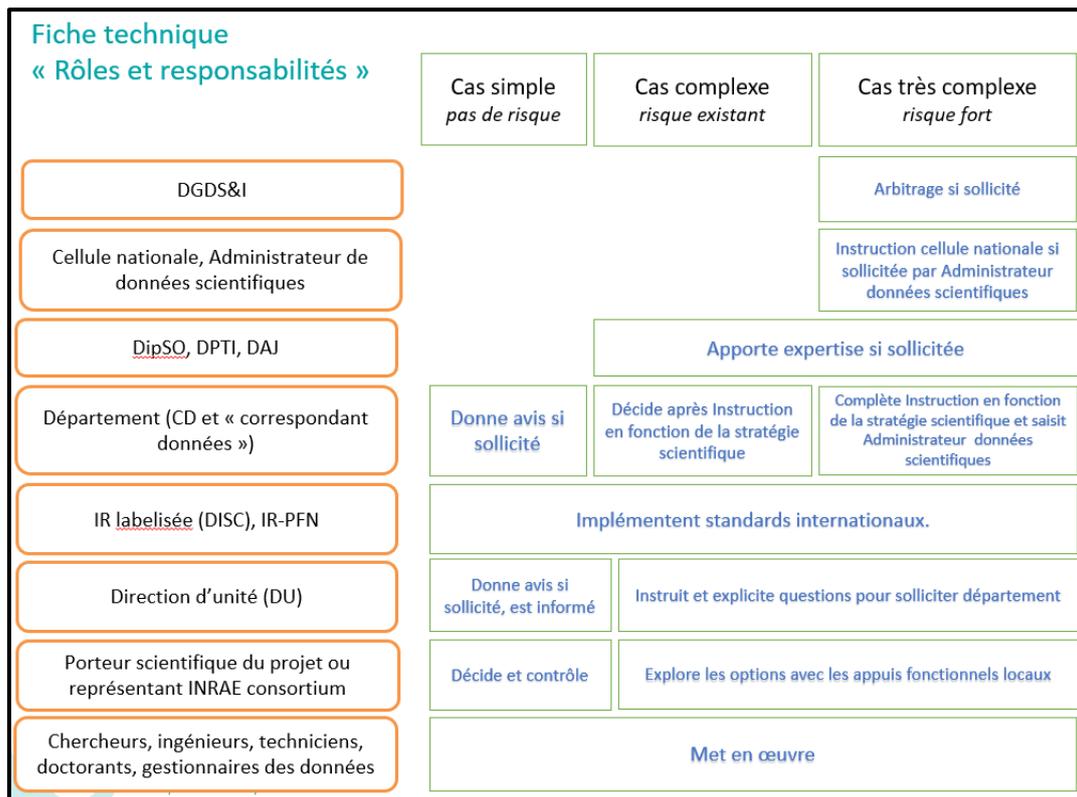


Schéma des rôles et responsabilités des acteurs de la donnée - INRAE, 2018.

Si le chercheur est identifié à l'origine de la production des données, apparaissant le plus à même de connaître ses données et les conditions de mise en œuvre de leur gestion, il est associé à un ensemble d'autres acteurs mobilisés dans une structure hiérarchique afin d'accroître la légitimité des décisions du devenir des données produites. Cette approche des rôles et des responsabilités dans la science ouverte se conçoit par les risques ; l'évaluation et la décision du partage, dans un modèle *bottom-up*, se répercutent progressivement aux échelons supérieurs pour ne pas laisser le producteur de la donnée seul ; la gouvernance des données nécessite donc l'appui indispensable de la direction. L'élaboration d'une politique de la donnée est alors primordiale, dès lors que l'on conçoit et accepte au plus haut niveau hiérarchique que l'environnement complexe, global et interdisciplinaire de la recherche scientifique suppose une gestion raisonnée des données à même de faire sereinement face aux nouveaux enjeux qu'elles imposent. Ainsi, si l'idéal de la science ouverte relève parfois de l'utopie, du mythe qui voudrait, selon Dominique Wolton, « que l'on puisse se passer des intermédiaires et établir une communication "directe" entre le producteur de connaissances et les usagers »<sup>1</sup>, il se confronte rapidement à la réalité complexe où un ensemble d'acteurs, aux côtés des chercheurs, est nécessaire. En effet, il ne peut y avoir « de connaissances sans professions intermédiaires, sans professeurs, journalistes, bibliothécaires, documentalistes, éditeurs... il faut nécessairement des professions qui médiatisent la connaissance »<sup>2</sup>.

<sup>1</sup> Dominique Wolton : « Abondance et gratuité : pour quoi faire et jusqu'où ? Entretien avec Joëlle Farchy, Pascal Froissart et Cécile Méadel », *op. cit.*, p. 34.

<sup>2</sup> *Ibid.*, p. 34.

## L'information scientifique et technique : une reconfiguration plurielle

Depuis le début des années 2000, les métiers de l'IST font face à des changements décisifs de leur environnement professionnel, doublement confronté aux mouvements d'ouverture et aux outils numériques qui leur imposent une critique réflexive sur leurs propres pratiques. « *Les professionnels IST, dit ainsi Odile Hologne en 2017, n'ont plus la parole toute puissante sur les questions des données* »<sup>1</sup>. Ils sont devenus un « maillon » d'une longue chaîne d'acteurs dont le travail sur les données est pour chacun d'entre eux indispensable. Les chercheurs les premiers, producteurs des données, attendent beaucoup d'eux face aux statuts changeant des données de la recherche en contexte de science ouverte, de *data deluge* et de relation renouvelée sciences et société. Le Comité consultatif commun d'éthique pour la recherche agronomique de l'Inra observe ainsi en 2016 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* »<sup>2</sup>. L'Institut répondra aux attentes des chercheurs en créant en 2020 la Direction pour la Science Ouverte au champ d'expertise très large.

Evidemment, les professionnels de l'IST de l'Inra et de l'Irstea n'ont pas attendu la création d'INRAE pour prendre en charge les enjeux des données de la recherche en environnement numérique et ouvert. Odile Hologne elle-même, travaillant dès les années 1990 au CEMAGREF (ex-Irstea) comme chef de projet multimédias sur la mise en place de systèmes experts, appréhendait les enjeux des données sur CD-Rom et finalement, faisait « *de l'hypertexte avant l'heure* »<sup>3</sup>. Mais son orientation vers les nouvelles technologies au sein des métiers de l'IST pouvait détonner au tournant des années 2000. Arrivée à l'Inra en 2004, elle constate en effet que les documentalistes et bibliothécaires, environ deux-cents à cette époque, peinent à légitimer leur place auprès de la direction et des équipes de recherche dont les outils numériques rendaient partiellement caduques leurs missions. En 1980, deux unités centrales de documentation de l'Inra géraient jusqu'à huit-cents titres de revues scientifiques, accueillaienent et accompagnaient en bibliothèque les lecteurs grâce à soixante agents, dont des bibliothécaires, des documentalistes, des magasiniers, des traducteurs, des photographes, bref des métiers participant à l'élaboration directe ou non de la science en train de se faire. Mais à partir des années 2000, « *le chercheur n'a plus besoin du lieu "bibliothèque" ni de la médiation du*

---

<sup>1</sup> Odile Hologne, « Conclusion », Séminaire « Données scientifiques », le 16 février 2017, Paris.

<sup>2</sup> Comité consultatif commun d'éthique pour la recherche agronomique, *Avis 8 sur les enjeux éthiques et déontologique du partage et de la gestion des données issues de la recherche*, Inra, 2016, fihal-02796585, p. 5.

<sup>3</sup> Odile Hologne, entretien du 16 juin 2021.

*bibliothécaire pour accéder à l'information* »<sup>1</sup>. Colette Cadiou au CEMAGREF dresse quant à elle un même constat : certains centres régionaux s'équipent de structures de documentation et de bibliothèques neuves mais à peine usitées tant les pratiques des chercheurs avaient déjà évolué grâce aux outils numériques<sup>2</sup>.

Les professionnels de l'IST de la recherche agronomique traversent donc une période de transition douloureuse en devant recomposer leur offre de service inadaptée aux nouvelles réalités de terrain. Un terrain d'ailleurs initialement très local, installé dans les centres régionaux déployés dès la création respective des deux instituts selon une stratégie scientifique et politique dont l'histoire en a restitué les dimensions<sup>3</sup>. Or au cours des années 2000, les professionnels de l'IST, isolés dans les centres de recherche et ayant peu souvent l'opportunité d'échanger avec leurs collègues, sont peu à peu sollicités sur des projets dépassant la seule échelle locale. Prenant alors conscience que « *c'était eux les mieux placés pour définir leur offre de service, sous couvert de la validation des orientations à la direction* » et acceptant « *de sortir de leur zone de confort, d'explorer de nouveaux territoires* »<sup>4</sup>, ils initient finalement une dynamique collective qui s'empare progressivement des sujets des données de la recherche et de la science ouverte, notamment par le prisme des dépôts institutionnels. L'Inra et le CEMAGREF introduisent ainsi respectivement au milieu des années 2000 « *une reconfiguration de l'IST pour organiser avec davantage d'efficacité la documentation scientifique et la gestion de la donnée* »<sup>5</sup>. Le projet « FuturIST », conduit de 2006 à 2009 par Odile Hologne de l'Inra et Françoise Boudet-Bône du CNRS, offre des perspectives à une profession dont les pratiques traditionnelles sont en déshérence par l'évolution des compétences, le développement de l'offre de services, l'investissement dans de nouveaux outils ainsi que « *la possibilité de nous ouvrir à des partenariats externes* »<sup>6</sup>.

Dans ce cadre porteur, le sujet de la *data* arrive sur la table au cours d'un séminaire qu'Odile Hologne organise en 2007 avec un collègue de l'INIST, Francis André. Il prend par la suite une dimension formelle avec la formation en 2009 du groupe de travail sur la gestion et le partage des données délivrant son rapport trois années plus tard. Celui-ci observe que « *les professionnels de l'IST pourraient assister les chercheurs dans tous les processus relatifs à la construction des métadonnées, la maintenance des référentiels (liste de termes, ontologies) et les questions de propriété intellectuelle (propriété des données, des bases de données, etc.). Les documentalistes disposent d'un socle de compétences et d'outils acquis*

---

<sup>1</sup> Odile Hologne, Françoise Boudet-Bône, « FuturIST : un projet pour accompagner l'évolution des métiers de l'information scientifique et technique », *Cahier des Techniques de l'Inra*, Inra, 2012, p. 5, fihal-01191338f.

<sup>2</sup> Colette Cadiou, entretien le 1<sup>er</sup> juillet 2021.

<sup>3</sup> Pierre Cornu, Egidio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra, entre science et politique*, Versailles, Editions Quae, 2018, 463 p.

<sup>4</sup> Odile Hologne, entretien du 16 juin 2021.

<sup>5</sup> Odile Hologne, *Mise en perspective de l'évolution des métiers de l'IST dans un organisme de recherche*, Frédoc 2013, Octobre 2013, Aussois, France, hal-02810258.

<sup>6</sup> Odile Hologne, Françoise Boudet-Bône, « FuturIST : un projet pour accompagner l'évolution des métiers de l'information scientifique et technique », *op. cit.*, p. 12.

dans la gestion des publications qui est transposable à la gestion des données »<sup>1</sup>. Toutefois, ils ne peuvent pas avoir la « connaissance intime de la nature des recherches, et des jeux de données associés » ; les professionnels de l'IST et les chercheurs doivent donc redéfinir les modalités de leur relation. Surtout, ils se retrouvent confrontés à de nouvelles données qui se caractérisent par leur hétérogénéité. Or, « sans une compréhension des différents types de données et des distinctions entre elles, le traitement qui leur est apporté peut être déficient »<sup>2</sup>. Les professionnels de l'IST doivent ainsi moduler leurs compétences et créer de nouveaux métiers à même de saisir la complexité des enjeux des données de la recherche en environnement numérique. C'est ainsi l'apparition à l'Inra des *data manager* et *data librarian*, dont leur mission, selon le groupe de travail sur la gestion et le partage des données, est « de mettre en relation des données non seulement avec l'amont (métadonnées sur le recueil et le contexte d'acquisition) mais aussi avec l'aval (mise en relation avec des champs de connaissance, des communautés de chercheurs, des attentes sociétales, etc.) »<sup>3</sup>. Les compétences des nouveaux professionnels de l'IST sont alors, pour Odile Hologne, un mélange entre savoirs traditionnels et connaissance de l'environnement numérique<sup>4</sup>. Le défi est conséquent : pour Esther Dzalé-Yeumo, actuellement responsable du pôle Num4Sci de la DipSO, il s'agit pour les professionnels de prendre la voie d'horizons parfois inconnus, tout en étant crédible dans un nouveau domaine d'expertise pour lequel ils doivent prouver leur valeur face à de nouveaux objets parfois simplement à inventer<sup>5</sup>.

Ainsi, bien que les effectifs des services IST aient diminué au cours de ces dernières années, passant à l'Inra de deux-cents personnes au début des années 2000<sup>6</sup> à cent-soixante en 2005<sup>7</sup> pour cent-vingt-et-une en 2017<sup>8</sup> et une soixantaine à la DipSO en 2020<sup>9</sup>, les professionnels ont su réaffirmer ces dernières années leur légitimité, et ce en multipliant les interactions avec des acteurs multiples, convoqués au sein de « réseaux transversaux »<sup>10</sup>, et sans lesquels le discours de l'IST sur le sujet des données de la recherche et de la science ouverte aurait été douteux<sup>11</sup>. La

---

<sup>1</sup> Christine Gaspin, Dominique Pontier, Laurence Colinet, Frédéric Dardel, Alain Franc, *et al.*, *Rapport du groupe de travail sur la gestion et le partage des données*, Avis du conseil scientifique, Inra, juin 2012, p. 38.

<sup>2</sup> Lyne Da Sylva, « Les données et leurs impacts théoriques et pratique sur les professionnels de l'information », *op. cit.*

<sup>3</sup> Comité Consultatif Commun d'Éthique Pour La Recherche Agronomique, *Avis 8 sur les enjeux éthiques et déontologique du partage et de la gestion des données issues de la recherche*, *op. cit.*, p. 20.

<sup>4</sup> Odile Hologne, « Données de la recherche : rôle des professionnels IST à l'Inra », *op. cit.*, p. 34.

<sup>5</sup> Esther Dzalé Yeumo, « Données de la recherche et recherche agronomique », stage ENSSIB, 07 juin 2016, p. 54.

<sup>6</sup> Odile Hologne, entretien du 16 juin 2021.

<sup>7</sup> Odile Hologne, Françoise Boudet-Bône, « FuturIST : un projet pour accompagner l'évolution des métiers de l'information scientifique et technique », *op. cit.*, p.

<sup>8</sup> Odile Hologne. « Quels services IST dans un contexte d'Open Science », *Bibliothèques et Chercheurs : quelles relations, quelle coopération ?*, 2015, p. 7.

<sup>9</sup> DipSO, « Une première année riche d'ouverture. Bilan 2020 », INRAE, juin 2021, p. 2, <https://www6.inrae.fr/dipso/content/download/3649/35674/file/Bilan-DipSO-2020.pdf>.

<sup>10</sup> Emmanuelle Jannès Ober, « Problématique de gestion, de diffusion des données », Séminaire « Données scientifiques », le 16 février 2017, Paris.

<sup>11</sup> Odile Hologne, entretien du 16 juin 2021.

création de la Direction pour la Science Ouverte, à la naissance d'INRAE, est un témoignage fort du nouveau rôle que les professionnels de l'IST ont su construire et s'approprier. L'organigramme, ci-dessous, présente clairement l'entremêlement complémentaire des nouveaux savoirs et savoir-faire aux pratiques de l'information scientifique et technique traditionnelles. La Direction pour la Science Ouverte, dépendante de la Direction générale déléguée à la science et à l'innovation, réunit ainsi les anciennes équipes de l'Inra et de l'Irstea, sous l'autorité couplée d'Odile Hologne, ex-Inra, et d'Emmanuelle Jannès-Ober, ex-Irstea. Cette direction « *organise la gouvernance des données, développe une capacité d'appui et d'aide à la décision du management scientifique (notamment via des services de veille et de scientiométrie) et accompagne l'évolution des compétences et des pratiques dont celles de la fonction IST* »<sup>1</sup>. Pour ce faire, elle se décompose en quatre pôles principaux dont l'organigramme est disponible en annexe (annexe 1) :

- Le pôle ACDC, « Accès, capitalisation et diffusion des connaissances », couvrant les axes de l'édition scientifiques, de la gestion documentaire et de l'archivage du patrimoine scientifique. Il gère les documents définitifs, papiers et numériques, issus de l'activité scientifique, les ressources documentaires nécessaires à la recherche agronomique, mais aussi le cycle de vie des documents d'activité de l'administration. Le pôle est doté de deux archivistes en charge d'engager les processus de traitement archivistique des documents d'activité, ainsi que des données de la recherche de concert avec Num4Sci.

- Le pôle ASTRA, « Analyses pour éclairer la stratégie », engagé sur les enjeux de veille et de scientiométrie. Ce pôle vient en appui de la recherche par ses capacités à appréhender l'environnement scientifique, technique, économique, politique ou encore juridique des projets de recherche.

- Le pôle SenS, « Sciences en société », dont les missions sont d'une part de développer les relations entre l'Institut et la société civile, notamment par le déploiement de relais dans le monde associatif et les organisations non gouvernementales, d'autre part de mettre en œuvre les sciences participatives sur lesquelles INRAE a pris depuis plusieurs années des engagements forts. Le nom du pôle témoigne clairement du virage *impliqué* de la science *en société*<sup>2</sup>.

- Le pôle Num4Sci, « Numérique pour la science », travaillant les enjeux numériques en termes techniques afin d'assister les départements de recherche et les directions d'appui dans la mise en œuvre des bonnes pratiques documentaires et de gestion des données de la recherche. Ce pôle permet d'agir sur les données, en faciliter l'accès, l'exploitation, l'archivage, et de développer de nouvelles compétences en environnement ouvert.

Le déploiement de la politique pour la science ouverte se réalise en outre en étroite relation avec l'administrateur des données scientifiques qui, aux côtés du pôle Num4Sci, met en œuvre la gouvernance des données.

---

<sup>1</sup> « Une direction pour la science ouverte », *inrae.fr*, <https://www6.inrae.fr/dipso/La-Direction-pour-la-science-ouverte>.

<sup>2</sup> Léo Coutellec, *La science au pluriel. Essai d'épistémologie pour des sciences impliquées*, *op.cit.*, 83 p.

Cette Direction de la science ouverte se compose ainsi d'un collectif de travail d'une soixantaine de personnes auquel s'ajoute une douzaine d'associés, chercheurs, techniciens, agents administratifs, affectés à leur unité respective mais qui « *œuvrent pour la science ouverte avec une lettre de mission DipSO à au moins 50%* »<sup>1</sup>. Les trois quarts des membres de la direction relèvent du champ de compétences de l'information, de la documentation, des technologies de l'information et du traitement des données scientifiques (BAP F). Répartis dans dix-sept centres d'INRAE, la DipSO s'engage auprès des unités de recherche où les fonctions d'appuis à la recherche « *sont profondément enracinées* ».

Les actions en faveur de la science ouverte sont plurielles, de l'accompagnement dans la mise en place de nouveaux modèles d'édition scientifique à l'assistance des équipes de recherche dans la mise en place de projets de recherche en environnement ouvert et de démarches de recherches participatives. Son champ d'expertise s'étend en outre du domaine de l'information et de la communication auprès des agents d'INRAE afin de promouvoir les actions d'ouverture par des fiches techniques, sites internet ou portails web, à l'élaboration de méthodologies et d'expertises pour l'engagement de processus de certification. La direction est dotée d'une enveloppe annuelle de trois millions d'euros dont deux tiers sont utilisés pour l'abonnement aux revues scientifiques, aux bases de données comme *Scopus* et *WoS*, ainsi qu'aux outils d'analyse bibliométrique. Le tier restant est consacré au développement des outils et processus de déploiement de la science ouverte au sein d'INRAE, à la formation aux instruments et services tels *DataINRAE*, *HAL INRAE*, le web sémantique, le thésaurus INRAE, *etc.* Pour ce faire, la DipSO travaille en étroite collaboration avec d'autres directions dont la Direction des services d'information (DSI). La fusion de l'Inra et de l'Irstea fut d'ailleurs l'opportunité de créer une relation nouvelle grâce au pôle Num4Sci se faisant l'interface entre les enjeux du numérique d'un point de vue technique et les enjeux de science ouverte en termes méthodologiques.

La Direction pour la Science Ouverte se compose ainsi de professionnels aux compétences hétérogènes pour répondre aux problématiques multiples auxquelles ils sont confrontés. De l'organisation à la préservation des données, en passant par leur *curation*<sup>2</sup>, ils couvrent le cycle de vie des données de la recherche comme appui auprès des communautés scientifiques, celles-ci restant encore trop souvent désemparées face aux bouleversements qu'induisent les mouvements d'ouverture et le numérique.

---

<sup>1</sup> DipSO, « Une première année riche d'ouverture. Bilan 2020 », INRAE, juin 2021, p. 12, <https://www6.inrae.fr/dipso/content/download/3649/35674/file/Bilan-DipSO-2020.pdf>.

<sup>2</sup> Lyne Da Sylva, « Les données et leurs impacts théoriques et pratique sur les professionnels de l'information », *op. cit.*

## Les producteurs de données : une sensibilisation de tous les instants

Lors de son discours de lancement du Premier plan pour la science ouverte, Frédérique Vidal, ministre de l'Enseignement supérieur, de la Recherche et de l'Innovation, exprime l'un des points nodaux consubstantiels à l'essor de la science ouverte : « *Le problème n'est pas principalement technique, il est d'abord, et fondamentalement, humain.* »<sup>1</sup> Si le numérique est le plus souvent pensé par le prisme de l'instrument, et *impensé* dans ses dimensions éthiques et politiques<sup>2</sup>, la science ouverte engage les communautés scientifiques sur une voie réflexive questionnant leurs propres habilités en environnement numérique. Au-delà de la problématique loin d'être anodine de la fracture numérique, révélée par la crise sanitaire du COVID-19 dans des dimensions parfois insoupçonnées au sein même de certains établissements d'ESR<sup>3</sup>, les chercheurs sont confrontés à un changement de paradigme, que ce soit dans le déploiement d'un projet de recherche ou la construction du fait scientifique, qui impose une sensibilisation de tous les acteurs aux enjeux de la science ouverte.

Le Deuxième Plan de la science ouverte, dans la continuité du premier où l'accent était d'ores et déjà mis sur la formation et la transmission des compétences, insiste sur le besoin de « *développer et reconnaître les compétences et métiers de la science ouverte* », et pour se faire, de « *transformer les pratiques pour faire de la science ouverte le principe par défaut* »<sup>4</sup>. Car bon nombre de ses principes applicatifs, ainsi la gestion des données de la recherche, demeurent encore trop souvent de l'ordre du « *bricolage* »<sup>5</sup> selon Marin Dacos, et ce tant du côté des communautés scientifiques que des métiers de l'IST. Dans un champ encore en construction, il importe donc de construire « *des réseaux transversaux* »<sup>6</sup> pour sensibiliser tous les acteurs de la science ouverte. Cette sensibilisation doit faire œuvre de patience et de pédagogie face à des communautés dont l'inertie peut être forte tant la science ouverte réinterroge, parfois en termes épistémologiques, l'activité scientifique, mais aussi son cadre formel de validation dont les limites ont depuis longtemps été franchies. En effet, « *si l'ouverture de la science et des résultats de recherche, notamment sous forme de publication, tend aujourd'hui à remporter une adhésion massive, les chercheurs hésitent encore, et restent réticents*

---

<sup>1</sup> Frédérique Vidal, « Plan national pour la science ouverte : discours de Frédérique Vidal », juillet 2018, mis à jour en octobre 2019, <https://www.enseignementsup-recherche.gouv.fr/cid132531/plan-national-pour-la-science-ouverte-discours-de-frederique-vidal.html>.

<sup>2</sup> Pascal Robert, *Mnémotechnologies. Une théorie générale critique des technologies intellectuelles*, op. cit., p. 310.

<sup>3</sup> Emilie Poirson, responsable du pôle Management Documentaire et Archives de l'INSA-Lyon, témoignage recueilli lors du stage de fin d'étude ENSSIB de mars à juin 2021 à l'INSA-Lyon.

<sup>4</sup> « Deuxième plan national pour la science ouverte. Généraliser la science ouverte en France. 2021-2024 », op. cit., p. 21.

<sup>5</sup> Marin Dacos, « Le plan national pour la science ouverte. Où en sommes-nous ? », dans *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>6</sup> Emmanuelle Jannès Obert, « Problématique de gestion, de diffusion des données », Séminaire « Données scientifiques », le 16 février 2017, Paris.

*lorsqu'il s'agit de donner accès à leurs données de recherche* »<sup>1</sup>. Odile Hologne à INRAE partage un constat semblable, les chercheurs craignant selon elle « *de se faire piller leurs résultats, que certains publient les bons articles avant eux justement en exploitant les données qu'ils auront ouvertes. Et puis il y a la problématique de la reconnaissance et certains considèrent beaucoup qu'il faut publier dans une revue à haut facteur d'impact pour être reconnu dans sa communauté.* »<sup>2</sup> Les problématiques, systémiques à ces niveaux-là, ne peuvent trouver des réponses dans la seule dimension technique des mouvements d'ouverture. Dans ces actions de sensibilisation, « *le savoir être est peut-être plus important que le savoir technique* » pour Emmanuelle Jannès-Ober.

La Direction pour la science ouverte d'INRAE, par les longues années d'expérience de chacun de ses membres, a bien intégré cette dimension intrinsèquement humaine de la science ouverte. Elle a développé ainsi tout un ensemble d'outils permettant d'accompagner les communautés scientifiques dans cette période de transition complexe. Les formations sont ainsi nombreuses pour tous les agents d'INRAE, en différents formats. Les sessions, d'une à deux heures, le plus souvent en ligne, introduisent les enjeux de la science ouverte qui, sur le versant de l'édition scientifique sont bien souvent connus, mais demeurent pour les problématiques des données de la recherche un terrain à apprivoiser. INRAE propose ainsi des formations au plan de gestion des données (PGD), devenant indispensable à tout projet de recherche, à l'utilisation de l'entrepôt de données institutionnel DataINRAE, ou encore aux *data papers* dont le Plan national pour la science ouverte en encourage l'appropriation par la publication d'articles sur les données afin de valoriser les résultats de la recherche<sup>3</sup>. Toutes les formations d'INRAE ont connu durant l'année 2020, en contexte de pandémie, un accroissement de leurs participants, passant de 351 participants pour 77 sessions en 2019 à 696 participants pour 125 séances l'année suivante<sup>4</sup>. Ces actions de formation, primordiales à l'échelle des communautés scientifiques, se complètent de la plateforme DoRANum (Données de la Recherche : Apprentissage NUMérique à la gestion et au partage), développée en 2016 par le réseau des Unités régionales de formation à l'information scientifique et technique (URFIST), l'INIST et la BSN, et avec la contribution d'organismes de recherche comme l'Inra. Les ambitions de ce dernier sont de concourir à l'apprentissage de la gestion des données de la recherche, tant pour les professionnels de l'IST que pour les chercheurs et doctorants.

Les efforts de sensibilisation et de formation se concentrent d'ailleurs fortement à l'échelle des futurs nouveaux chercheurs. A INRAE, une formation bien plus lourde est dédiée aux doctorants de l'Institut, formation dite *Papyrus* entièrement en ligne, initiant les chercheurs en devenir aux enjeux de la science

---

<sup>1</sup> » Bernard Jacquemin, Joachim Schöpfel et Renaud Fabre, « Libre accès et données de recherche. De l'utopie à l'idéal réaliste », *Études de communication* [En ligne], 52 | 2019, DOI : <https://doi.org/10.4000/edc.8468>.

<sup>2</sup> Odile Hologne, entretien du 16 juin 2021.

<sup>3</sup> INRAE, « Formations », [En ligne], consulté le 26 juin 2021, <https://datapartage.inrae.fr/On-vous-signale/Formations>.

<sup>4</sup> DipSO, « Une première année riche d'ouverture. Bilan 2020 », INRAE, juin 2021, p. 6, <https://www6.inrae.fr/dipso/content/download/3649/35674/file/Bilan-DipSO-2020.pdf>.

ouverte en une trentaine d'heures<sup>1</sup>. Bien sûr, le travail de formation ne peut se situer à la seule échelle des doctorants que l'on incite alors par exemple à publier dans des épijournaux. Pour Antoine Petit, cela serait « *trop facile de dire aux jeunes chercheurs : "nous on a fait notre carrière en publiant dans Nature et dans Science et toi maintenant qui débute ta carrière, il suffit que tu publies dans les épijournaux et tu verras tout va bien se passer"* »<sup>2</sup>. Il y a donc une « *valeur de l'exemple* » essentielle à mettre en œuvre auprès des chercheurs confirmés qui appréhendent parfois difficilement les dimensions techniques que recouvre la science ouverte.

La DipSO s'est ainsi dotée d'un forum « *dédié à la science ouverte et au numérique au sens large* »<sup>3</sup> permettant les échanges entre agents des unités sous tutelle INRAE. Son objectif est ainsi de « *dé-siloter les échanges entre membres de réseaux distincts* » et, de cette manière, de participer à l'engagement d'une dynamique à l'échelle de l'Institut, à Paris et en province. Les communautés PEPI, (Partage d'expérience et de pratique en informatique) créées à l'Inra en 2010, participent par ailleurs à ce cadre formateur afin « *d'échanger, partager ses pratiques et ses questions, monter collectivement en compétences, expérimenter et mutualiser des outils ou méthodes, découvrir de nouvelles technologies, se soutenir et s'entraider... Et ainsi aller plus loin, gagner du temps et de l'énergie* »<sup>4</sup>, et ce grâce notamment au développement d'un wiki très riche témoignant de la dynamique d'ouverture des communautés scientifiques d'INRAE.

INRAE met ainsi actuellement en œuvre un ensemble de dispositifs permettant d'accompagner tous les acteurs de la recherche agronomique dans le déploiement de la science ouverte. Un accompagnement qui se traduit par la mise en place d'un réseau resserré de compétences de la donnée par la nomination de référents dits stratégiques ou opérationnels, au niveau des quatorze départements de l'Institut ou à l'appui des unités de recherche, et en mesure de répandre la bonne parole. Cette sensibilisation, doublée d'efforts d'accompagnement, est une nécessité pour l'Institut selon son Président-directeur général Philippe Mauguin<sup>5</sup>, que retranscrivent les nouveaux recrutements au sein d'INRAE. Si au début des années 2000, les chercheurs étaient recrutés sur la base de leurs compétences en génétique, les compétences liées au numérique sont majeures dans les profils aujourd'hui accueillis. La science ouverte est ainsi « *un engagement stratégique* » de l'Institut, qui se fonde sur un ensemble de dispositifs techniques et de management permettant de mettre en œuvre la gouvernance des données.

---

<sup>1</sup> « Parcours d'accompagnement pour publier et communiquer en science (Papyrus) », *ist.Inrae.fr*, <https://ist.inrae.fr/produit/papyrus/>.

<sup>2</sup> Antoine Petit, « La science ouverte, une révolution nécessaire », *La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

<sup>3</sup> « Bienvenue sur Discourse », Forum DipSO, avril 2020, <https://forum.dipso.inrae.fr/t/bienvenue-sur-discourse/7>.

<sup>4</sup> « Bienvenue sur le wiki de la communauté PEPI 2G », INRAE PEPI 2G, <https://pepi2g.wiki.inrae.fr/>.

<sup>5</sup> Philippe Mauguin, « Introduction », Colloque de restitution de l'étude prospective sur la transition numérique, Inra-Agreenium, le 27 juin 2019, Paris, <https://www.inrae.fr/actualites/quelle-transition-numerique-lenseignement-superieur-recherche>.

# LA GESTION DES DONNEES DE LA RECHERCHE : DE L'INFORMATION SCIENTIFIQUE A LA TECHNIQUE

## Le cycle de vie des données à INRAE

La gestion des données de la recherche en environnement ouvert appelle à la mise en place de dispositifs susceptibles de répondre aux enjeux pluriels qui traversent la société dans son ensemble. Les communautés scientifiques, au sens large, doivent donc mettre en œuvre ces dispositifs de traitement des données de la recherche pour garantir en toute fin leur qualité intrinsèque à même d'assurer simultanément un regain de transparence et de confiance au sein de la recherche scientifique et la régénération des relations sciences agronomiques et société. Dans un contexte de *data deluge* et de crises multiples, la question de la qualité de la donnée est vive, d'autant que la puissance de l'environnement numérique doit composer avec sa propre fragilité bien trop souvent impensée. La qualité des données désigne, pour les auteurs de la prospective de la transition numérique à l'Inra, « *l'aptitude de l'ensemble des caractéristiques intrinsèques des données (fraîcheur, disponibilité, cohérence fonctionnelle ou technique, traçabilité, sécurisation, exhaustivité) à satisfaire des exigences internes (pilotages, prise de décision...) et des exigences externes (réglementations...) à l'organisation* »<sup>1</sup>. Nous rajouterions aux exigences internes les enjeux propres au mouvement d'ouverture envers la société engagée dans un nouveau rapport aux sciences qui, peu à peu, se font sciences impliquées. C'est donc une problématique forte pour la recherche scientifique qui nécessite effectivement, d'une part, « *l'interopérabilité de systèmes complexes et des jeux de données qui les composent* »<sup>2</sup>, d'autre part une « *traçabilité qui se manifeste dans de nombreux secteurs scientifiques et économiques : santé et pharmacie, agroalimentaire et grande distribution...* ». Une traçabilité d'autant plus requise pour INRAE que son secteur d'activité couvre les enjeux environnementaux et alimentaires qui, progressivement depuis plusieurs décennies, concentrent l'attention de la société de plus en plus critique sur les méfaits des sciences modernes.

Les communautés scientifiques ont ainsi développé au cours de ces deux dernières décennies des outils de maîtrise des données de la recherche mis en œuvre sous forme de processus ; le cycle de vie des données de la recherche retranscrit cette approche. Le schéma suivant, élaboré par Esther Dzalé Yeumo à l'Inra<sup>3</sup>, s'appuie sur le modèle du *UK Data Service* du Royaume-Uni<sup>4</sup>.

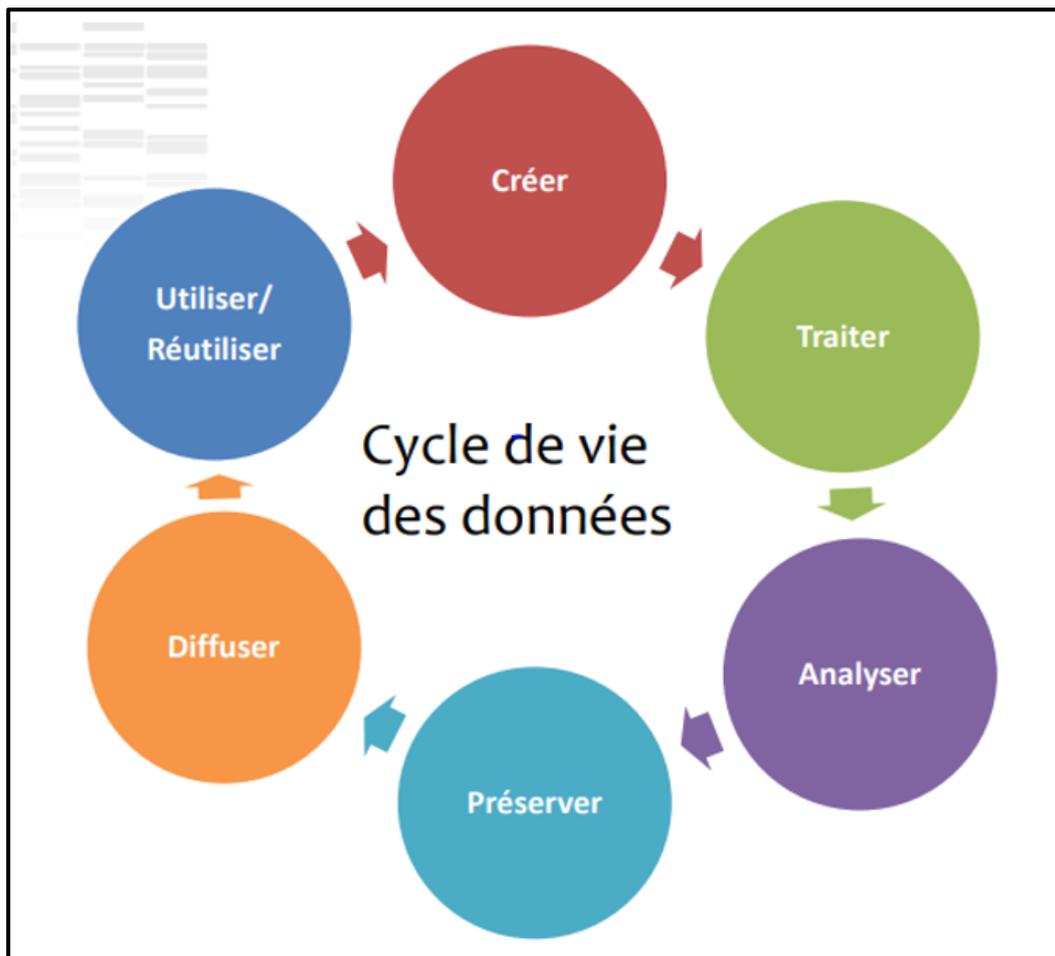
---

<sup>1</sup> *Transition numérique et pratiques de recherche et d'enseignement supérieur en agronomie, environnement, alimentation, sciences vétérinaires à l'horizon 2040*, Résumé de la prospective, Inra, 27 juin 2019, p. 42.

<sup>2</sup> *Ibid.*, p. 43.

<sup>3</sup> Esther Dzalé Yeumo, Dominique L'Hostis, « Open Science. Gestion et partage des données de la recherche », Formation URFIST, 22 janvier 2015, Paris, p. 38, <https://hal.inrae.fr/hal-02800107/document>

<sup>4</sup> UK Data Service, « Research data lifecycle », <https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx>.



Cycle de vie des données de la recherche – Esther Dzalé Yeumo, 2015.

Il se compose de ces six étapes identifiées de la vie des données de la recherche produites en environnement ouvert. La science ouverte se fait en effet ici le cadre de mise en œuvre d'une gestion raisonnée des données, par la mise en application des principes FAIR qui sont la garantie de leur qualité et de leur intelligibilité. Chacune des étapes doit appliquer les principes ou tendre vers eux ; ils sont les piliers pour améliorer « *la gestion, la mise en qualité et l'ouverture des données* »<sup>1</sup>. En effet, les quatre principes (*Findable, Accessible, Interoperable, Reusable*) visent en finalité le partage des données : les données doivent donc être « *faciles à retrouver par tous, sont dans tous les cas récupérables par leur identifiant à l'aide d'un protocole standard de communication, les jeux de données utilisent des métadonnées contextuelles précises, les contenus et formats respectent les standards internationaux, et enfin, les données sont mises à disposition selon une licence explicite et accessible* »<sup>2</sup>. Or, si l'intelligibilité des données de la recherche est, selon les sociologues Jérôme Denis et Samuel Goëta, « *intrinsèquement liée aux*

<sup>1</sup> Batifol V., Burnel L., Cardona A., Johany F., « Cycle de vie des données : un outil pour améliorer la gestion, la mise en qualité et l'ouverture des données », Affiche, Réseau Qualinous & Mission RGPD, département ACT – INRAE, DOI : 10.15454/hsc3-b796.

<sup>2</sup> Inra, *Transition numérique et pratiques de recherche et d'enseignement supérieur en agronomie, environnement, alimentation, sciences vétérinaires à l'horizon 2040*, Résumé de la prospective, 27 juin 2019, p. 43.

*conditions locales de leur production et de leur usage* »<sup>1</sup>, alors leur gestion impose des efforts de tous les acteurs impliqués au cours de leur cycle de vie. C'est en effet un processus durant lequel « *une ou plusieurs personnes interagissent avec les données, de leur création à la publication jusqu'à leur réutilisation ultérieure* »<sup>2</sup>.

Les chercheurs, producteurs des données, sont ainsi les premiers à mettre en œuvre les principes FAIR, la gouvernance des données à INRAE actant le principe selon lequel « *la responsabilité de l'ouverture des données reste portée en premier lieu par le chercheur qui doit en particulier* :

- *Veiller au cadre éthique (valeurs sociétales, valeurs d'établissement, ...) et déontologique (rigueur scientifique, reproductibilité, ...) dans lequel s'inscrit la production et le partage des données ;*

- *Déposer ses données dans des entrepôts disciplinaires ou dans <http://data.inrae.fr> afin de leur attribuer un identifiant unique (ex : DOI, URI, etc.) pour permettre de les citer ;*

- *Décrire clairement les données sous forme de métadonnées pour contribuer à les rendre trouvables (findable) et intelligibles pour les réutilisateurs potentiels. La qualité des métadonnées, comme celle des données doivent être garanties par le chercheur ;*

- *Pouvoir valoriser son travail en préservant sa capacité à le poursuivre. Ainsi lorsque les données sont « ouvrables » selon les critères définis par les autres principes, il est possible d'en différer l'ouverture afin de protéger une exploitation scientifique des données par les partenaires du projet ou pour des raisons de sécurité publique. »<sup>3</sup>*

Ces actions, dont le chercheur se fait le premier acteur, jalonnent donc le cycle de vie des données de la recherche ; il est toutefois accompagné par les professionnels de l'IST dont l'expertise soutient le déploiement de chacune de ces étapes, et ce avant même la création des données. En effet, celle-ci est précédée de l'élaboration d'un plan de gestion des données (PGD) qui présente les modalités de collecte, de traitement, d'organisation, de stockage, de sécurisation, de partage ainsi que de pérennisation des données tout au long de leur cycle de vie. Le PGD devient aujourd'hui un prérequis indispensable à tout projet de recherche financé sur fonds publics (ANR) ou européens, bien que sa mise en œuvre puisse demeurer encore délicate pour les communautés scientifiques. Une étude d'INRAE réalisée en 2020 auprès de ses chercheurs révèle un besoin d'accompagnement de leur part notamment dans la rédaction des plans malgré le déploiement d'outils d'assistance

---

<sup>1</sup> Jérôme Denis, Samuel Goëta. « Les facettes de l'Open Data : émergence, fondements et travail en coulisses », Pierre-Michel Menger, Simon Paye (dir.), *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, Paris, Collège de France, 2017, p. 121.

<sup>2</sup> INRAE, « Cycle de vie des données : un outil pour améliorer la gestion, la mise en qualité et l'ouverture des données », DoRaNUM, publié le 25/06/2021, DOI : 10.13143/GDBG-CF63.

<sup>3</sup> Direction pour la Science Ouverte, « Principes pour la gouvernance des données », INRAE, 10 décembre 2020, p. 2.

et de formation par la DipSO<sup>1</sup>. Les professionnels de l'IST jouent ainsi un rôle primordial dans ce cycle de vie des données, à la fois dans l'assistance des chercheurs, mais aussi en propre, au cours de chacune des autres étapes.

L'étape de création, visible dans le schéma ci-dessus, vise la collecte des données et l'inscription des premières métadonnées, dans le cadre défini par le PGD. L'étape de traitement, fondamental pour assurer l'intelligibilité des données, nécessite de les vérifier et de les valider, ainsi que de compléter leurs métadonnées. Les modalités de stockage et d'échange sécurisé sont alors initiées à moment-là et pour le reste du projet. L'étape d'analyse vient renforcer ces actions indispensables de documentation des données et surtout introduit une première sélection ; cette étape marque la distinction, à INRAE, entre données brutes et données achevées. Les données brutes, dans ce cadre-là, se distinguent des données achevées par le caractère partiel de leur documentation : des « *données personnelles non anonymisées* »<sup>2</sup>, par exemple, doivent être ici supprimées si cela le requiert. Intervient ensuite l'étape de préservation entamant le stockage sécurisé et les actions de pérennisation des données sur le long terme, telles que la structuration et l'organisation des *jeux de données* documentés ; le choix d'un entrepôt est requis de la part du chercheur qui, selon les principes de la gouvernance, doit par ce dépôt initier les actions d'identification pérenne des données. C'est ainsi l'attribution d'un identifiant qui marque l'étape suivante de diffusion, celle-ci permettant de clairement situer une donnée ou un ensemble de données et de réfléchir à leur diffusion au regard de leur nature juridique et des droits d'accès attribués. Enfin intervient au terme du cycle la réutilisation des données, dont les enjeux se situent à la fois à l'échelle des communautés scientifiques, qui peuvent par l'ensemble des actions de documentation qui précède valoriser leurs recherches et leurs résultats, et à l'échelle de la société dont nombre de ses acteurs, économiques, politiques ou associatifs, peuvent se saisir des données à leurs propres fins.

Le cycle de vie des données est ainsi un cycle de *documentation* des données : il vise à transformer des données dites brutes en données achevées. C'est d'ailleurs « *une caractéristique indispensable à vérifier avant toute décision d'ouverture et une condition sine qua non de la mise en "Open data". Elle est gage de fiabilité et de qualité de la donnée : le traitement de la donnée est finalisé et l'on s'est assuré de la complétude des métadonnées et des documents liés à sa traçabilité* »<sup>3</sup>. En environnement de production des données massives, cette distinction est essentielle, dissociant les ensembles caractérisés par leur quantité de ceux appréhendés par leur qualité. Dès 2004, des chercheurs pointaient du doigt la nécessité d'engager la gestion des données de la recherche dans des processus raisonnés, faisant de ces

---

<sup>1</sup> Sylvie Cocard, Dominique L'Hostis, « Bilan d'enquête sur les plans de gestion de données à INRAE », INRAE DipSO, février 2021, p. 9, [https://datapartage.inrae.fr/content/download/3840/40724/version/1/file/PGD\\_EnqueteINRAE-2020\\_Bilan.pdf](https://datapartage.inrae.fr/content/download/3840/40724/version/1/file/PGD_EnqueteINRAE-2020_Bilan.pdf).

<sup>2</sup> Batifol V., Burnel L., Cardona A., Johany F., « Cycle de vie des données : un outil pour améliorer la gestion, la mise en qualité et l'ouverture des données », affiche, Réseau Qualinous & Mission RGD, département ACT – INRAE, DOI : 10.15454/hsc3-b796.

<sup>3</sup> INRAE, « Données de la recherche et science ouverte : le principe d'un accès "aussi ouvert que possible, aussi fermé que nécessaire" », *op. cit.*, p. 2.

pratique « *une part de la méthode scientifique* »<sup>1</sup> Le cycle de vie des données est ainsi un engagement éthique envers les communautés scientifiques et la société, un engagement qui déborde de la notion même de « donnée », ce dont témoigne les principes pour une gouvernance des données d'INRAE : « *Qu'elles soient destinées à être librement accessibles (ouvertes) ou à usage restreint, les données doivent être considérées comme un patrimoine scientifique commun, à l'échelle de l'établissement, qu'il convient de gérer pour les rendre FAIR.* »<sup>2</sup>

## **L'environnement technique de la gouvernance des données à INRAE**

A INRAE, les professionnels de l'IST ont développé au fur et à mesure des années un ensemble d'outils afin de mettre en œuvre le cycle de vie des données de la recherche et d'accompagner le déploiement de la science ouverte au sein de l'Institut. Pour ce faire, ils ont dû restructurer leurs compétences pour créer une combinaison fructueuse de connaissances traditionnelles et de savoirs nouveaux. Selon Odile Hologne, les premiers les positionnent en avant-garde dans l'appréhension des nouveaux enjeux de l'édition, des modalités de citation et d'évaluation, des connaissances juridiques (licence Creative Commons...), vocabulaires, *etc.* ; les seconds impliquent une acculturation des objets numérique tels que les identifiants pérennes (DOI, URI...), les langages de données et de métadonnées (Dublin Core...), les enjeux de leur interopérabilité (OAI-PMH...)<sup>3</sup>. Cette réarticulation des compétences en information scientifique et technique a conduit les professionnels de l'Inra et de l'Irstea à proposer au fur et à mesure des années des boîtes à outils à l'intention des communautés scientifiques afin que celles-ci se saisissent des enjeux de la science ouverte. La naissance d'INRAE permit la fusion de leurs dispositifs, tel le site internet Datapartage, alimenté par la DipSO, mettant à la disposition des membres de l'Institut les services, outils et bonnes pratiques recommandées, afin de les initier à l'élaboration des PGD, la gestion des identifiants ou encore l'utilisation des entrepôts de données, des catalogues de documents et des archives ouvertes.

La plateforme HAL INRAE, dont la trajectoire a été abordée précédemment, est créée en 2019 lorsqu'elle est retenue comme dépôt institutionnel, conduisant en juin 2020 à la migration de l'ensemble des notices et documents des deux instituts. Cette archive ouverte est complétée d'un ensemble d'autres outils dont Bel-INRAE, catalogue des fonds documentaires de l'Institut, qui traverse toutefois une crise existentielle. En effet, selon Raymond Schiano, en charge notamment des questions

---

<sup>1</sup> « *Data archiving is a process, not an end state where data is simply turned over to a repository at the conclusion of a study. Rather, data archiving should begin early in a project and incorporate a schedule for depositing products over the course of a project's life cycle and the creation and preservation of accurate metadata, ensuring the usability of the research data itself. Such practices would incorporate archiving as part of the research method.* » James Jacobs, Charles Humphrey, "Preserving research data", *Communications of the ACM*, vol. 47, n° 29, septembre 2004, p. 29. 10.1145/1015864.1015881, cité dans Dzalé

<sup>2</sup> Direction pour la Science Ouverte, « Principes pour la gouvernance des données », *op. cit.*, p. 3.

<sup>3</sup> Odile Hologne, « Données de la recherche : rôle des professionnels IST à l'Inra », *op. cit.*, p. 34.

d'archivage numérique, si le logiciel libre *PMB* à la source du catalogue est encore performant, c'est tout simplement au niveau des compétences de documentaliste que le bât blesse. Bel-INRAE fédère en effet quatre-vingt-cinq bibliothèques pour 100 000 documents dont 17 500 numériques. Or, « *le métier de documentaliste à l'ancienne*, dit-il, *est un peu en déshérence* »<sup>1</sup> ; il n'y a plus de recrutements de documentalistes en charge spécifiquement des bibliothèques depuis des années et les départs à la retraite ne sont pas remplacés sur ces postes. Les bibliothèques sont ainsi peu à peu fermées et « *il n'y plus personne pour aller chercher un bouquin, une revue à l'ancienne* ». Dès lors, l'avenir du catalogue, qui demeure une ressource essentielle à la recherche, invite à la réflexion dans la prise en charge commune des fonds papier et numériques, d'autant qu'il est interfacé avec les autres ressources d'INRAE dont Aureli (Accès unifié aux ressources électroniques d'INRAE).

Aureli est le moteur de recherche de l'Institut pour le requêtage de ses ressources documentaires, publications, données, information scientifique et techniques. Dans le cadre de l'avènement de la science ouverte, cette interface, gratuite et ouverte à tous, se veut une simplification du travail de recherche des ressources disponibles, que ce soit les ressources externes auxquelles l'institut a un abonnement ou qui sont en libre accès (revues, dépôt d'éditeur, bases de données bibliographiques, archives ouvertes, *etc.*), ou les objets internes ayant suivi un processus de documentation, documents numériques ou papier, produits multimédias, ainsi que données de la recherche. Ces dernières, d'ailleurs, sont stockées au sein de l'entrepôt institutionnel DataINRAE.

Un entrepôt de données est « *une interface que l'on peut consulter et interroger et qui est capable de stocker, gérer, maintenir et préserver les données ou objets numériques* »<sup>2</sup>. S'ils peuvent être généralistes ou disciplinaires, INRAE a fait le choix d'un entrepôt institutionnel dans la continuité immédiate des plateformes d'origine des deux instituts. Certains chercheurs, ou équipes de recherche, peuvent faire le choix de recourir à des plateformes web dédiées, bien que les exigences de pérennité et de fiabilité de ces interfaces soient hautement discutables. Dès 2016, l'Inra alertait ses agents, qui étaient « *fortement encouragés à déposer leurs écrits en texte intégral dans cette archive institutionnelle et à les diffuser librement et sans restriction sous réserve de disposer des droits de diffusion. Cette pratique doit être privilégiée par rapport au dépôt dans les plateformes de type « réseau social » tel que Research Gate ou Academia* »<sup>3</sup>. Ces interfaces, faussement gratuites, sont détenues par des entreprises privées qui font de la vente des données leur fonds de commerce, et dont le modèle de partage du savoir interroge dans une dynamique d'*open science*.

---

<sup>1</sup> Raymond Schiano, entretien du 1<sup>er</sup> juillet 2021.

<sup>2</sup> « *Repository (aka Data Repository or Digital Data Repository) is a searchable and queryable interfacing entity that is able to store, manage, maintain and curate Data/Digital Objects.* » Définition du groupe d'intérêt RDA Data Foundation and Terminology Interest Group cité par Esther Dzalé Yeumo, « Entrepôts de données. Pierre angulaire du partage des données scientifiques », ANF Renatis, 8 juillet 2016, <https://datapartage.inrae.fr/content/download/3429/35028/version/1/file/Entrepots-ANFRenatis.pdf>.

<sup>3</sup> Inra, « Charte pour le libre accès aux publications et aux données », *op. cit.*, p. 5.

Les entrepôts de données institutionnels sont ainsi à privilégier, ce qu'a choisi INRAE avec DataINRAE. Celui-ci est fondé sur l'outil DataVerse développé par Harvard ; particulièrement utilisé aux Etats-Unis et peu à peu adopté en Europe<sup>1</sup>, INRAE, en choisissant cette solution logicielle, rejoint un environnement porteur qui répond dès lors aux enjeux d'interopérabilité fondamentaux en contexte de science ouverte. La plateforme se décompose en trois axes principaux : gestion, partage et recherche. DataINRAE a en effet pour fonctionnalité première la gestion des données passant par leur dépôt, leur documentation par métadonnées, leur stockage et archivage, leur liaison à d'autres ressources internes ou externes, et finalement le contrôle du processus de publication des données, normalisant en cela les actions aux échelles désirées. L'axe Partage passe par l'identification pérenne, renforçant les actions de citation et de traçabilité des données, pour en finalité notamment produire les *data papers*. Enfin l'axe Recherche engage par l'ensemble de ces processus une meilleure lisibilité et visibilité des données de la recherche ; les métadonnées qui leur sont associées sont indexées, facilitant dès lors leur connaissance et leur réutilisation. Les principes FAIR se situant au centre du cycle de vie des données de la recherche, DataINRAE est développé afin d'être en conformité avec ceux-ci, avec la réglementation en vigueur et les attentes des financeurs publics<sup>2</sup>.

L'ensemble de ces actions se répartissent en deux fonctions principales au sein de la solution logicielle : la fonction entrepôt « *pour stocker, organiser et documenter les données dès leur collecte* »<sup>3</sup> et la fonction annuaire « *qui favorise la découverte, l'accès et la citation des données dès leur publication* ». La première fonction permet de constituer des *datasets* ou jeux de données, qui sont décrits par des métadonnées, identifiés par un DOI, et citables comme objet unique. Aux niveaux supérieurs, les jeux de données peuvent être rassemblés pour former des *dataverse*, organisant et gérant des ensembles de *datasets*, propres à des projets ou collectifs. La gestion des droits d'accès devient ici primordiale sur les actions de dépôt, de documentation, de modification et de publication des données<sup>4</sup>. La seconde fonction de DataINRAE, celle d'annuaire, appréhende la solution logicielle comme le moyen de décrire les données (grâce à des langages de métadonnées normalisés tel Dublin Core), d'identifier et de citer les données (grâce à l'attribution de DOI et la génération de citations), enfin de les rechercher et de les découvrir, que ce soit pour les communautés scientifiques ou pour tout internaute dès lors que les droits d'accès le permettent. Tous les objets numériques ne peuvent toutefois être stockés, gérés et recherchés au sein de DataINRAE ; ainsi, l'institut recommande « *l'usage d'une forge logicielle institutionnelle pour le code* »<sup>5</sup>, tandis que sa description à

---

<sup>1</sup> Esther Dzalé Yeumo, « DOI & Portail Dataverse : Retour d'expérience de l'Inra », Séminaire « Données scientifiques », le 16 février 2017, Paris.

<sup>2</sup> Fanny Dedet, Anne-Sophie Martel, Windpouire Esther Dzale Yeumo, Sylvie Cocaud, « Le portail Data Inra et les services associés », *Cahier des Techniques de l'Inra*, Inra, 2018, ffhal-02622475f.

<sup>3</sup> *Ibid.*, p. 2.

<sup>4</sup> *Ibid.*, p. 3.

<sup>5</sup> DataINRAE, « FAQ », consulté le 27 juillet 2021, [https://ist.blogs.inrae.fr/dataInrae-guide/faq/#Type\\_de\\_contenu](https://ist.blogs.inrae.fr/dataInrae-guide/faq/#Type_de_contenu).

proprement parler peut l'être dans l'entrepôt de données. Elle sera alors documentée par un ensemble de métadonnées, et identifiée grâce à un DOI afin de la rendre citable et référençable en liaison avec le code lui-même.

Le rôle de l'identifiant est donc fondamental dans la gestion des données de la recherche, permettant à chaque objet numérique d'être individualisé dans l'environnement ouvert au sein duquel il navigue. S'il existe de nombreux identifiants, de l'URI sur le web à l'ARK pour les objets de la BNF, DataINRAE utilise les DOI grâce à une API, interface applicative dédiée à la création et la gestion des DOI à partir de la solution logiciel DataCite. L'attribution de DOI, activée dès la publication du *dataset*, peut se faire sur des objets très divers, des données de la recherche aux articles scientifique, en passant par les logiciels, les objets physiques, la littérature grise, ou encore les PGD eux-mêmes dont la version finale peut être déposée dans DataINRAE qui le liera aux *datasets* associés. Le DOI crée ainsi « *un lien pérenne vers une ressource identifiée, ce qui permet d'augmenter sa visibilité, de faciliter sa découverte, de fiabiliser sa citation et d'améliorer sa traçabilité* »<sup>1</sup>, et *in fine*, de relier les ressources entre elles.

Cette perspective de liaison des ressources se retrouve par ailleurs dans le projet de thésaurus. Entré en service en mars 2021, ce thésaurus INRAE s'appuie sur les travaux de l'Irstea depuis les années 1990<sup>2</sup> pour construire aujourd'hui un référentiel commun à toutes les applications d'INRAE, dans l'objectif d'indexer des contenus textuels, des jeux de données, des pages webs, *etc.*<sup>3</sup>. S'il était répandu de penser que les moteurs de recherche pouvaient maintenant se suffire à eux-mêmes, les avantages des thésaurus sont dans l'information scientifique et technique à nouveau mis à profit pour créer les bonnes affiliations, les relations justes entre ressources par l'apposition de mots-clefs. Pour ce faire, en environnement numérique et de web sémantique, des URI sont attribués aux 15 289 concepts répartis en 12 domaines et 64 microthésaurus<sup>4</sup> ; ils permettent ainsi de créer des identifiants uniques et pérennes, interopérables, dans l'idéal entre l'ensemble des dispositifs d'INRAE. Le rôle du documentaliste, au sens plein du terme, est ici particulièrement visible : il s'agit très concrètement, selon les mots de Colette Cadiou, de « *documenter les données* »<sup>5</sup>, de leur donner la profondeur documentaire de leur insertion dans un univers d'informations dont le numérique en permet la liaison.

Enfin, parmi tous les outils développés par INRAE, un projet est fondamental dans la mise en œuvre d'une gestion des données de la recherche engagée dans la

---

<sup>1</sup> Fanny Dedet, Anne-Sophie Martel, Windpouire Esther Dzale Yeumo, Sylvie Cocard, « Le portail Data Inra et les services associés », *op. cit.*, p. 8.

<sup>2</sup> Colette Cadiou, entretien du 1<sup>er</sup> juillet 2021.

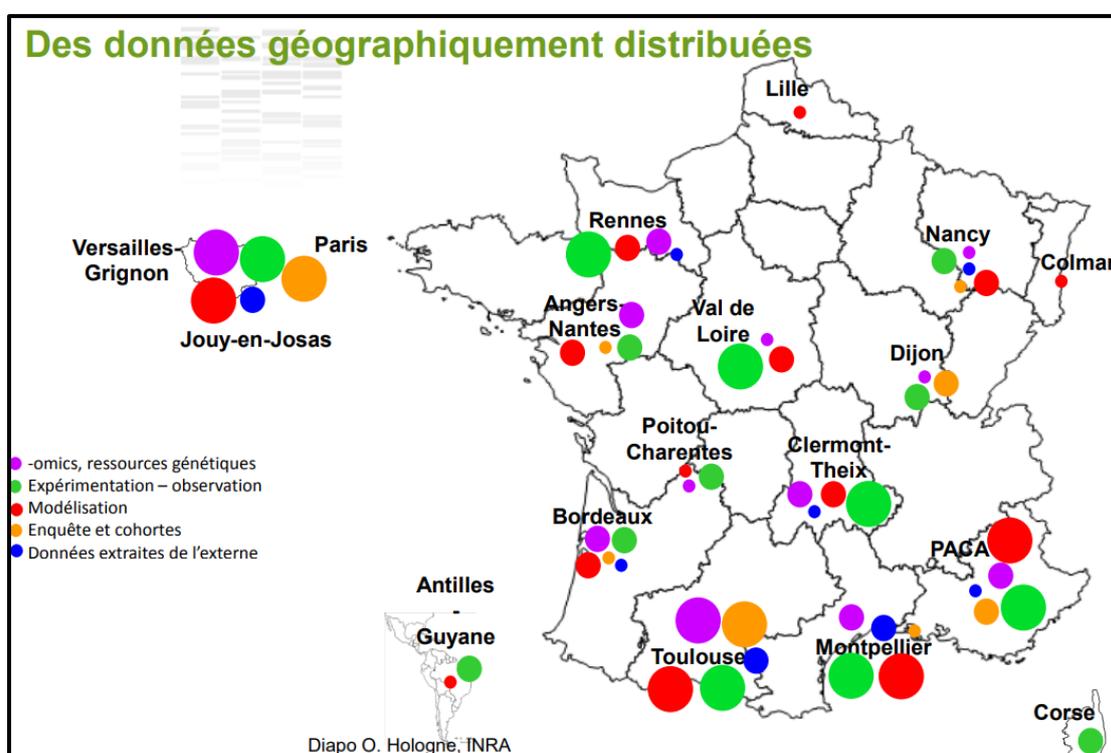
<sup>3</sup> Sophie Aubin, « Le thésaurus INRAE, une nouvelle ressource pour l'Institut », *Vocabulaires Ouverts @ INRAE*, le 29 mars 2021, consulté le 15 juin 2021, <https://vocabulaires-ouverts.inrae.fr/2021/03/29/le-thesaurus-inrae-une-nouvelle-ressource-pour-linstitut/>.

<sup>4</sup> « A propos du thésaurus INRAE », *Vocabulaires Ouverts @ INRAE*, consulté le 15 juin 2021, <https://vocabulaires-ouverts.inrae.fr/a-propos-du-thesaurus-inrae/>.

<sup>5</sup> Colette Cadiou, entretien du 1<sup>er</sup> juillet 2021.

science ouverte et à même de faire face aux enjeux des *data deluge* : AgroDataRing. Issu des résultats d'une étude du groupe de travail « e-infra Storage » en 2016, le projet AgroDataRing se veut « répondre aux besoins actuels et futurs de stockage moyen et long terme des données scientifiques » grâce à la « co-construction d'une infrastructure partagée et mutualisée »<sup>1</sup> de stockage des données de la recherche, et ce « dans le cadre d'une démarche collaborative en partenariat entre plusieurs Départements ». La masse croissante des données, ingérable à l'échelle de l'individu, impose en effet l'élaboration d'une solution technologique particulièrement pérenne.

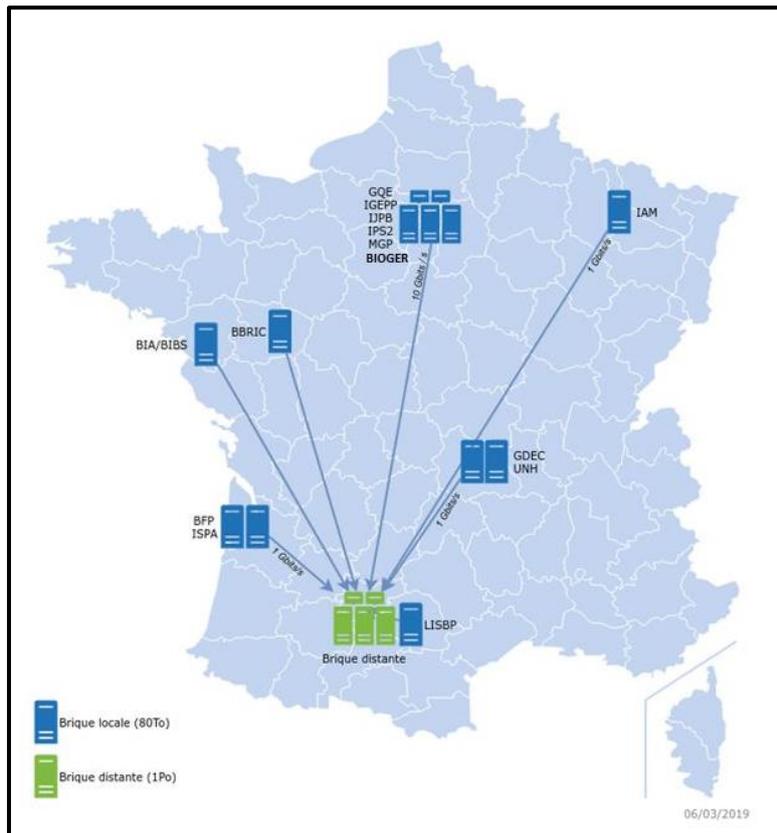
Les données étaient encore en 2016 géographiquement réparties au siège et dans chacun des centres régionaux sur les infrastructures de stockage locales et les *datacenters* de l'Inra en Ile-de-France, ce qu'illustre la carte suivante distinguant en outre la typologie des données.



Distribution géographique des données à l'Inra - Esther Dzalé Yeumo, 2016.

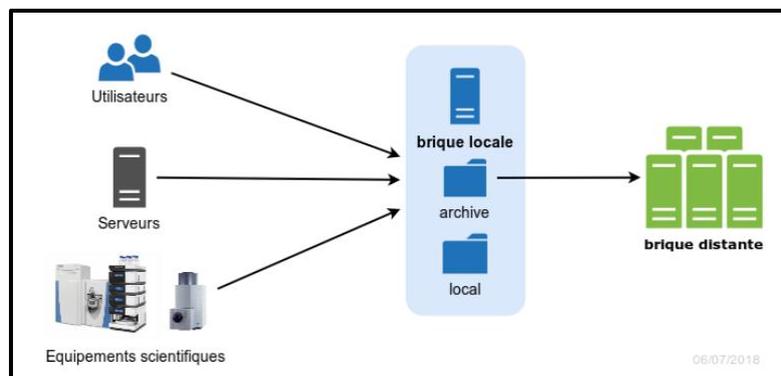
La solution AgroDataRing propose, quant à elle, l'infrastructure de stockage illustrée ci-dessous, faite de briques locales installées à proximité des producteurs de données, dans un contexte à la fois local/régional à hauteur de 30% de l'infrastructure, et de briques distantes, DataCenter INRAE (pour 70% de l'infrastructure).

<sup>1</sup> Pierre Adenot, Stéphane Bansard, David Benaben, Veronique Brunaud, Christophe Caron, *et al.*, « AgroDataRing : une infrastructure partagée et mutualisée pour le stockage longue durée », *Cahier des Techniques de l'Inra*, Inra, 2018, p. 2.



Distribution géographique de l'infrastructure AgroDataRing - INRAE, 2020.

La volumétrie des briques locales s'élève à 80 To ; elles engageant des actions de réplication sur les briques distantes à la volumétrie individuelle de 1 Po, pour un coût annuel au To de l'ordre de 45 €<sup>1</sup>. Le flux des données se réalise comme suit :



Flux des données - « Fiche technique AgroDataRing », INRAE, 2020.

Par ailleurs, l'approche collaborative n'est pas anodine, permettant de mettre en œuvre une solution qui se veut « évolutive »<sup>2</sup> afin de garantir la pérennisation des données stockées avec la contribution des communautés scientifiques soumises à des évolutions rapides de leur environnement de recherche.

Toutefois, cette infrastructure de stockage, proposant une solution pour le recueil des mégadonnées, doit s'accompagner d'une réflexion politique de définition

<sup>1</sup> *Ibid.*, p. 5.

<sup>2</sup> *Ibid.*, p. 8.

des « *données stratégiques* » de l'Institut. En effet selon Odile Hologne, « *pour des questions d'incertitudes à ce stade sur ce qu'il nous faut réellement archiver* »<sup>1</sup> parmi les données de la recherche, rien n'a encore pour le moment abouti. Il existe certes une large offre de solutions logicielles externes proposées par INRAE à ces agents selon leurs besoins et attentes : la GED *Nextcloud*, la solution intranet *SharePoint* de Microsoft – toutefois peu réputée en interne pour son ergonomie et ses performances<sup>2</sup> –, l'outil *ForgeMIA* pour les codes sources, et bien d'autres<sup>3</sup>. De même, parallèlement aux engagements de la France dans le Plan national pour la science ouverte, INRAE s'est investi au sein de plusieurs programmes internationaux dans le développement d'*e-infrastructures* en mesure de prendre en charge, à l'échelle européenne notamment, la gestion des données de la recherche. « *L'objectif est que toutes les infrastructures qui produisent, manipulent, traitent et/ou échangent des données puissent être interconnectées de façon sécurisée, tout en veillant à respecter la souveraineté européenne.* »<sup>4</sup> Le programme *European Cloud Initiative*, qui se décline en l'*European Data Infrastructure* et l'*European Open Science Cloud* (EOSC) est initié en 2016 par la Commission européenne et financé à hauteur de 600 millions d'euros. Son ambition est de « *changer la donne en matière de science ouverte en Europe* » afin de faire de celle-ci « *le véritable leader de la science ouverte et le lieu le plus propice à l'innovation* »<sup>5</sup>. INRAE participe ainsi au programme en mettant en œuvre une interface de test liée au programme EOSC-Pillar 5b afin d'archiver les jeux de données provenant de DataINRAE vers les serveurs du CINES<sup>6</sup> déjà utilisés dans le cadre de HAL-INRAE. L'interfaçage AgroDataRing et DataINRAE, quant à lui, a été envisagé dès 2019<sup>7</sup>.

Mais toutes ces réflexions sont cependant encore récentes, d'autant qu'un travail a été récemment engagé avec l'administrateur des données d'INRAE sur la meilleure qualification de ces données stratégiques. En effet, « *quand on aura une meilleure idée de la façon de considérer nos données sous cet angle stratégique, explique Odile Hologne, on sera davantage capable de raisonner l'archivage numérique à long terme en se focalisant justement sur ce qui est important et pas sur tout. Parce qu'il ne faut pas tout garder.* »<sup>8</sup> Le stockage des données de la recherche n'est pas leur archivage, imposant donc une réflexion à multiples niveaux pour engager la pérennisation des données qui doivent impérativement « *être considérées comme un patrimoine scientifique* »<sup>9</sup> à préserver.

---

<sup>1</sup> Odile Hologne, entretien du 16 juin 2021.

<sup>2</sup> Raymond Schiano, entretien du 1<sup>er</sup> juillet 2021.

<sup>3</sup> INRAE, « Stocker les données – Etats des lieux », Datapartage, <https://datapartage.inrae.fr/Gerer/Stockage-les-donnees/Contexte-Etat-des-lieux>.

<sup>4</sup> Alain Beretz, *Stratégie nationale des infrastructures de recherche*, 2018, p. 3.

<sup>5</sup> T. Skordas, « European Open Science Cloud Council Conclusions », *Digital Single Market Blog Posts*, 2018, cité dans Violaine Rebouillat, *Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs*, thèse de doctorat en sciences de l'information et de la communication, Conservatoire national des arts et métiers, 2019, p. 84.

<sup>6</sup> Raymond Schiano, entretien du 1<sup>er</sup> juillet 2021.

<sup>7</sup> « AgroDataRing », <https://www.ingenum.inrae.fr/adr>.

<sup>8</sup> *Ibid.*

<sup>9</sup> Direction pour la Science Ouverte, « Principes pour la gouvernance des données », *op. cit.*, p. 3.

# LES SCIENCES AGRONOMIQUES EN PERSPECTIVE : HISTOIRE ET EPISTEMOLOGIE EN ENVIRONNEMENT NUMERIQUE

---

## FAIRE ARCHIVE : INTENTIONNALITE ET TRACE DE LA DONNEE PATRIMONIALE SCIENTIFIQUE

### Archives des sciences, archives de chercheur, archives de la recherche

Les données de la recherche, telles qu'elles ont été abordées jusqu'à présent au sein d'INRAE, se caractérisent principalement par leur gestion élaborée selon les principes de la science ouverte en environnement numérique. Cette gestion raisonnée advient par l'organisation d'une gouvernance des données à même de les engager un cycle de vie consolidant leur cadre de création, de traitement, d'analyse, de conservation, de diffusion et de réutilisation qu'il suppose. Ce cycle de vie maintient ainsi les données de la recherche dans une boucle de sollicitations continues, aux fondements des sciences agronomiques *impliquées*, et dont le principe majeur et recteur est leur préservation selon un ensemble de modalités et de dispositifs qui en assure leur pérennisation sur le long terme. Il ne s'agit pas seulement de créer les données, pour ensuite les stocker et enfin les ouvrir ; la gestion des données de la recherche suppose un travail pointu dont les principes et la méthodologie empruntent certes à l'information scientifique et technique, mais aussi à l'archivistique, interrogeant dès lors leur statut dans le paysage des archives contemporaines.

*« Les archives sont l'ensemble des documents, y compris les données, quels que soient leur date, leur lieu de conservation, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité. »<sup>1</sup>*

Cette définition légale des archives selon le Code du patrimoine introduit la notion de donnée depuis la modification de la loi LCAP du 7 juillet 2016<sup>2</sup>. Particulièrement large dans ces acceptations, cette approche fait de tout document une archive à sa création, sans distinction entre le support matériel de manifestation et la forme sémiotique d'expression<sup>3</sup>. En complétant explicitement le terme de document de celui de donnée, l'environnement numérique est reconnu légitime dans la production, la diffusion et la lecture de document. En effet, *« qu'il soit papier ou électronique, explique Céline Guyon à la suite de cette redéfinition, un document est porteur d'un ensemble d'informations déterminées par un contexte et un*

---

<sup>1</sup> Article L211-1 du Code du patrimoine, [https://www.legifrance.gouv.fr/codes/article\\_lc/LEGIARTI000032860025/](https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000032860025/).

<sup>2</sup> Article 59 de la loi n° 2016-925 du 7 juillet 2016 relative à la liberté de la création, à l'architecture et au patrimoine, [https://www.legifrance.gouv.fr/loda/article\\_lc/LEGIARTI000032855756/2016-07-09](https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000032855756/2016-07-09).

<sup>3</sup> Bruno Bachimont, *Patrimoine et numérique : technique et politique de la mémoire*, op. cit., p. 17.

*environnement de production* »<sup>1</sup>. Les données de la recherche, dans ce cadre, sont ces ensembles d'informations dont la production passe par un « *principe de manipulation, appliqué à un contenu, qui en reconfigure le rapport au sens et les conditions de son interprétation* »<sup>2</sup> Un principe qui, de fait, réinterroge en profondeur le rapport au document, c'est-à-dire à l'enregistrement d'un événement qui en garde la trace, la mémoire. Elles partagent ainsi certaines caractéristiques communes avec les archives traditionnellement appréhendées, ce que Lyne Da Sylva liste de la sorte :

« • *Elles sont produites par un organisme (un chercheur ou un groupe de recherche) dans l'exercice de ses fonctions.*

• *Elles sont un produit unique, comme une pièce d'archive ; toute production ultérieure n'est pas une "réédition", mais un nouveau produit unique.*

• *Quand de multiples versions sont créées, le chercheur peut être appelé à décider quelle version est la version officielle, laquelle doit être conservée, etc.*

• *Elles ne sont pas publiées en exemplaires multiples (bien qu'elles puissent être téléchargées de multiples fois) ; en fait, elles ne sont pas généralement publiées dans le sens traditionnel de la publication de livres ou d'articles.*

• *Elles doivent être interprétées dans leur contexte de production.*

• *Elles sont vouées à être conservées et gérées à long terme (curated) pour une période déterminée (ou indéterminée).*

• *Comme préconisé par l'approche d'archivistique intégrée, on reconnaît que la gestion des données doit être pensée dès leur conception.* »<sup>3</sup>

Ces caractéristiques, qui recourent les éléments parcourus jusqu'à présent, laissent à penser que les sciences de l'archive, de l'archivistique à la diplomatique<sup>4</sup>, ont bien un rôle à jouer dans la gestion des données de la recherche. Nous l'avons déjà explicité plus tôt, la section Aurore de l'AAF définissait en 2014, après plusieurs mois d'après discussions, les données de la recherche comme « *l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. A ce titre, elles constituent une partie des archives de la recherche* »<sup>5</sup>. Définies en ce sens, le travail de caractérisation des données de la recherche pourrait alors être clos, à les considérer comme « archives de la recherche ». Mais dès lors que l'on entre sur le terrain de la science, les concepts, les notions, les pratiques s'entremêlent. En effet, qu'est-ce qu'une archive de la recherche ?

---

<sup>1</sup> Céline Guyon, « Le concept d'archives : d'une définition à une autre », 21 novembre 2016, hal-01401138.

<sup>2</sup> Bruno Bachimont, *Patrimoine et numérique : technique et politique de la mémoire*, op. cit., p. 77.

<sup>3</sup> Lyne Da Sylva, « Les données et leurs impacts théoriques et pratique sur les professionnels de l'information », op. cit.

<sup>4</sup> Marie-Anne Chabin, « Peut-on parler de diplomatique numérique ? », Valentine Frey, Matteo Treleani (dir.), *Vers un nouvel archiviste numérique*, Paris, L'Harmattan, 2013, p. 108 à 129.

<sup>5</sup> Julien Pomart, « AAF / Section Aurore: Un groupe de travail sur les données de la recherche », op. cit.

Sa définition est loin d'être si aisée d'autant que la recherche scientifique recouvre des *mondes* faits de collectifs de pensée contraires, de groupes sociaux divers, de pratiques et de techniques structurantes, de contraintes politiques, économiques et financières, bref d'histoires complexes et enchevêtrées dont les archives qui en restituent les trajectoires ne peuvent être en conséquence monolithiques. En France, Thérèse Charmasson l'explique, de premières réflexions institutionnelles émergent dès les années 1960 sur le sujet des archives des sciences<sup>1</sup>. Elles sont d'ailleurs appréhendées par le terme de « *source de l'histoire des sciences* », une définition éloignée de la manière de les traiter aujourd'hui qui entretient le lien entre les enjeux de leur conservation et leur utilité scientifique, juridique ou administrative qu'elles pourraient toujours avoir. La question de leur gestion produit plusieurs rapports au cours des années 1970 pour finalement, durant la décennie 1980 prendre une nouvelle envergure au sein des organismes de recherche eux-mêmes. Marqués par les commémorations de leur création, que ce soit l'Institut Pasteur, le Conservatoire national des arts et métiers (CNAM) ou encore le CNRS, ce sont les débuts des considérations patrimoniales des produits de la recherche faisant écho aux nouveaux intérêts pour les sujets du patrimoine et de la mémoire<sup>2</sup>.

Aux « archives scientifiques », dont l'attribut est ambigu, sont préférées les « archives des sciences », soit « *tout document d'archives permettant d'étudier l'évolution générale des politiques de recherche et d'enseignement scientifiques, l'évolution de telle ou telle discipline scientifique particulière ou encore l'apport de tel ou tel scientifique au développement des connaissances* »<sup>3</sup>. Elles peuvent donc recouvrir les dimensions administrative et scientifique de la recherche, mais aussi politique et sociale afin d'en restituer la complexité par une histoire et une sociologie des sciences renouvelées qui se détachent des perspectives traditionnelles d'appréhension du savoir. En effet, les années 1980 sont aussi marquées par les travaux de chercheurs qui étudient les sciences à l'aune d'un ensemble de concepts tranchant avec une histoire linéaire, sinon cyclique<sup>4</sup>, qui proposait une perspective cumulative des savoirs extraits de tout environnement politique, et dont n'étaient retenus que les styles de pensée victorieux par la domination de leur régime de véridicité, pourtant historiquement situé, sur tous les autres. Il s'agit alors de saisir le processus de la recherche dans toute sa complétude, et de le faire à la manière de Bruno Latour distinguant, telles les deux faces de Janus<sup>5</sup>, la science faite, celle des manuels et des savoirs achevés, de la science en train de se faire, celles des doutes, des errements et des controverses. Or, plutôt qu'« archives des sciences », d'aucuns parlent des « archives des savoirs » qui peuvent introduire un autre rapport à la trace,

---

<sup>1</sup> Thérèse Charmasson, « Archives scientifiques ou archives des sciences : des sources pour l'histoire », *La revue pour l'histoire du CNRS*, 14, 2006, mis en ligne le 03 mai 2008, consulté le 13 août 2021, DOI : <https://doi.org/10.4000/histoire-cnrs.1790>

<sup>2</sup> Pierre Nora, *Les lieux de mémoire*, Paris, Gallimard, 1984, 3 vl.

<sup>3</sup> Thérèse Charmasson, « Archives des sciences », *Conservation et valorisation du patrimoine des organismes de recherche*, Saint-Etienne, Publications de l'Université de Saint-Etienne, 2007, p. 27

<sup>4</sup> Thomas S. Kuhn, *La structure des révolutions scientifiques*, Paris, Flammarion, 1983, 284 p.

<sup>5</sup> Bruno Latour, *Le métier de chercheur. Regard d'un anthropologue*, Paris, Editions INRA, 2001, p. 12.

parfois « *cohérent avec les représentations qu'ont les chercheurs de leurs archives de recherche* »<sup>1</sup> selon Margot Georges, doctorante à l'Université d'Angers et archiviste. « *Ainsi pour eux il est plus important de conserver "le savoir", le produit de leur travail, que la "recherche", le processus qui aboutit au savoir. Les archives des savoirs seraient donc les tirés-à-part, articles et publications qui sont bien souvent disponibles en bibliothèque* ». Bien que les « *contours terminologiques des "archives des savoirs" [soient] dessinés par des positionnements divers* »<sup>2</sup>, c'est une définition incomplète lorsqu'on envisage une histoire des sciences par le prisme des processus de recherche.

Seront ainsi travaillées les « archives de la recherche », qui se détachent du produit fini pour en comprendre l'élaboration, pour en restituer le processus auquel contribuent les données de la recherche, soit les traces qui, usuellement invisibilisées dans la science faite, semblent pouvoir prétendre en environnement ouvert et numérique au statut d'archives. Ces archives peuvent en cela rejoindre les « archives de chercheurs » qui invitent à appréhender, selon l'historien Jean-François Bert, « *l'univers des savants par la prise en compte de la dynamique des savoirs* »<sup>3</sup>. En effet, « *se confronter à une archive de chercheur consiste à entrer dans une démarche, à expérimenter l'espace de l'autre, à l'éprouver en essayant de décrire le plus finement possible des séries de gestes, de discours, de pratiques, d'agencements, mais aussi les stratégies mobilisées dans le temps par un chercheur pour donner sens à ses réalisations théoriques* ». Tous les chercheurs évidemment ne permettent pas d'établir ce rapport, *a posteriori*, avec les archives. Denis Poupardin le remarquait à l'Inra chez certains chercheurs, « *les archives de leurs prédécesseurs et les leurs constituaient "un trésor de guerre" auquel ils étaient fort attachés. Mais pour d'autres, les archives étaient associées à l'idée de vieilleries sales et encombrantes, dont il convenait de se débarrasser au plus vite* »<sup>4</sup>. Les archives de chercheur interrogent donc la distinction entre l'archive privée, personnelle, et l'archive publique, institutionnelle, dans le champ de la recherche scientifique, avec évidemment des écarts d'appréciation significatifs entre les sciences de la nature et du vivant et les sciences humaines et sociales. Thérèse Charmasson le relevait en 2007, les archives personnelles des scientifiques étaient jusqu'à très récemment le plus souvent considérées comme archives privées au sens juridique du terme, quand bien même elles étaient produites au sein des organismes de recherche. Toutefois, note-t-elle, « *pour la période la plus contemporaine, les archives personnelles des scientifiques semblent se confondre avec les archives des laboratoires et il apparaît souvent de plus en plus difficile d'établir une distinction formelle entre archives personnelles et archives de laboratoires* »<sup>5</sup>.

---

<sup>1</sup> Margot Georges, « Institutions de recherche, chercheurs et archives des savoirs : quelle(s) définition(s) par les producteurs ? », Jean-François Bert, Marc Ratcliff (dir.), *Frontière d'archives : recherches, mémoires, savoirs*, Paris, Editions des archives contemporaines, 2015, p. 36.

<sup>2</sup> *Ibid.*, p. 36.

<sup>3</sup> Jean-François Bert, *Qu'est-ce qu'une archive de chercheur ?*, Marseille, OpenEdition Press, 2014, p. 71.

<sup>4</sup> Denis Poupardin, « Communication », Colloque « Archorales » du 2 décembre 2008, Inra, *op. cit.*

<sup>5</sup> Thérèse Charmasson, « Archives des sciences », *op. cit.*, p. 35.

L'environnement ouvert et numérique impose ainsi au chercheur un nouveau rapport au document et à la donnée, un rapport qui l'extrait d'une relation exclusive et individuelle à ceux-ci pour le réintroduire dans un univers de flux et de partage.

Dès lors, lorsque l'Inra en 2016 fait des données « *un patrimoine qu'il faut gérer et valoriser* »<sup>1</sup>, il les fait traces des activités de la recherche agronomique qu'il importe de préserver comme archives de la recherche, afin non plus seulement de les *communiquer dans l'espace*, mais aussi de les *transmettre dans le temps*.

## **Des archives orales au Comité d'Histoire : INRAE et ses archives**

Les enjeux actuels de l'archivage à INRAE viennent pour partie des préoccupations mémorielles de l'Inra, qui émergent au cours des années 1990. L'Irstea a, lui aussi, développé un vif intérêt pour la préservation de ses archives, notamment numériques, mais l'Inra, par son envergure scientifique et politique dans le champ de la recherche agronomique, engage un travail d'introspection peu à peu conséquent. Pourtant, nous le disions en introduction, l'Institut présente des « *lacunes archivistiques très importantes* »<sup>2</sup> comme le rapportent les auteurs de *l'Histoire de l'Inra, entre science et politique*. Des lacunes caractérisées par « *la perte de fonds archivistiques importants, le retard des versements aux Archives nationales ou départementales, et enfin l'extrême dispersion et l'absence d'inventaire des archives encore conservées par l'Inra* ». Cette situation résulte notamment de l'inondation du siège de l'Inra à Paris en mai 2000<sup>3</sup>, mais aussi, au cours de son histoire, des opérations immobilières successives, des déménagements et des réaffectations de locaux, ou encore de l'organisation originelle de l'Institut composé de plusieurs centres régionaux à l'autonomie affirmée sur la gestion administrative des archives, et où le degré de conscience dans le souci de les conserver était très variable<sup>4</sup>. Toutefois, toutes ces situations traduisent une absence explicite et, au cours des années 1990 assumée, de vision politique sur la problématique des archives, qu'elles soient papier ou numériques.

Pourtant, dès 1989, Bruno Galland, conservateur général aux Archives nationales, est chargé de réaliser une enquête dans les centres d'Île-de-France. Le rapport final réalise un état des lieux de l'archivage et délivre de premières recommandations qui, bien qu'elles n'aient pas été suivies en tant que telles<sup>5</sup>, permirent le versement de premiers fonds d'archives au centre des archives contemporaines de Fontainebleau en 1990, dont ceux de personnalités distinguées

---

<sup>1</sup> Inra, « Charte pour le libre accès aux publications et aux données », *op. cit.*, p. 5.

<sup>2</sup> Pierre Cornu, Egizio Valceschini, Odile Maeght-Bournay, *L'histoire de l'INRA, entre science et politique*, *op. cit.*, p. 12.

<sup>3</sup> Florence Descamps, *L'historien, l'archiviste et le magnétophone : De la constitution de la source orale à son exploitation*, Paris, Institut de la gestion publique et du développement économique, 2005, nouvelle édition, DOI : <https://doi.org/10.4000/books.igpde.104>.

<sup>4</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>5</sup> Fanny Audous, « L'archivage dans les centres d'Île-de-France de l'INRA », Christine Cazenave, Françoise Girard, *Conservation et valorisation du patrimoine des organismes de recherche*, Saint-Etienne, Publications de l'Université de Saint-Etienne, 2007, p. 89.

de l'histoire de l'Inra (Jacques Poly, Jean Rebeschung). Cette première initiative toutefois ne tombe pas définitivement dans l'oubli et est suivie trois années plus tard d'un appel, dans les colonnes mêmes de l'*Inra mensuel*, de deux conservateurs du patrimoine des Archives nationales demandant à la direction de l'Institut de se saisir sérieusement du sujet<sup>1</sup>. L'institution nationale ne reste ainsi pas indifférente au sort des archives de l'Inra qui, au même titre que tout document public, sont soumises à la loi n° 79-18 du 3 janvier 1979.

La même année, la direction de l'Inra charge Denis Poupardin, économiste de formation et directeur de recherche au département d'économie et de sociologie rurales de l'Institut, d'entamer une réflexion sur les enjeux de l'archivage par une lettre de mission appelant à définir, « *en liaison avec l'administration des Archives nationales, une politique de collecte et de conservation de ses archives administratives et scientifiques* »<sup>2</sup>. Son étude vise ainsi à étudier les objectifs à atteindre et les moyens à mettre en œuvre préalables à l'élaboration d'une politique d'archivage au sein de l'Institut, d'autant plus nécessaire que « *les archives de l'Inra constituent la "mémoire" de l'institution et permettent la transmission des savoirs et savoir-faire* »<sup>3</sup> : leur perte ou leur destruction serait véritablement préjudiciable à la fois pour la recherche agronomique et pour l'histoire de celle-ci. La stratégie de Denis Poupardin est alors de proposer la mise en place progressive d'une politique d'archivage de concert avec les départements de recherche qui, par leur autonomie, pouvaient « *avoir un rôle important à jouer dans la conservation de leurs résultats* »<sup>4</sup>. Pour ce faire, l'économiste va à la rencontre des chercheurs qui lui délivrent des témoignages précieux pour appréhender les enjeux de gestion inhérents aux archives de la recherche. Toutefois, la direction fait le choix de ne pas poursuivre en l'état les conclusions de son rapport. Une forme de politique d'archivage est bien élaborée au cours de l'année 1995 avec l'aide de la direction des Archives de France<sup>5</sup>, produisant pour les centres de recherche une documentation d'appui ; mais l'état des fonds d'archives travaillés au cours des années 2000 et 2010 par les historiens laissent à penser que la mise en œuvre fut délicate.

Et c'est bien le constat paradoxal que nous a dressé l'historien Pierre Cornu. L'Inra a selon lui « *une obsession de son identité en tant qu'institut agronomique, une culture des pères fondateurs, des grandes figures, de l'esprit "maison"* »<sup>6</sup> que témoignent notamment les premiers versements réalisés en 1990. Mais cette obsession ne passe pas par l'archive, qui n'a jamais été appréhendée que « *jusqu'à très tardivement comme un support valide de mémoire, et encore moins d'histoire* » ;

---

<sup>1</sup> Texte de Chantal Boucher et Mireille Jean, rapporté par Denis Poupardin, Communication, Colloque «Archorales» du 2 décembre 2008, Inra, *op. cit.*

<sup>2</sup> « Lettre de mission du 5 mai 1993 », cité dans Mireille Jean, Denis Poupardin, « Les archives des unités de recherche : le point de vue des scientifiques de l'Institut national de la recherche agronomique (I.N.R.A.) », *op. cit.*, p. 28.

<sup>3</sup> Denis Poupardin, *Une politique d'archivage à l'Inra ?*, rapport d'étude, 1994, 118 p.

<sup>4</sup> Denis Poupardin, « Communication », Colloque «Archorales» du 2 décembre 2008, Inra, *op. cit.*

<sup>5</sup> Fanny Audous, « L'archivage dans les centres d'Île-de-France de l'INRA », *op. cit.*, p. 90.

<sup>6</sup> Pierre Cornu, entretien du 11 juin 2021.

au contraire, elle « *passé par l'oralité* ». En effet, au moment même où on entend commémorer les cinquante ans de l'Institut, notamment par la publication d'un ouvrage qui en retrace l'histoire – écrit par un économiste, ce n'est pas anodin<sup>1</sup> –, la question de l'archive entre par une porte étonnante dans les murs de l'Inra : la captation des témoignages oraux des membres de l'Institut.

« *C'est dans ce contexte particulier, et avec l'idée que dans notre Institut, c'était étrangement les écrits qui s'envolaient et les paroles qui restaient* »<sup>2</sup>, qu'est née la mission Archives orales (Archorales). Une mission de collecte d'archives orales qui, de 1994 à 2004, compile 700 heures de témoignages des membres de l'Inra, et depuis à continuer son œuvre : en 2011, ce sont ainsi 357 témoignages enregistrés pour 284 versements aux Archives nationales. 110 témoignages ont fait l'objet d'un travail de numérisation, réalisé dans le cadre d'un des plans nationaux de numérisation du ministère de la Culture<sup>3</sup>. Ce fonds d'archives orales interroge ainsi, au même titre que les données de la recherche, la pérennisation des archives de la recherche en environnement numérique. D'ailleurs, par cette collecte, Denis Poupardin entend « *aider à faire resurgir dans la mémoire des résultats de recherche tombés fâcheusement dans l'oubli* »<sup>4</sup> ; un objectif dont les préoccupations actuelles de la science ouverte se font les échos. Toutefois, le projet ambitionne surtout de palier une documentation dispersée ou perdue en constituant directement un fonds d'archives qui, par l'oralité, est un des meilleurs moyens « *pour éveiller les témoins à une conscience historique, pour acclimater une démarche patrimoniale et historique dans une organisation* »<sup>5</sup>, selon l'historienne Françoise Descamps. Il s'agit en effet d'inscrire « *la présence d'archivistes-historiens dans la durée et dans la familiarité* » par la création d'« *un premier lieu d'échange et de confrontation* » propice au développement d'instances prenant en compte et en charge les enjeux de l'archivage et, par extension, de la mémoire de l'Institut.

En 2005, la directrice de l'Inra Marion Guillou co-écrit l'édito du onzième tome de la revue *Archorales* dédiée depuis le premier volume à la publication de quelques-uns des nombreux entretiens retranscrits. Valorisant le travail de Denis Poupardin, alors persuadé, selon les mots des auteurs, « *que la place qui serait reconnue plus tard à un organisme de recherche dans l'histoire des sciences risquait fort d'être déterminée in fine par le soin qu'il aurait mis à conserver et à classer ses propres archives* »<sup>6</sup>, la direction de l'Institut manifeste par ce texte la prise de conscience de l'importance des « *différentes sources, écrites, orales, photographiques* »<sup>7</sup> grâce auxquelles « *les historiens pourront conduire leurs*

---

<sup>1</sup> Jean Cranney, *Inra, 50 ans d'un organisme de recherche*, Paris, Inra, 1996, 526 p.

<sup>2</sup> Denis Poupardin, « Communication », Colloque « Archorales » du 2 décembre 2008, Inra, *op. cit.*

<sup>3</sup> Christian Galant, « Archives orales INRAE », INRAE, 20 octobre 2011, mis à jour le 14 avril 2021, <https://www6.inrae.fr/comitedhistoire/ARCHORALES>.

<sup>4</sup> *Ibid.*

<sup>5</sup> Florence Descamps, *L'historien, l'archiviste et le magnétophone : de la constitution de la source orale à son exploitation*, *op. cit.*

<sup>6</sup> Marion Guillou, Bernard Hubert, « Editio », *Archorales*, tome 11, INRA, 2005, p. 1.

<sup>7</sup> *Ibid.*, p. 1.

*investigations avec rigueur et nous aider à développer un regard critique sur les mythes et les “vérités de tout temps” qui habillent notre espace professionnel* ». La direction de l’Inra, alors engagé au même moment dans la commémoration de son sixième anniversaire, comprend ainsi que l’entreprise de sauvegarde de la mémoire de l’Institut ne peut se contenter des seuls témoignages oraux. Il importe alors de développer un véritable souci de l’archive pour nourrir le travail des historiens. Un travail d’autant plus nécessaire que l’Inra accuse un retard certain vis-à-vis d’autres organismes de recherche, tels l’Inserm et le CNRS qui ont sollicité l’aide d’historiens pour l’écriture de leur histoire. L’Inra, actant par les mots de Marion Guillou le caractère historiquement situé des sciences agronomiques, s’ouvre ainsi à la recherche historique en créant un Comité d’Histoire.

Fusionnant le souci du témoignage et celui de l’archive, le Comité d’Histoire de l’Inra est créé en 2005 afin de participer à la collecte et à la conservation des fonds d’archives institutionnels et scientifiques. La mission Archorales n’est certainement pas étrangère à cette création, Florence Descamps soulignant que la collecte des archives orales permet « *d’obtenir le soutien tactique de la communauté des “anciens”, dont certains continuent de graviter autour de l’organisation savent faire passer les “messages” opportuns au moment où il faut* »<sup>1</sup>. En effet, la singularité de ce comité est d’être formé d’anciens cadres de la maison, directeurs de recherche, responsables de centre, retraités qui, ayant pris conscience des enjeux mémoriels et historiques de l’archivage, cherchent à développer leur propre stratégie de préservation des traces de l’activité de l’Institut. Or, Pierre Cornu le souligne, l’Inra est une « *maison fondée sur le principe de loyauté* »<sup>2</sup> ; les membres de ce comité, même retraités, avaient directement accès au bureau de Marion Guillou – et ont toujours aujourd’hui cette relation à la direction – afin d’initier des projets et de solliciter des moyens. Un dialogue étroit s’est donc immédiatement installé entre la direction générale et le Comité d’Histoire. Celui-ci intègre par ailleurs d’autres acteurs, avec la présence dès 2005 des historiens Christophe Bonneuil, Jean-Luc Mayaud, et Thérèse Charmasson, et d’une archiviste Fanny Audous<sup>3</sup>.

Toutefois, si la Fonction Archives est mise en place en 2004 au sein de l’Inra<sup>4</sup>, notamment à la suite du recrutement d’une archiviste l’année précédente, l’Institut éprouve quelques difficultés à pérenniser le poste dont les recrutements fluctuent au fur et à mesure des années, à l’image de bien d’autres organismes de recherche. En 2010, encore peu d’établissement d’ESR « *sont dotés d’archivistes pour ordonner, référencer et ranger ces documents stockés dans de bonnes conditions alors qu’ils représentent des kilomètres linéaires* »<sup>5</sup>. A l’Inra, il faut attendre le milieu de la décennie pour voir un véritable engagement politique sur le sujet des archives, à

---

<sup>1</sup> Florence Descamps, *L’historien, l’archiviste et le magnétophone : de la constitution de la source orale à son exploitation*, op. cit.

<sup>2</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>3</sup> « Composition du Comité d’Histoire », INRAE, <https://www6.inrae.fr/comitedhistoire/Presentation/Composition>.

<sup>4</sup> Fanny Audous, « L’archivage dans les centres d’Île de France de l’INRA », op. cit., p. 89.

<sup>5</sup> Christian Michau, « Avant-propos », *La gestion des archives au sein d’un établissement d’enseignement supérieur et de recherche*, Paris, AMUE, 2010, p. 2.

l'appui du Comité d'Histoire présidé depuis 2011 par l'économiste Egizio Valceschini qui en fait une instance devant contribuer à répondre aux problématiques qui traversent la recherche agronomique en temps de crise.

Le Comité d'Histoire renoue un dialogue étroit avec les Archives nationales en 2016 par la présence du conservateur du patrimoine et responsable du pôle agriculture et PTT Henri Pinoteau<sup>1</sup>. L'Inra, au même moment, stabilise les recrutements en développant la Mission Archives marquée par l'arrivée d'une nouvelle archiviste Lina Sbeih qui s'engage dans le traitement des années d'arriérés d'archives et dans un travail de sensibilisation des unités de recherche et des départements aux enjeux de l'archivage<sup>2</sup>. Dès lors, deux visions parallèles se développent, et parfois s'opposent : la logique administrative de l'archivage qui confronte l'Inra aux enjeux de la gestion réglementée des documents d'activité, tenant compte de leur cycle de vie dans l'objectif de la mise en application de leur sort final ; la logique de valorisation de l'archive portée par les ambitions du Comité d'Histoire d'écrire l'histoire et de valoriser la mémoire de l'Institut. Toutefois, la Mission Archives fait face à un Everest d'archives à traiter ; c'est ainsi le Comité d'Histoire, par sa proximité avec la direction, qui se fait l'instance légitime sur l'archive définitive. Egizio Valceschini, Pierre Cornu et Odile Maeght-Bournay s'engagent alors dans la constitution de fonds documentaires leur permettant d'écrire *L'histoire de l'Inra : entre science et politique*.

Ce chantier, qui les occupe de 2015 à 2018, les conduit à entreprendre eux-mêmes la collecte des archives de l'Institut en l'absence d'inventaires et de versements réguliers aux Archives. Pourtant, même à l'issue de leur enquête historique, les auteurs observent qu'« *il demeure difficile de se faire une idée précise du corpus potentiel pour une histoire de l'Inra* »<sup>3</sup>, soulignant en cela l'ampleur des travaux qui restent à entreprendre. Le corpus qu'ils ont constitué pour leurs propres besoins s'est certes « *avéré riche d'une extrême diversité, foisonnant d'archives manuscrites ou électroniques, d'ouvrages, de revues, de documents imprimés aux statuts divers, d'images, de photographies, pour beaucoup encore jamais exploités* »<sup>4</sup>. Mais Pierre Cornu nous le raconte, ils se sont heurtés, dans cette enquête historique, au caractère multisupport de ces archives très contemporaines, posant tout un ensemble de questions sur le statut des documents en environnement numérique, sur les données personnelles accessibles dans certaines archives, soit finalement des problématiques que le Comité d'Histoire, sans l'appui des historiens, n'avait pas l'habitude de traiter.

La fusion de l'Inra et de l'Irstea, entamée à la fin de la décennie, marque alors un tournant majeur pour la question de l'archivage. En effet, l'Irstea a, selon Pierre Cornu, une culture fondamentalement différente de la documentation et de l'archive.

---

<sup>1</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>2</sup> Lina Sbeih, Fanny Dedet, Patrick Moreau, Esther Dzale, « L'archivage des données de la recherche à l'Inra. Eléments de réflexion, démarche et perspectives », *Cahier des Techniques de l'INRA*, INRA, 2020, p. 4, fihal-02861909f.

<sup>3</sup> Pierre Cornu, Egizio Valceschini, Odile Maeght-Bournay, *L'histoire de l'Inra, entre science et politique, op. cit.*, p. 12.

<sup>4</sup> *Ibid.*, p. 13

Se faisant historiquement l'organe de recherche du ministère de l'agriculture sur les questions territoriales, l'Irstea « *s'est vécu comme la mémoire de l'action publique* »<sup>1</sup> portée par le corps des ingénieurs généraux du génie rural des eaux et forêts. Ce corps d'ingénieurs, très ancien dans la République, est en effet « *le bras armé de l'Etat dans toutes les politiques territoriales depuis des décennies* », estimant dès lors qu'ils sont les dépositaires de cette mémoire de l'action publique dans les territoires. Les collections de l'Irstea sont ainsi riches de nombreuses données, de cartes, d'images satellitaires, de cadastres, de photographies du reboisement des terrains de montagne au XIX<sup>e</sup> siècle qui côtoient celles des cours d'eaux actuels, bref, de documents qui attestent « *de la connaissance intime de chaque parcelle du territoire national* ». Or, autant les agronomes se fondent sur « *le modèle de la culture du champ de blé, vous le semez à l'automne ou au printemps, vous récoltez l'été et puis c'est fini : on est sur un cycle annuel* » ; les forestiers, eux, « *raisonnent par demi-siècle* ». Dès lors, ce qui induit cette relation différenciée à l'archive, c'est finalement le rapport à l'histoire et au temps.

Bien sûr, à l'instar de l'Inra, l'Irstea a tardé à développer en propre une Fonction Archives. Mais, confrontée au développement massif des outils du numérique, la question prend forme au cours des années 1990, notamment sous l'impulsion de Raymond Schiano, recruté en 1984 au CEMAGREF et formé, au fur et à mesure de sa carrière, à la gestion des systèmes documentaires d'information scientifique et technique et à l'archivage numérique. Dès 2000, le documentaliste participe notamment au groupe de travail « *Systèmes d'informations, informatique et forêt méditerranéenne* » interrogeant le devenir des « *informations, observations, références et expertises* »<sup>2</sup>, appuyant les décisions de gestion des forêts et devant être « *collectées, organisées et vérifiées afin d'être opérationnelles et de permettre de réaliser des états des lieux et d'en assurer la mise à jour nécessaire* ». La gestion des données de la recherche apparaît bien dès lors primordiale pour permettre d'appréhender, dans le temps long, ces objets d'étude. Mais c'est à partir de la fin des années 2000, début de la décennie suivante, que l'Irstea s'investit politiquement dans les enjeux de l'archivage. Raymond Schiano en 2012 produit un travail de recherche sur « *l'archivage des documents engageants à l'Irstea* »<sup>3</sup>, dont les réflexions portaient doublement sur les documents papier et numériques d'une équipe de recherche. Il y travaille alors les notions de *records management*, étudie la norme OAIS, engage une réflexion sur la conservation des bases de données, des mails, et la mise en place des bonnes pratiques documentaires et notamment les recommandations en matière de format de fichier (PDF/A). Ses efforts, parallèles à l'engagement d'un dialogue avec le SIAF depuis 2010, débouchent sur l'élaboration d'une politique d'archivage signée en 2015 par la direction, initiant dès lors la mise en place d'une gestion raisonnée des archives de la recherche. Celle-ci, telle qu'elle

---

<sup>1</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>2</sup> « *Conclusions du groupe de travail "Systèmes d'informations, informatique et forêt méditerranéenne"* », *Forêt méditerranéenne*, t. 21, n° 1, mars 2000, p. 131.

<sup>3</sup> Raymond Schiano, *L'archivage des documents engageants à l'Irstea d'Aix-en-Provence : propositions pour le traitement homogène des dossiers papier et numériques au sein d'une équipe d'experts*, Théorie de l'information, CNAM, 2011-2012

est définie, vise tant « *la gestion courante d'un service* » afin de « *disposer en permanence des informations utiles* », que la sauvegarde de la mémoire, « *en constituant les matériaux de l'histoire* »<sup>1</sup>. L'Irstea s'engage ainsi avec force sur les enjeux de l'archivage, et notamment de l'archivage numérique, nourrissant la DipSO de son expertise à la naissance d'INRAE.

## **INRAE, une nouvelle dynamique transverse**

Sans être de l'ordre du mariage de la carpe et du lapin, l'historien Pierre Cornu souligne que la fusion de l'Inra et de l'Irstea fut l'occasion pour chacun d'entre eux de se nourrir des savoirs et savoir-faire de l'autre<sup>2</sup>. Certes, en termes scientifiques, le premier a avalé le second, alors que le rapport à la donnée de la recherche est particulièrement développé à l'Inra qui se positionnait en fer de lance de la science ouverte. Mais en termes de maîtrise de la documentation et de l'archivage, l'Irstea apporte une expertise reconnue en environnement papier et numérique, équilibrant les apports de la DipSO comme appui à la recherche. Ainsi, la responsable du pôle Accès, capitalisation et diffusion des connaissances (ACDC), où se concentrent les enjeux de l'archivage et de la valorisation du patrimoine scientifique, est Colette Cadiou, ex-Irstea ; le pôle Numérique pour la science (Num4Sci), où se posent les problématiques de gestion des données de la recherche, est sous la responsabilité d'Esther Dzalé Yeumo, ex-Inra. Les données de la recherche ne sont donc pas le cœur du périmètre des missions du pôle ACDC composé des fonctions Edition, Ressources documentaires et Archivages. Mais elles ne leur sont pas étrangères, dès lors que chacune de ces fonctions se déploie dans le cadre de la science ouverte ; elles sont ainsi gérées, comme le souligne Colette Cadiou, au sein de la DipSO et à cheval entre plusieurs pôles<sup>3</sup>.

C'est ainsi par le biais d'un groupe de travail commun « stockage et archivage » que les problématiques nécessairement transverses des données de la recherche sont approfondies. Créé en 2019 à l'Inra en relation avec le déploiement de DataInra, ce groupe prolonge à INRAE ses objectifs de réflexions sur les enjeux conjoints du stockage et de l'archivage pérenne. Il réunit ainsi des membres de Num4Sci et d'ACDC dont les archivistes Raymond Schiano<sup>4</sup>, ex-Irstea, et Lina Sbeih, ex-Inra, particulièrement impliquée sur le sujet depuis l'été 2020<sup>5</sup>. Leur ambition est de proposer une offre de service pour l'Institut répondant aux problématiques plurielles et différenciées du stockage et de l'archivage en contexte d'inflation des données et des documents. Conscients que toutes les données ne pourront être conservées, ils distinguent « *l'opportunité d'archiver* » de « *la*

---

<sup>1</sup> Raymond Schiano, « Archivage à l'Irstea », diaporama de formation, Clermont-Ferrand, 2 juillet 2019, p. 3.

<sup>2</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>3</sup> Colette Cadiou, entretien du 1<sup>er</sup> juillet 2021.

<sup>4</sup> Raymond Schiano, entretien du 1<sup>er</sup> juillet 2021.

<sup>5</sup> Lina Sbeih, Fanny Dedet, Patrick Moreau, Esther Dzale, « L'archivage des données de la recherche à l'Inra. Eléments de réflexion, démarche et perspectives », *Cahier des Techniques de l'INRA*, INRA, 2020, 10 p., ffhal-02861909f.

*capacité d'archiver* »<sup>1</sup> : la première dépend de l'élaboration de la gouvernance des données comprenant un volet archivage qui définit « *les critères de choix des jeux de données qui feront l'objet d'un archivage au-delà de dix ans* » ; la seconde, quant à elle, est déterminée par la nature des formats des données et leur degré de documentation suffisant pour préserver leur intelligibilité. L'entrepôt DataINRAE se fait alors le lieu de mise en œuvre des bonnes pratiques de gestion des données de la recherche « *favorable à l'archivage pérenne* »<sup>2</sup>, et ce grâce à l'étude fine, comparable à l'archivistique, de la valeur de preuve, de l'intérêt scientifique, historique ou patrimonial des données dont la destruction se réalise par « *une évaluation des risques et les conséquences de [leur] perte d'exploitation* ». Le document « L'archivage des données de la recherche à l'Inra. Eléments de réflexion, démarche et perspectives », élaboré à l'été 2020, pose ainsi les premières bases méthodologiques abouties d'un archivage numérique pérenne empruntant aux savoir-faire complémentaires de l'information scientifique et technique et des sciences de l'archive. L'expertise des archivistes est à ce titre légitime, et dans le cas d'INRAE, renforcée par les compétences complémentaires de Lina Sbeih et Raymond Schiano. En effet, ce dernier, par la trajectoire de sa carrière au sein de l'Irstea, se positionnait davantage sur le versant scientifique des archives de la recherche quand Lina Sbeih, engagée en 2015 au siège de l'Inra, a développé ses missions sur l'archivage des documents d'administration<sup>3</sup>. Réunis en janvier 2021 au pôle ACDC, les deux archivistes œuvrent à l'élaboration du cycle de vie des documents d'activité au sein de l'Institut et à la sensibilisation aux enjeux de l'archivage numérique, le tout en relation avec le Comité d'Histoire qui, lui-même par l'action de Pierre Cornu notamment, entend se faire connaître auprès des unités de recherche afin de notifier aux chercheurs leur vif intérêt pour leur mémoire<sup>4</sup>. Le Comité d'Histoire se fait alors une instance de collecte des archives de la recherche complémentaire, bien que ses activités empruntent encore trop davantage au modèle du pompier intervenant sur des feux.

INRAE, en effet, ne dispose pas de système d'archivage électronique (SAE) en mesure de prendre en charge la pérennisation des documents et données sur le long terme. Déjà à l'époque de l'Irstea, sur la question combinée de l'archivage et du stockage, Raymond Schiano souligne qu'ils avaient botté en touche en l'absence de moyens pour s'équiper de logiciels en mesure d'engager les actions de préservation idoines. Or aujourd'hui, lorsqu'il est nécessaire d'intervenir sur un fonds numérique à la demande d'un agent de l'Institut, les archivistes ne peuvent leur proposer que du « *stockage capacitif* »<sup>5</sup> sur les serveurs d'INRAE dont l'accès se réalise, en façade, par l'intermédiaire de leur GED. C'est, selon Raymond Schiano, une forme d'« *archivage intermédiaire* » bien loin d'être satisfaisante mais en mesure de faire face, à court et moyen terme, aux enjeux de la pérennisation des

---

<sup>1</sup> *Ibid.*, p. 5.

<sup>2</sup> *Ibid.*, p. 5.

<sup>3</sup> Raymond Schiano, entretien du 1<sup>er</sup> juillet 2021 ; Pierre Cornu, entretien du 11 juin 2021.

<sup>4</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>5</sup> Raymond Schiano, entretien du 1<sup>er</sup> juillet 2021.

archives numériques qui à INRAE recouvrent une typologie très vaste. La solution logicielle actuellement utilisée est *Nextcloud*, sur laquelle est en train d'être migré l'ensemble des documents des GED des anciens instituts. Si l'Irstea gérait les documents d'activité sur les plateformes *Alfresco* puis *Stratus*, l'Inra, quant à lui, avait engagé en 2016 une étude sur le déploiement la gestion électronique de documents en son sein<sup>1</sup>. INRAE tente ainsi de répondre aux besoins des deux anciens instituts en la matière.

Toutefois, INRAE est encore jeune. L'ensemble des acteurs que nous avons rencontrés ont ainsi souligné à de multiples reprises la juvénilité de tous leurs projets autour de l'archivage numérique qui, se nourrissant des savoirs et savoir-faire de chacun, prennent toutefois une nouvelle ampleur sous la direction de la DipSO. Son champ de compétences très vaste et ses ambitions en science ouverte imposent en effet une phase d'apprivoisement des enjeux et des personnes dont INRAE, par son envergure politique et scientifique, ne peut faire l'impasse.

## **Les enjeux de la patrimonialisation des archives de la recherche**

Le caractère patrimonial des données de la recherche est une thématique loin d'être anecdotique dans les ambitions d'ouverture d'INRAE. La DipSO identifie clairement quatre principaux enjeux à ce nouveau paradigme du partage des savoir qui sont, dans l'ordre de leur présentation, patrimoniaux, économiques, scientifiques et sociétaux<sup>2</sup>. Les enjeux patrimoniaux ouvrent ainsi l'énoncé – ce n'est pas anodin – et s'accompagnent de deux attributs visant à expliciter ce qu'ils sous-tendent. Par enjeux patrimoniaux doit en effet être compris, selon la DipSO, la prise en compte et en charge des valeurs de « *preuve* » et de « *mémoire* », la première aux fondements éthiques de tous les autres enjeux, la seconde comme visée à la fois immédiate et diachronique justifiant leur déploiement. Aborder la dimension patrimoniale des données de la recherche n'est toutefois pas un sursaut ponctuel dans les réflexions de l'Institut sur le sujet de la science ouverte ; c'est un axe de travail récurrent qui traverse l'Inra et l'Irstea depuis les années 2010.

La charte de déontologie de l'Inra, nous l'avons déjà souligné, fait ainsi en 2016 des données de la recherche, aux côtés « [des] *savoir-faire et collections*, [de] *la propriété industrielle* », « *un patrimoine de qualité à préserver et valoriser* »<sup>3</sup>. Une vision partagée en 2017 à l'Irstea où « *la gestion patrimoniale est une entrée importante de la question* »<sup>4</sup> du partage des données de la recherche selon Emmanuelle Jannès-Obert, aujourd'hui directrice adjointe de la DipSO. Il n'est donc

---

<sup>1</sup> Fanny Dedet, Olivier Schneider, *Cadrage Gestion électronique de documents (GED)*, note de cadrage, 2016, 97 p., hal-02798909.

<sup>2</sup> DipSO, « Les enjeux pour INRAE », *datapartage.inrae.fr*, INRAE, le 19 février 2015, mis à jour le 30 mars 2020, <https://datapartage.inrae.fr/Partager-Publier/Les-enjeux-pour-INRAE>.

<sup>3</sup> INRAE, « Charte de déontologie de l'Inra », *op. cit.*, p. 3.

<sup>4</sup> Emmanuelle Jannès Obert, « Problématique de gestion, de diffusion des données », Séminaire « Données scientifiques », le 16 février 2017, Paris.

pas étonnant de constater que sept années plus tard, lors de l'élaboration des principes pour une gouvernance des données à INRAE, celles-ci soient réaffirmées comme « *un patrimoine scientifique commun* »<sup>1</sup> : elles doivent en effet être engagées dans un mouvement de gestion raisonnée à des fins d'ouverture pour pouvoir, *in fine*, être *transmises*. Car le savoir scientifique, par essence, se communique et c'est bien à cela qu'aspire INRAE lorsqu'il fait des données de la recherche, en science ouverte, un patrimoine : il initie un mouvement de *transmission, dans le temps*, qui enrichit la quête initiale de *communication, dans l'espace*, des sciences.

Or la mise en patrimoine connaît depuis la seconde moitié du XX<sup>e</sup> siècle un bouleversement dans son appréhension par la société, marquée par une forme de rupture dans la transmission. Pierre Nora l'expliquait alors, « *on est passé d'un patrimoine étatique et national à un patrimoine de type social et communautaire où se déchiffre une identité de groupe ; et, donc, d'un patrimoine hérité à un patrimoine revendiqué. De matériel et visible, le patrimoine est devenu invisible et symbolique [...] le patrimoine a quitté son âge historique pour entrer dans son âge mémoriel* »<sup>2</sup>. Un âge mémoriel qui ne contraint plus à l'oubli, bien au contraire, et où l'environnement numérique devient un « *substitut à la mémoire* »<sup>3</sup> en se faisant le lieu commun du « *mirage de l'exhaustivité (tout garder) et de l'exactitude (fidélité de l'enregistrement)* »<sup>4</sup>. Plongé dans une mémoire devenue archivistique qui « *s'appuie tout entière sur le plus précis de la trace, le plus matériel du vestige, le plus concret de l'enregistrement, le plus visible de l'image* »<sup>5</sup> et entend « *à la fois la conservation intégrale de tout le présent et la préservation intégrale de tout le passé* », le numérique renforce l'ambition de la société avide de mémoire de lui déléguer « *le soin de se souvenir pour elle* ». La patrimonialisation des données de la recherche, sous la valeur explicite de « *mémoire* » revendiquée par INRAE, interroge donc le rapport à la connaissance scientifique qui advient progressivement selon les modalités de la science ouverte. Stimulée par ses idéaux de partage et d'ouverture, celle-ci se déploie dans ce présent *omniprésent*, cette crise du temps où viennent « *à perdre de leur évidence les articulations du passé, du présent et du futur* »<sup>6</sup>.

Ainsi, le terme travaillé par INRAE sur les données de la recherche est celui de « *préservation* », attribut phare du présentisme aux côtés de conservation ou encore de prévention selon François Hartog<sup>7</sup>. Toutefois, la « *préservation* » des données patrimoniales scientifiques s'inscrit dans une démarche de « *valorisation* » et de mise en « *commun* », tentant en cela de réinvestir la recherche

---

<sup>1</sup> Direction pour la Science Ouverte, « Principes pour la gouvernance des données », *op. cit.*, p. 3.

<sup>2</sup> Pierre Nora, « Conclusion des Entretiens », Pierre Nora (dir.), *Science et conscience du patrimoine. Entretiens du patrimoine*, Paris, Editions Fayard, 1997, p. 391.

<sup>3</sup> Bruno Bachimont, *Patrimoine et numérique. Technique et politique de la mémoire*, *op. cit.*, p. 219.

<sup>4</sup> *Ibid.*, p. 218.

<sup>5</sup> Pierre Nora, Pierre Nora, « Entre mémoire et histoire. La problématique des lieux », Pierre Nora, *Les lieux de mémoire*, Paris, Gallimard, vol. 1 « La République », 1984, p. 30.

<sup>6</sup> François Hartog, *Régimes d'historicité. Présentisme et expériences du temps*, Paris, Editions du Seuil, 2012 [2003], p. 38

<sup>7</sup> François Hartog, *Chronos. L'Occident aux prises avec le Temps*, Paris, Gallimard, 2020, p. 288.

agronomique d'un peu de temporalité grâce à de nouveaux acteurs, extérieurs à leur mise en patrimoine, qui peuvent et pourront s'en saisir pour produire de nouveaux savoirs et savoir-faire. Cette dynamique, caractéristique de la science ouverte, est ainsi portée par l'environnement numérique qui induit un « *désintermédiation de la mémoire* »<sup>1</sup> dans un double mouvement de numérisation du patrimoine et de patrimonialisation du numérique. Une désintermédiation qui reste cependant apparente selon Louise Merzeau, professeure en sciences de l'information et de la communication, car la captation des traces en environnement numérique, où se dépose la mémoire, se réalise par le biais d'un ensemble d'artefacts et d'acteurs dont les attributions ne sont pas « *explicitement liées au patrimoine ou à l'archive, mais à la communication et à l'information* »<sup>2</sup>.

Or à INRAE, ce sont bien encore par des approches info-communicationnelles que les actions de transmission des données de la recherche se réalisent. En effet, outre le groupe de travail transverse « stockage et archivage », il n'existe pas à l'heure actuelle de dispositif de valorisation des données de la recherche à proprement parler dans lequel la fonction « Archives – Valorisation du patrimoine scientifique » de la DipSO aurait pu pleinement s'engager. Non pas que la Direction pour la Science ouverte n'en ait pas la motivation, ni les compétences, bien au contraire, mais la jeunesse de l'Institut, de la DipSO et de ses équipes, qui s'approprient encore après avoir initié les premiers travaux en pleine pandémie, impose du temps et de la place pour l'avènement de projets aboutis. Toutefois, si la question patrimoniale des données de la recherche devait un jour émerger en dehors de son cadre info-communicationnel, elle s'appuierait sur l'expertise développée jusqu'à présent dans le champ général de la valorisation des archives de la recherche qui, tant à l'Inra qu'à l'Irstea, ont été bien pris en compte ces dernières années au travers de nombreux projets.

En effet, INRAE, par le biais des deux anciens instituts, est notamment engagé dans une politique de numérisation de collections de littératures blanche et grise, ainsi que de photographies, « *afin d'en assurer la conservation, d'en faciliter la diffusion et d'en permettre la réutilisation* »<sup>3</sup>. Cette première forme de valorisation qu'est la numérisation se réalise depuis plusieurs années déjà au sein de l'Irstea, faisant appel pour l'occasion à des prestataires externes formés<sup>4</sup>. Pour l'Inra, le travail est plus délicat. Le caractère dense de ses archives, pourtant dispersées et parcellaires, et la longue absence d'une volonté politique claire sur leur gestion raisonnée ont induit des dispositifs de traitement différenciés selon les attentes, les acteurs et les ambitions. Ainsi le Comité d'Histoire, par exemple, s'est rapidement engagé dans des actions de numérisation afin de répondre à l'urgence face à des

---

<sup>1</sup> Louise Merzeau, « Faire mémoire des traces numériques », *E-dossiers de l'audiovisuel*, INA, juin 2012, <http://www.ina-expert.com/e-dossier-de-l-audiovisuel-sciences-humaines-et-sociales-et-patrimoine-numerique/faire-memoire-des-traces-numeriques.html>.

<sup>2</sup> *Ibid.*

<sup>3</sup> « Patrimoine numérisé », *ist.inrae.fr*, consulté le 26 juin 2021, <https://ist.inrae.fr/le-libre-acces-a-inrae/patrimoine-numerise/>.

<sup>4</sup> Raymond Schiano, entretien du 1<sup>er</sup> juillet 2021.

fonds risquant d'être détruits, ou lorsque des chercheurs, partant à la retraite avec leurs archives mais les laissant à disposition pour consultation avant de les récupérer, autorisent les membres du comité à les numériser. Les historiens eux-mêmes, parfois assistés d'étudiants, entreprennent ces actions de numérisation des documents avant de les stocker dans des *drives* pour leurs recherches ultérieures. Insatisfaisant à la fois sur le fond et la forme, l'archivage numérique pour les travaux d'histoire est ainsi pour le moment encore « *du bricolage* »<sup>1</sup> de l'aveu de Pierre Cornu. Toutefois, « *la DipSO, elle, ne bricole pas ; elle travaille professionnellement* ». Le Comité d'Histoire travaille ainsi en étroite collaboration avec la DipSO et attend de la part de ses membres qu'ils viennent « *professionnaliser, sécuriser et déontologiser la maîtrise de la question de l'archive numérique* »<sup>2</sup> et de leurs procédures.

En effet, les membres de la DipSO ont initié au fur et à mesure de ces dernières années des projets de valorisation du patrimoine scientifique leur permettant de d'étudier les méthodologies, les dispositifs, leurs avantages et leurs contraintes. Dès 2014, la Délégation à l'information scientifique et technique de l'Inra fait le choix, dans le cadre d'un partenariat avec la BSN et d'un financement du ministère de l'Enseignement supérieur et de la recherche, de numériser le fonds documentaire des *Avertissements agricoles*, puis de le déposer sur *Internet Archives*. Ce sont 50 000 pages numérisées pour environ 26 000 entrées sur le site internet en libre accès. Cette option de valorisation, qui vise ainsi le partage des documents à un public dépassant le cadre des spécialistes, s'est réalisée à une époque où l'Inra, récemment engagé sur ces missions de patrimonialisation, n'avait pas encore alloué les moyens financiers idoines<sup>3</sup>. La solution *Internet Archives* avait ainsi été jugée plus simple et moins chère que *Gallica* de la Bibliothèque nationale de France (BNF). L'Institut, pourtant, est associé à *Gallica* depuis 2011, en tant que partenaire d'ESR et contributeurs du marché de numérisation des imprimés<sup>4</sup>. Mais il faudra attendre 2018 pour que le partenariat prenne concrètement forme par la numérisation du fonds Thuret, fonds de la bibliothèque de Gustave Thuret (1817-1875) constituée de 160 mètres linéaires de rayonnages occupés dont 280 livres rares ont été numérisés (80 000 pages) et rendus accessibles sur *Gallica*<sup>5</sup>. C'est toutefois avec le projet *Gallica marque blanche* qu'INRAE entre dans une nouvelle dimension de la valorisation de son patrimoine scientifique.

*Gallica*, créé en 1997 en tant que bibliothèque numérique collective par la BNF dont les missions sont de « *collecter, conserver, enrichir et communiquer le patrimoine documentaire national* »<sup>6</sup>, développe au début des années 2010

---

<sup>1</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>2</sup> *Ibid.*

<sup>3</sup> Colette Cadiou, entretien du 1<sup>er</sup> juillet 2021.

<sup>4</sup> Direction des services et des réseaux, *Les partenaires de Gallica en 2017*, Bibliothèque nationale de France, février 2018, p. 22, [https://www.bnf.fr/sites/default/files/2018-11/partenaires\\_gallica.pdf](https://www.bnf.fr/sites/default/files/2018-11/partenaires_gallica.pdf).

<sup>5</sup> « Institut national de la recherche agronomique – Fonds Thuret », 2018, <https://gallica.bnf.fr/edit/und/institut-national-de-la-recherche-agronomique-inra-0>.

<sup>6</sup> Alix Lamé-Bergis, Isabelle Mangou, *Gallica marque blanche*, BNF, été 2020, p. 1.

l'initiative *Gallica marque blanche*, « forte du résultat de ces années investies dans la création et l'évolution de sa bibliothèque numérique »<sup>1</sup>. La BNF souhaite en effet « mutualiser les moyens qui ont été alloués au développement de celle-ci et faire ainsi bénéficier ses partenaires de son savoir-faire ». *Gallica marque blanche* offre ainsi l'opportunité aux établissements partenaires d'engager des actions de numérisation afin d'intégrer les collections numériques nationales en bénéficiant d'une interface web au nom et à l'identité graphique personnels au travers de laquelle accéder et consulter les fonds documentaires versés. Cette offre se réalise dans une « logique d'enrichissement mutuel des collections »<sup>2</sup> sous réserve d'un « appréciation scientifique des collections numérisées proposées par les partenaires et la capacité d'intégration de ces documents dans la bibliothèque numérique ».

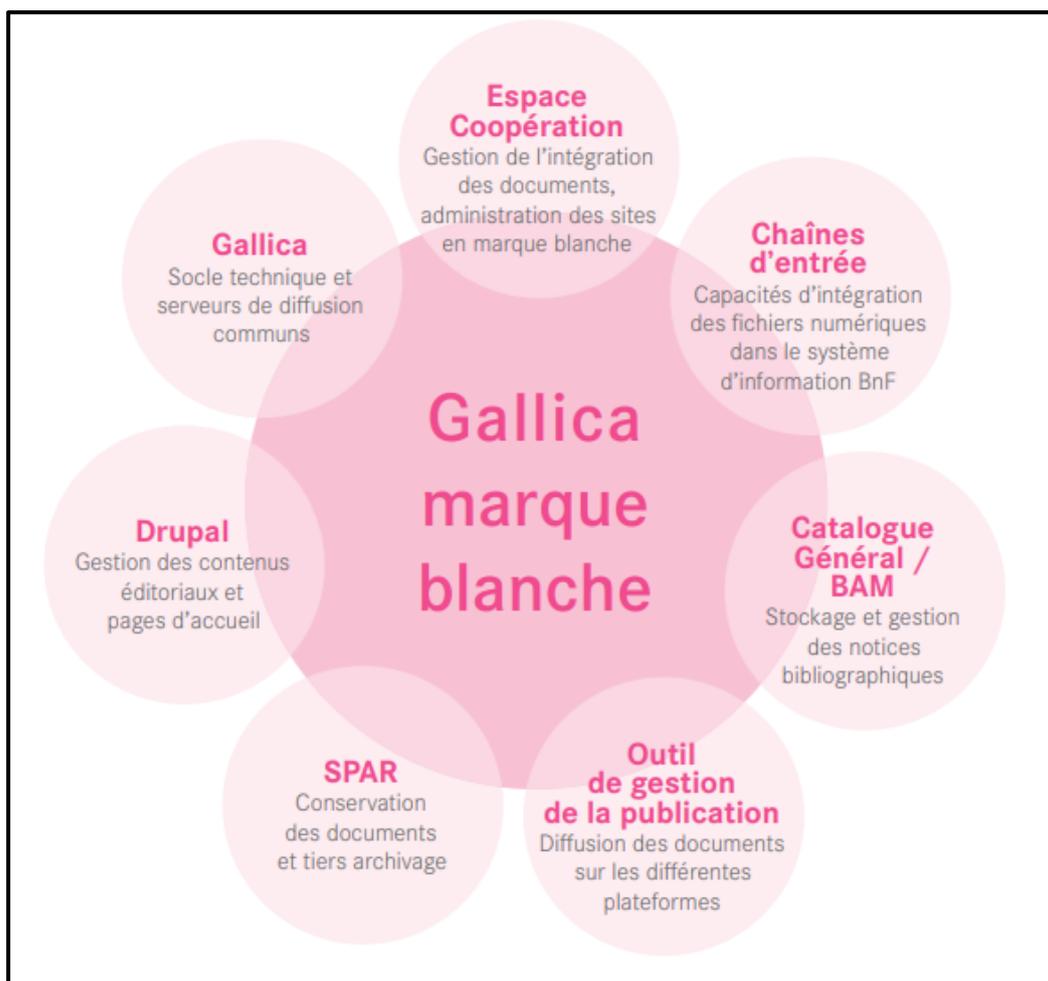


Schéma de présentation de *Gallica marque blanche* - Alix Lamé-Bergis, Isabelle Mangou, 2020.

En 2020, ce sont ainsi une vingtaine de bibliothèques numériques en marque blanche réalisées, dont le Centre de coopération internationale en recherche agronomique pour le développement (Cirad) auquel a longtemps été associé l'Inra, notamment sur les enjeux de mémoire ; le Comité d'Histoire de l'Institut, créé en 2005, devient en effet en 2008 le Comité d'Histoire de l'Inra et du Cirad. La

<sup>1</sup> *Ibid.*, p. 1.

<sup>2</sup> *Ibid.*, p. 2.

bibliothèque numérique du Cirad, *NumBA*, est consacrée à l'histoire de l'agronomie des régions tropicales et méditerranéennes à partir de la fin du XIX<sup>e</sup> siècle, donnant ainsi accès aux textes, cartes, iconographies de leurs champs de recherche agronomique. De manière semblable, INRAE est donc actuellement engagé dans ce projet en marque blanche qui concentre les efforts de certains membres de la DipSO en vue d'en faire la vitrine du patrimoine scientifique de l'Institut. Surtout, selon Raymond Schiano, *Gallica marque blanche* offre l'opportunité, comme « *produit d'appel* »<sup>1</sup>, de promouvoir en interne cette dynamique de patrimonialisation auprès des chercheurs : « *On leur dit, "si vous avez encore des fonds dont vous estimez qu'ils sont rares, originaux, intéressants, ou des choses qui risquent d'être perdues, il y a ça".* »

Cette bibliothèque numérique interviendra, une fois mise en production, en complément des autres offres numériques de l'Institut dont HAL INRAE. Si nous avons déjà abordé l'organisation de l'archive ouverte, conçue peu à peu comme espace de relais des productions scientifiques contemporaines, il importe de souligner qu'elle participe à la mise en patrimoine de la recherche agronomique en rendant disponible en libre accès les collections complètes des revues dont INRAE est propriétaire et qui, pour nombre d'entre elles, ont marqué l'histoire de la recherche agronomique française et internationale. Certaines d'entre elles très anciennes, et/ou n'étant plus éditées, sont ainsi conservées et pérennisées sur les serveurs du CINES, telles les *Annales de l'Abeille* (1958-1968), *Agronomie* (1981-2004), *Le lait* (1921-2004), *Le Courrier de l'environnement* (1993-...), *Veterinary Research* (1993-...), et bien d'autres.

D'autres projets participent par ailleurs, de manière plus ponctuelle, à la dynamique de valorisation du patrimoine scientifique d'INRAE. Ses fonds photographiques, particulièrement conséquents, commencent ainsi à être appréhendés, en sachant que leur problématique, selon Pierre Cornu, est de traiter à la fois des négatifs et des fichiers numériques qui bien trop souvent ne sont ni indexés ni référencés<sup>2</sup>. L'Inra a toutefois en 2015 traité le fonds photographique de l'artiste photographe Jean Weber réalisant une grande partie de sa carrière à l'Inra<sup>3</sup>. Ce sont ainsi 23 501 photographies déposées sur la plateforme *Flickr* dont 10 000 diapositives numérisées pour un total de 1,4 millions de vues depuis 2015<sup>4</sup>. Cette valorisation exceptionnelle est en partie due au photographe lui-même racontant d'ailleurs en 2015 que le numérique avait bouleversé ses pratiques photographiques, le conduisant à opter pour un « *archivage chronologique* » de ses prises de vue. « *L'ensemble des photos se trouvent dans un dossier nommé "Images", dans lequel j'ai créé un dossier par année, puis dans chaque année un dossier par mois, et enfin un dossier par événement.* »<sup>5</sup> Une rigueur qui permet la valorisation de ce fonds

---

<sup>1</sup> Raymond Schiano, entretien du 1<sup>er</sup> juillet 2021.

<sup>2</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>3</sup> Jean Weber, Christian Galant, « Jean Weber : témoignage », *Archorales*, 18, Editions INRA, 2018, 192 p., hal-02790495.

<sup>4</sup> Fonds Jean Weber, [https://www.flickr.com/people/inra\\_dist/](https://www.flickr.com/people/inra_dist/).

<sup>5</sup> *Ibid.*, p. 184.

photographique conséquent, toutefois réalisée sur une plateforme privée, certes dédiée à l'hébergement à la fois amateur et professionnel de photographies, mais dont la pérennisation est bien loin d'être assurée à long terme, soulignant toujours davantage le rôle de ces entreprises privées qui ont su, au cours des deux dernières décennies, s'approprier certaines fonctions régaliennes usuellement dévolues aux Etats et échapper de la sorte aux logiques institutionnelles classiques.

Cet enjeu se pose d'ailleurs avec une certaine acuité dans le cadre de la patrimonialisation du numérique, face inverse de la numérisation du patrimoine abordée précédemment. Ainsi soulignerons-nous l'engagement de réflexions de la part de la DipSO sur le sujet de la collecte et de la conservation des sites web d'INRAE pourvu de quelques 1500 sites divers et variés qui, pour bon nombre d'entre eux, risquent de voir leur contenu disparaître dès changement de plateforme ou fermeture du site<sup>1</sup>. Ce sont toutefois des réflexions qui se réalisent en association avec les actions de la BNF pour le compte duquel le pôle ACDC de la DipSO, selon Colette Cadiou, peut éventuellement donner des recommandations sur le devenir de tel ou tel site. Cet axe de « mise en patrimoine » et de valorisation des contenus web demeure pour le moment à l'état de projet d'autant que des infrastructures d'ampleur se sont d'ores et déjà engagées dans l'archivage du web. On notera ainsi la présence sur *Internet Archives* d'une première collecte du site institutionnel de l'Inra *inra.fr* au 7 octobre 1997, qui restitue notamment la présentation générale de l'Institut bien éloigné des préoccupations de science ouverte et des problématiques environnementales, sanitaires et alimentaires<sup>2</sup>. Les historiens des sciences ne sauraient à ce jour faire l'impasse sur ces ressources pour étudier l'histoire de la recherche agronomique.

En effet, comme le soulignent Francesca Musiani et Valérie Schafer, le patrimoine, qu'il soit numérisé ou nativement numérique, « *incite aujourd'hui les chercheurs et le monde des archives à se tourner vers les outils promus par les humanités numériques, que ce soit dans le contexte de la création de corpus, de leur documentation ou encore de leur exploitation* »<sup>3</sup>. Le numérique, comme *milieu*, offre par les outils qu'il déploie l'opportunité de repenser la recherche historique, d'autant plus lorsque celle-ci, fille des Lumières et de la rationalité, prend les sciences contemporaines pour objet d'étude.

---

<sup>1</sup> Colette Cadiou, entretien du 1<sup>er</sup> juillet 2021.

<sup>2</sup> « Présentation générale de l'INRA », *inra.fr*, snapshot *Internet Archives* du 7 octobre 1997, [https://web.archive.org/web/19971007182126/http://www.inra.fr/PRESENTATION\\_INRA/index.html](https://web.archive.org/web/19971007182126/http://www.inra.fr/PRESENTATION_INRA/index.html).

<sup>3</sup> Francesca Musiani, Valérie Schafer, « Patrimoine et patrimonialisation numériques », *RESET*, n° 6, 2017, <http://journals.openedition.org/reset/803> ; DOI : 10.4000/reset.803.

# ÉCRIRE L'HISTOIRE DE LA RECHERCHE AGRONOMIQUE EN ENVIRONNEMENT NUMÉRIQUE

## L'histoire en environnement numérique et ouvert

L'histoire à l'ère numérique est-elle une histoire numérique ? Cette question, délicate à cerner, résulte de ce qui sous-tend l'attribut de numérique à toute discipline. Bruno Bachimont le souligne, une histoire numérique, une linguistique numérique ou encore « *une philologie numérique aborde-t-elle les problèmes philologiques posés par le numérique ou veut-elle traiter ses objets habituels en enrichissant ses méthodes d'outils numériques ?* »<sup>1</sup> La question de la place du numérique dans l'activité de recherche est toutefois loin d'être nouvelle au sein de la communauté des historiens, ceux-ci aimant à citer Emmanuel Leroy Ladurie prédisant dès 1968 que « *l'historien de demain sera programmeur ou ne sera pas* »<sup>2</sup>. Mais cette idée – ou idéal peut-être – d'un historien programmeur renvoie à la seule dimension instrumentale du numérique qui aujourd'hui, bien au contraire, se fait environnement plein, nous l'abordons en première partie. Les humanités numériques, certes, portent en premier lieu un vif intérêt aux outils qui modifient les pratiques de recherche ; d'ailleurs en histoire, « *la transformation la plus souvent mise en évidence est la possibilité nouvelle d'accéder immédiatement à de vastes réservoirs de données et d'articles* »<sup>3</sup>. Mais toute œuvre d'historien advient aujourd'hui dans le *milieu* numérique, modifiant inéluctablement en cela les pratiques de recherche, les conditions de production des connaissances et la nature même des savoirs.

Sans surprise, les sciences de l'information et de la communication ont très tôt appréhendé le numérique pour interroger de manière collective leur relation à celui-ci. L'histoire, elle, a été moins vivace mais se saisit aujourd'hui pleinement des enjeux « *pour la fabrique et l'écriture de l'histoire* »<sup>4</sup> d'autant que, « *à rebours de discours sur la radicale nouveauté de ces enjeux pour les sciences, la tradition de réflexivité des SHS fournit un ensemble de concepts critiques qui en relativise la portée* »<sup>5</sup>. Les sciences humaines et sociales, se faisant humanités numériques, ont en effet rapidement compris qu'il ne suffisait pas de déployer des outils et d'accroître les compétences en informatique pour renouveler la recherche ; il s'agit davantage « *de constater, d'assumer et au besoin de critiquer* »<sup>6</sup> la charge théorique du renouvellement épistémologique de la recherche induit par le numérique. Un

---

<sup>1</sup> Bruno Bachimont, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques. Principes pour une science des données », *op. cit.*, p. 386.

<sup>2</sup> Emmanuel Le Roy Ladurie, « La fin des érudits », *Le Nouvel Observateur*, 8 mai 1968, p. 38.

<sup>3</sup> Philippe Rygiel, *Historien à l'âge numérique*, Villeurbanne, Presses de l'ENSSIB, 2017, p. 180.

<sup>4</sup> Frédéric Clavert, Valérie Schafer, « Les humanités numériques, un enjeu historique », *Quaderni*, vl. 1 (n° 98), 2019, DOI : 10.4000/quaderni.1417.

<sup>5</sup> Clément Mabi, Jean-Christophe Plantin, Laurence Monnoyer-Smith, « Interroger les données en SHS à partir de leur écosystème », *op. cit.*, p. 67.

<sup>6</sup> *Ibid.*, p. 70.

double travail d'assomption et de critique de cette technologie intellectuelle qui font des humanités numériques, selon Frédéric Clavert et Valérie Schafer, un « lieu de bricolage »<sup>1</sup> marqué par la valeur d'expérimentation et où penser « la place du tâtonnement, de l'artisanat, des impasses dans les processus de recherche ». Les humanités numériques ouvrent en effet de nouveaux champs des possibles qui doivent inviter « le chercheur à retrouver le goût historique de l'expérimentation. [...] “Un historien à d'ailleurs suggéré de faire droit au temps des démonstrations : les sociétés savantes du XVIII<sup>e</sup> siècle n'ont-elles pas posé les fondements des sciences modernes en faisant se côtoyer le sérieux et le douteux, le spectacle technique et l'expérience scientifique, sous le signe de l'expérimentation.” »<sup>2</sup> Une valeur de l'expérimentation que Pierre Cornu promeut dans ses pratiques de recherche en milieu numérique qui sont, dans cette dualité semblable du sérieux et du douteux, « à la fois du bricolage et de la rigueur méthodologique »<sup>3</sup>.

L'environnement numérique, par la valorisation de l'expérimentation, donne ainsi l'opportunité à l'historien de s'essayer à de nouvelles formes d'écriture de l'histoire. Des formes plurielles qui, du carnet de recherche numérique à l'histoire collaborative<sup>4</sup> entamant un travail de désintermédiation du savoir et de son édition, sont toutes marquées du sceau du partage et de l'ouverture. Le numérique, nous l'avons abordé au cours de cette étude, se fait en effet le catalyseur de la science ouverte : il en définit, aux fondements, les conditions d'existence et de réalisation de ses ambitions de partage des savoirs. Ce partage, dans une dynamique d'ouverture, n'entend toutefois pas l'être dans une relation seulement verticale à autrui ; c'est bien un travail à plat fait d'itérations successives que certains chercheurs en sciences humaines et sociales produisent. Pierre Cornu, en écrivant l'histoire de la recherche agronomique du temps présent, se risque à cette expérimentation :

*« A la longue ça fonctionne en miroir : c'est à la fois une opportunité pour moi d'accéder aux archives de la recherche et c'est une manière différente de faire de l'histoire. Ce que je pratique à INRAE, c'est une histoire collaborative. C'est une histoire qui discute des questions qu'elle pose, des méthodes qu'elle met en œuvre, des produits qu'elle en tire, et qui accepte la critique des acteurs, des témoins, des tiers sur ce qu'elle produit. J'ai un livre qui va sortir sur la systémique agraire à l'automne ; j'en ai montré toutes les étapes aux gens que ça intéressait. Je travaille de manière ouverte, transparente et les étapes d'élaboration, les questionnements, les méthodes, les concepts mis en œuvre se discutent. C'est une démarche qui me plaît. C'est pour ça aussi que j'ai accepté d'être en délégation à INRAE. C'est que cette culture de la science ouverte me semble intéressante pour enrichir la méthode historique, notamment dans l'histoire du temps présent où on ne peut pas être un*

---

<sup>1</sup> Frédéric Clavert, Valérie Schafer, « Les humanités numériques, un enjeu historique », *op. cit.*

<sup>2</sup> Aurélien Berra, « Pour une histoire des humanités numériques », cité dans Frédéric Clavert, Valérie Schafer, « Les humanités numériques, un enjeu historique », *op. cit.*

<sup>3</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>4</sup> Patrick Fraysse, Julie Deramond, Jessica de Bideran, « La construction info-communicationnelle des médiations historiques et mémorielles », *Sciences de la société*, n° 99, 2016, p. 3 à 12.

*historien extérieur, en surplomb, qui porte un jugement sans interagir avec les acteurs. On a besoin d'un accès aux acteurs, on a besoin d'une relation qui n'est pas de connivence, mais de donnant donnant, de don et contre don, dans la relation à la production de connaissance. »<sup>1</sup>*

L'historien d'INRAE œuvre ainsi auprès de ses témoins pour enrichir une histoire qui prend sens par les archives et les témoignages conjointement. Florence Descamps le souligne, « *le recours aux sources orales dévoile combien l'histoire n'est pas une connaissance isolée ou sanctuarisée du passé, elle est elle-même traversée de mémoires sociales plurielles* »<sup>2</sup>. Des mémoires qui, à INRAE, sont traditionnellement transmises par l'oralité. C'est donc une histoire des sciences agronomiques impliquée que propose Pierre Cornu :

*« J'ai été invité à écrire un chapitre sur la prise en compte de la question de l'environnement à l'Inra. C'est un texte que j'ai co-écrit avec Egizio Valceschini et qu'on a fait relire, étape après étape, à des acteurs de la prise en compte de la question de l'environnement dans divers départements de recherche. On a envoyé notre texte en état de brouillon à sept ou huit chercheurs, de manière séparée, qui ont eu des responsabilités d'une manière ou d'une autre sur l'environnement. Et ils nous ont fait des retours en disant : "Je ne suis pas d'accord avec vous parce qu'il s'est passé ça, et vous demandez à un tel, et moi je peux témoigner que..." Et donc on s'est retrouvés avec des documents électroniques surchargés de commentaires, de bulles dans les marges, de gens qui nous apportaient des informations historiques majeures, qui étaient en gros des témoignages suscités par la lecture de notre texte, et qui pouvaient être parfaitement en contradiction avec le témoin A et le témoin B qui avaient lu le même texte. »*

Par ces modalités de collecte de bribes de mémoire qui se confrontent, l'historien crée une nouvelle archive qui nourrit le récit de l'histoire :

*« On a fait une compilation de notre texte avec tous les commentaires, notre texte tout écrasé et un nuage de commentaires dans la marge de tous les gens qui se sont passionnés pour la question et qui voulaient que leur version de l'histoire soit la bonne.*

*Et je l'ai archivé ça. En format PDF, pour photographier. Pour moi, ça montre à la fois que c'est de la science ouverte, de l'histoire en train de se faire, de l'histoire collaborative, et c'est de la fabrique autoalimentée de document électronique. Mon propre texte co-signé s'est enrichi ; il a été pris dans un champ de tir entre les gens disant "non mais on ne peut pas écrire ça, oui mais vous oubliez..." », etc. et on voit comment l'histoire se construit dans cette architecture compliquée. »*

Cet exercice de l'historien ne relève pas d'un simple jeu d'acrobate. Devant « *accepter l'idée que l'on vit dans un autre monde* »<sup>3</sup>, il se confronte à cette

---

<sup>1</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>2</sup> Florence Descamps, *Archiver la mémoire. De l'histoire orale au patrimoine immatériel*, Paris, Editions EHESS, 2019, p. 156.

<sup>3</sup> Pierre Cornu, entretien du 11 juin 2021.

technologie intellectuelle qui interroge les conditions de *connaissance* du passé<sup>1</sup>, modifie le mode de raisonnement et change la manière d'écrire : « *Quand on peut sans cesse bousculer des paragraphes, insérer des mots, on ne pense plus pareil : on pense en hypertexte. Et quand on a pour métier de penser la temporalité, et qu'on peut sans cesse revenir en arrière et retoucher les textes comme si l'encre était encore fraîche, ça change complètement l'écriture.* »<sup>2</sup> C'est que le numérique advient comme « *fonction constitutive dans notre rapport à la temporalité* »<sup>3</sup>, comme milieu qui se fait cadre de compréhension du monde et du temps selon ses propres modalités, en considérant « *comme donné et présent ce qui ne l'est pas, moyennant l'exécution du calcul* »<sup>4</sup>. Bien sûr le numérique ne se réduit pas au calcul ; ce nouveau rapport au temps et au monde qu'il institue crée une culture de la complexité selon Bruno Bachimont, une complexité technique notamment qu'il faut assumer et tenter d'alléger.

« *Pour moi, explique Pierre Cornu, il faut considérer que c'est un enrichissement de l'expérience de la temporalité. Ce n'est pas confortable, mais c'est passionnant. Et les historiens ont tort de penser qu'ils sont toujours extérieurs aux choses et qu'ils maîtrisent tout. Pour moi il y a un péché d'orgueil dans le métier d'historien qui est l'illusion d'être extérieur et en surplomb par rapport aux choses que l'on étudie. Faire de l'histoire du temps présent, ça m'oblige à me rendre compte que la production sinon du vrai, du moins du vraisemblable est un combat permanent, mais que je préfère l'inconfort d'une intelligence partagée des faits historiques, qui est sans cesse remis sur le métier, au fait de travailler tout seul dans sa bulle sur des choses que personne ne lit.* »<sup>5</sup>

L'historien, versant volontairement ou non dans l'histoire numérique, œuvre donc aux côtés d'acteurs qui enrichissent ses recherches. Les témoins pour l'histoire du temps présent en premier lieu, évidemment, d'autant que, dans le champ de l'histoire des sciences, l'historien fait face à des savoirs scientifiques qu'il n'est pas en mesure de pleinement saisir seul ; dans chaque archive des sciences se discerne l'arrière-boutique de la fabrique du savoir qui ne peut se révéler qu'à l'aide des chercheurs eux-mêmes. Mais, l'historien s'associe également aux archivistes et bibliothécaires qui participent à construire ces objets frontières de l'environnement numérique. En effet, nous avons eu l'opportunité de le voir au cours de cette étude, de nombreux savoirs et savoir-faire, mobilisés aujourd'hui en humanités numériques, ont d'ores et déjà été réfléchis et développés par ces corps de métier. Nous sommes entrés dans une ère où bibliothécaires, archivistes et chercheurs

---

<sup>1</sup> Bruno Bachimont, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques. Principes pour une science des données », *op. cit.*, p. 112.

<sup>2</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>3</sup> Bruno Bachimont, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques. Principes pour une science des données », *op. cit.*, p. 211.

<sup>4</sup> *Ibid.*, p. 212.

<sup>5</sup> Pierre Cornu, entretien du 11 juin 2021.

« peuvent produire en co-construction et nourrir réciproquement leur travail »<sup>1</sup>, apportant « une tension dialectique où chacune des parties interfère avec l'autre ».

Toutefois, dans cet univers transdisciplinaire de co-construction des savoirs par une instrumentation sophistiquée, que reste-t-il du « goût de l'archive » ? En 1989, Arlette Farge décrit la relation intime que forge tout historien et archiviste avec l'archive. Une archive qui est « excès de sens là où celui qui la lit ressent de la beauté, de la stupeur »<sup>2</sup> en se liant aux personnages du passé. Il y a alors un double mouvement d'appropriation physique et intellectuelle du document qui se réalise dans un lieu dédié, le centre d'archives<sup>3</sup>. Or en environnement numérique, de nouvelles médiations adviennent entre l'archive et l'historien, et le centre d'archives, sa boîte et son fauteuil ne sont plus de mises lorsque les recherches peuvent s'effectuer ailleurs. « Et s'il peut y avoir un déplacement aux archives, lorsque celles-ci ne sont pas encore numérisées, alors le chercheur n'est plus assis à lire les documents, mais debout à prendre des dizaines de photos. »<sup>4</sup> L'archive entre dans l'ère de la donnée, qui s'apprécie par sa quantité ; ce changement de paradigme modifie la façon de la lire. « La lecture aujourd'hui est une recherche à partir de mots clés, s'effectue à partir de résumés et de pages balayées après avoir été photographiées. »<sup>5</sup> Pourtant le « goût de l'archive à l'ère numérique » demeure, bien qu'elle s'apprécie autrement<sup>6</sup>. Certes des historiens, à l'instar d'Odile Maeght-Bournay, co-auteurice de *L'histoire de l'Inra, entre science et politique*, préfèrent « mille fois avoir le document entre les mains »<sup>7</sup>, afin notamment d'avoir « une notion spatiale de là où se trouvent les choses ». Mais les outils numériques recèlent aussi des surprises, engageant tout autant l'historien dans sa quête frénétique de l'archive : « Je m'émerveille, explique Odile Maeght-Bournay, de toutes les archives que je peux finalement trouver en ligne, parfois des choses improbables. J'ai beaucoup de plaisir à faire des recherches et, sur la quinzième page du moteur de recherche, à trouver le document qui correspondait exactement à ce que je voulais. » Un goût de la quête et de la découverte qui transcende donc les technologies intellectuelles. Pierre Cornu nous le partage en ces termes :

« Pour moi, le goût de l'archive vient de la découverte, du don et du partage. Je n'ai aucun problème à apprécier, et à être excité comme un étudiant débutant : “on m'a envoyé ça, pourquoi on me l'a envoyé, qu'est-ce que je peux en faire”. Ce que j'adore dans la méthode historique, c'est le fait que l'on a une méthode pour identifier toutes les intentionnalités, toute la vie que porte un bout de papier, un document qui peut être un fichier électronique. Par exemple, dans les archives que

---

<sup>1</sup> Dominique Naud, « Le numérique : beaucoup de gestes pour un meilleur partage », Frédéric Clavert, Caroline Muller, *Le goût de l'archive à l'ère numérique*, 2018, co-écriture en ligne, <http://www.gout-numerique.net/>.

<sup>2</sup> Arlette Farge, *Le goût de l'archive*, Paris, Editions du Seuil, 1989, p. 42.

<sup>3</sup> Frédéric Clavert, Caroline Muller, « Introduction : le goût de l'archive à l'ère numérique », *op. cit.*

<sup>4</sup> Suzanne Dumouchel, « Compte rendu de la séance introduction du séminaire #EPHN2017 (Invité : Sean Takats) », *Modéliser et virtualiser*, 18 janvier 2017, <https://movi.hypotheses.org/129>.

<sup>5</sup> *Ibid.*

<sup>6</sup> Frédéric Clavert, Caroline Muller, *Le goût de l'archive à l'ère numérique*, 2018, co-écriture en ligne, <http://www.gout-numerique.net/>.

<sup>7</sup> Odile Maeght-Bournay, entretien du 19 juillet 2021.

*j'adore lire, il y a les fils d'échange de mails. C'est incroyable comme archives : quand vous avez deux chercheurs qui se disputent sur l'interprétation d'une question et qu'ils vous mettent en copie, moi je prends les mails, je les prends bout à bout, et j'en fais une archive pour dire : "qu'est-ce qui s'est joué là". »<sup>1</sup>*

L'histoire à l'ère numérique est donc une histoire numérique qui renouvelle ses approches tant en termes techniques qu'épistémologiques ; la construction du savoir historique en est nécessairement bouleversée. Toutefois, l'histoire des sciences, elle, doit peut-être prolonger et approfondir son rapport au numérique. En effet, la recherche scientifique contemporaine, à l'aune de la science ouverte, se restructure toujours davantage dans cet environnement numérique total dont l'étude ne peut dissocier les sciences de leur technologie intellectuelle de subsistance.

## **L'histoire numérique des sciences**

Depuis les années 1980, les *Science Studies* ont réalisé un travail titanesque de délégitimation de l'épistémologie traditionnelle, en identifiant dans le processus de production des savoirs scientifiques la participation d'acteurs, d'instruments et des contraintes exogènes jusqu'alors volontairement minorés. L'histoire des sciences, dès lors, s'est renouvelée en se mêlant « *de mille façons à l'histoire politique, à l'histoire administrative, à l'histoire religieuse, à l'histoire économique, à la métaphysique, bien sûr, à l'histoire du droit, à la psychologie, à l'histoire des sensibilités, du corps, de la subjectivité* »<sup>2</sup>. Toutefois, l'histoire des sciences agronomiques, elle, impose un rapport supplémentaire qui complexifie davantage sa tâche. Elle se lie en effet nécessairement à l'histoire rurale dont l'objet d'étude, depuis la seconde moitié du XX<sup>e</sup> siècle, est à la source d'un empilement de textes scientifiques, de statistiques, de cartes, de discours pluriels dont il concentre vivement l'attention. Pierre Cornu le souligne, « *peu de territoires connaissent un tel déséquilibre entre production de sens et poids économique et démographique* »<sup>3</sup>. Certes la fin du XX<sup>e</sup> siècle marque une déshérence de l'intérêt des sciences humaines et sociales pour les territoires ruraux ; mais la prise de conscience généralisée des problématiques environnementales, réinterrogeant alors les frontières de l'urbain et du rural, engage l'histoire sur de nouveaux terrains de recherche.

Au cours des années 2000, les historiens des sciences et du rural appellent en effet conjointement à l'écriture « *d'une histoire des acteurs et des institutions des sciences et techniques de l'agriculture et de l'alimentation* »<sup>4</sup> ; ils constatent un espace commun dominé par les idéologies de progrès et de modernité, dont l'étude et la critique devaient être d'autant plus engagées que les sociétés occidentales

---

<sup>1</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>2</sup> Bruno Latour, « Il ne faut plus qu'une science soit ouverte ou fermée », *Rue Descartes*, vol. 41, n° 3, 2003, p. 69.

<sup>3</sup> Pierre Cornu, « Les caractères originaux de l'histoire rurale française à l'épreuve de la seconde moitié du XX<sup>e</sup> siècle », Séminaire de recherche, 27 mars 2019, Laboratoire d'études rurales, Lyon.

<sup>4</sup> Christophe Bonneuil, Gilles Denis, Jean-Luc Mayaud, « Pour une histoire des acteurs et des institutions des sciences et techniques de l'agriculture et de l'alimentation » dans Christophe Bonneuil, Gilles Denis, Jean-Luc Mayaud, *Sciences, chercheurs et agriculture : pour une histoire de la recherche agronomique*, Paris, Éditions Quæ, 2008, 304 p.

étaient entrées dans cette modernité réflexive atteinte par ses propres réalisations. Les historiens Christophe Bonneuil et Jean-Luc Mayaud le soulignent : « *Le progrès a perdu sa qualité de discours historique, il a gagné le statut d'objet d'histoire* »<sup>1</sup>, aux côtés du concept de ruralité, lui-même sous-tendant le concept de nature nouvellement approché.

Ce concept de nature, nous l'avons souligné en première partie<sup>2</sup>, se construit au cours des derniers siècles selon un nouveau mode de composition du monde : une cosmogonie fondée, selon l'anthropologue Philippe Descola, sur « *l'affirmation d'une différence de nature, et non plus de degré, entre les humains et les non-humains, une différence qui met l'accent sur le fait que les premiers partagent avec les seconds des propriétés physiques et chimiques universelles, mais s'en distinguent par leurs dispositions morales et cognitives* »<sup>3</sup>. C'est donc une nature domestiquée qui, en tant que telle, est travaillée par les sciences, une nature qui se distingue ontologiquement de l'homme et sur laquelle celui-ci s'en octroie le pouvoir absolu ; une nature portée par cette mystification ne cachant plus aujourd'hui ses apories : « *La nature comme ressource infinie permet une croissance infinie grâce au perfectionnement infini des techniques* »<sup>4</sup>. Le présent est alors marqué par l'ère de l'Anthropocène, cette nouvelle étape de l'histoire de la Terre où la nature réapparaît sous forme de chaos provoqué par l'activité humaine. D'aucuns parlent même de Capitalocène, afin de pointer les logiques intrinsèques au capitalisme et d'identifier la responsabilité à un autre niveau, ou encore de Technocène, ciblant le rapport des sociétés réflexives à la rationalité techno-scientifique<sup>5</sup>. Toutefois, « *ce ne sont pas les découvertes scientifiques qui ont provoqué le changement de l'idée de Nature. C'est le changement de l'idée de Nature qui a permis ces découvertes* »<sup>6</sup>. Les sciences sont donc les héritières de cette cosmogonie naturaliste, qui s'en est faite leur condition d'existence en entamant un travail d'épuration progressive des rapports entre nature et culture. Et l'histoire, fille de Lumières, n'a pas échappé à ce mouvement de rationalisation.

Construite en tant que discipline rationnelle, elle se trouve en effet dans un court-circuit intellectuel lorsqu'elle veut prendre la rationalisation pour objet d'étude, lorsqu'elle veut étudier la recherche agronomique qui lui présente la science faite sous ses plus beaux atours, et dissimule la science en train de se faire où la rationalité n'est pas aussi domptée qu'on le voudrait. Pour Pierre Cornu, quitter l'impasse de ces contradictions impose de « *réapproprier la complexité du réel aux*

---

<sup>1</sup> Christophe Bonneuil, Gilles Denis, Jean-Luc Mayaud, « Pour une histoire des acteurs et des institutions des sciences et techniques de l'agriculture et de l'alimentation », *op. cit.*, p. 35.

<sup>2</sup> 1.2.1. « Les sciences en crise », p. 28.

<sup>3</sup> Philippe Descola, « Humain, trop humain ? », *Esprit*, n° 420, décembre 2015, <https://esprit.presse.fr/article/philippe-descola/humain-trop-humain-38537>.

<sup>4</sup> *Ibid.*

<sup>5</sup> FEP Fondation de l'écologie politique, « "Capitalocène ou Technocène" (Part 1) : Intervention de Christophe Bonneuil », ajouté le 18 novembre 2015. <https://www.youtube.com/watch?v=MAQICfVEaxY>.

<sup>6</sup> Maurice Merleau Ponty, cité dans Philippe Descola, *Par-delà nature et culture*, *op. cit.*, p. 132.

acteurs »<sup>1</sup>, c'est-à-dire de se défier de l'illusion rétrospective de la fatalité oblitérant les contingences de l'histoire. C'est une histoire dite compréhensive « *qui saisit la complexité des individus, des singularités irréductibles des cadres dominants qui ne sont jamais optimalement performants. Les cadres, les dirigeants, les patrons, les élus, l'Etat, les communautés scientifiques, etc. peuvent être (souvent) dans l'erreur, dans le faux, dans l'égarement* » et c'est à cette histoire de réattribuer de manière juste les intentionnalités et les absences d'intentionnalités, d'autant plus dans le champ de la recherche agronomique qui, engagée aujourd'hui dans la science ouverte, est sollicitée sur le devant de la scène pour répondre aux enjeux environnementaux, sanitaires, alimentaires, technologiques, etc. Ces efforts permettent alors de comprendre les trajectoires et de rejeter l'implacabilité du discours du progrès et de la rationalité par la connaissance du passé et de sa complexité.

Or quelle tangibilité de la connaissance de ce passé qui, originellement contingent, est maintenant révolu ? Comment, afin de défaire les œuvres de la rationalité, comprendre la complexité de ce qui n'est plus ? Pour répondre, Bruno Bachimont propose une *épistémologie de la trace*<sup>2</sup>, soit une théorie en capacité de faire sereinement face à ce paradoxe de la connaissance du passé. La trace est « *la clef pour la mémoire de l'événement* »<sup>3</sup>, elle en est l'enregistrement en document, en archive, afin de faire de l'événement une persistance. Mais l'intelligibilité de cette trace ne se réalise qu'à la seule condition de sa juste compréhension, celle-ci se fondant dès lors sur l'empathie, soit cette capacité à faire sien ce qui arrive à autrui et par la trace d'en appréhender la complexité et d'en construire l'histoire. Bruno Bachimont rejoint en cela l'histoire compréhensive d'un Pierre Cornu qui co-construit sa connaissance du passé avec l'aide de ses propres acteurs, et fait dès lors de l'histoire une science *impliquée*, refusant le « *péché d'orgueil dans le métier d'historien qui est l'illusion d'être extérieur et en surplomb par rapport aux choses que l'on étudie* »<sup>4</sup>. L'environnement numérique de la science ouverte œuvre en cela à l'écriture de l'histoire de la recherche, bien qu'il s'agisse toutefois de garder le formalisme nécessaire à une restitution *juste* des traces du passé. Ces considérations interrogent en effet l'acte de rationalité de l'histoire dont certains en proposent un renouvellement épistémologique par l'entrée du récit, déjà bien travaillé par Paul Ricoeur<sup>5</sup>. Yvan Jablonka, notamment, fait du récit une méthode historique valable<sup>6</sup> quand Bruno Bachimont, de manière semblable, travaille son épistémologie de la trace par le récit dont l'intelligibilité de la langue et du discours permet dans le temps et l'espace le partage de l'expression et du ressenti du vécu<sup>7</sup>.

---

<sup>1</sup> Pierre Cornu, « Les caractères originaux de l'histoire rurale française à l'épreuve de la seconde moitié du XX<sup>e</sup> siècle », *op. cit.*

<sup>2</sup> Bruno Bachimont, *Patrimoine et numérique. Technique et politique de la mémoire*, *op. cit.*, p. 108.

<sup>3</sup> *Ibid.*, p. 95.

<sup>4</sup> Pierre Cornu, entretien du 11 juin 2021.

<sup>5</sup> Paul Ricoeur, *Temps et récit. L'intrigue et le récit historique*, Paris, Editions du Seuil, 1991 [1983], vl. 1, 404 p.

<sup>6</sup> Yvan Jablonka, *L'Histoire est une littérature contemporaine*, Paris, Editions du Seuil, 2014, 352 p.

<sup>7</sup> Bruno Bachimont, *Patrimoine et numérique. Technique et politique de la mémoire*, *op. cit.*, p. 109.

Or l'intelligibilité des traces, qui permettent cette expression et ce ressenti du vécu, ne peut être assurée que par des caractéristiques absolues qui fortifient le lien entre elles et l'événement, un lien sans lequel la connaissance du passé n'est plus assurée : authenticité, fiabilité, intégrité, ainsi qu'identité et traçabilité sont ainsi les maîtres-mots des stratégies de préservation de tout document<sup>1</sup>. Ces maîtres-mots, en environnement numérique, ont toutefois quelques aspérités dès lors que cette technologie intellectuelle neuve reconfigure notre rapport aux contenus, et par extension, au passé. En science ouverte, ils nourrissent l'appréhension des produits de la recherche qui deviennent des objets *situés* par un travail complet de documentation. De ces objets émergent, *in fine*, des archives dont la formation aboutie, conditionnée à leur gestion raisonnée en environnement ouvert et numérique, est primordiale pour nourrir cette connaissance du passé et la compréhension de sa complexité. Ainsi, un cycle de vie « vertueux » du document et de la donnée de la recherche, qui engage le document et la donnée de la recherche dans des boucles de traçabilité, de citation et d'hypertexte, participe à saisir dans une profondeur renouvelée les relations, interactions, transactions, les luttes, les doutes, les errements qui adviennent au sein de la recherche scientifique, et enrichit en cela l'écriture de l'histoire des sciences agronomiques. Le travail de documentation des produits de la recherche, qui pour l'histoire des sciences se font traces, fonde donc un savoir historiquement situé sur lequel la connaissance du passé peut s'appuyer.

L'histoire des sciences, en cela, se fait histoire numérique des sciences. A un premier niveau, elle se nourrit de l'approche de science ouverte pour à la fois enrichir son travail d'enquête à partir des objets documentés et produire, à l'instar de Pierre Cornu, une co-construction du récit de l'histoire qui restitue aux acteurs la complexité de leur trajectoire. Mais plus en profondeur, l'histoire numérique des sciences est une histoire qui prend les sciences, et notamment les sciences agronomiques, dans leur environnement numérique total et interroge le rôle de la science ouverte dans leur déploiement. En effet, la science ouverte n'est pas une nouvelle région des sciences ; confrontée aux crises plurielles qui la sollicitent, elle est peu à peu la condition d'existence et de réalisation de la recherche scientifique. Et l'histoire numérique des sciences doit, aux côtés des concepts de progrès, de modernité, de nature ou encore de ruralité, interroger le concept de partage qui la sous-tend.

---

<sup>1</sup> *Ibid.*, p. 87.

## CONCLUSION

---

Au cours de ces deux dernières décennies, la recherche agronomique s'est engagée sur la voie de la science ouverte qui en a modifié progressivement les conditions d'existence. Confrontées à un changement de paradigme scientifique et technique, INRAE entend fonder ses activités scientifiques sur cette dynamique d'ouverture afin de leur faire regagner une certaine légitimité à dire le vrai sur les sujets de l'environnement, de l'alimentation, du vivant qui, en temps de crise, subvertissent l'ordre des champs et du laboratoire. La production des savoirs ne peut plus en effet être l'apanage des chercheurs d'autant que les enjeux scientifiques et éthiques, mais aussi politiques et économiques sont grands. Ce travail mobilise donc un ensemble complexe et complémentaire d'acteurs, qui participent à produire, préserver et partager les connaissances scientifiques. A ce titre, le rôle de la DipSO à INRAE est essentiel, d'autant que « *la professionnalisation de la communication joue un rôle important dans le processus de production considéré comme la fabrication d'un ensemble d'objets matériels (articles, ouvrages, bases de données, images, etc.) orienté vers un public* »<sup>1</sup>. Les professionnels de l'IST ont longtemps joué un rôle d'intermédiaire invisibilisé dans ce processus mais la science ouverte, par le renouvellement éthique de la recherche, impose une rigueur pour les sciences *impliquées* dont les enjeux dépassent le cadre des communautés de chercheurs. Ils opèrent ainsi une mise en relation d'objets, de savoirs, de personnes, et se font des intermédiaires visibles, assumés et nécessaires grâce à l'environnement numérique qui les supporte.

Toutefois, Dominique Wolton le souligne, « *aucune "nouvelle" technologie de communication ne peut à elle seule provoquer une rupture. Il faut la rencontre entre un progrès technique et un projet politique pour qu'il y ait vraiment une "révolution". La technique à elle seule ne suffit pas à transformer les rapports humains et sociaux.* »<sup>2</sup> Le poids de la science ouverte au sein d'INRAE est en effet porté par une vision politique forte qui entend positionner l'Institut, aux échelles nationale et internationale, en fer de lance de la science ouverte dans la recherche agronomique. Une des faiblesses de cette présente étude se manifeste alors ici. Ne disposant ni du recul, ni du temps, ni des sources nécessaires à l'étude exhaustive et critique du déploiement de la science ouverte au sein de la recherche agronomique, le corpus documentaire élaboré pour ce travail se compose pour partie de documents institutionnels dont la dimension performative engagée sur l'ouverture gagnerait à être nuancée de l'envers du décor. Au-delà de cette limite, il importe cependant de souligner que les valeurs d'ouverture sont loin d'être artificielles au sein de l'Inra, de l'Irstea puis d'INRAE, et les professionnels de l'IST, en appliquant à eux-mêmes les principes de la science ouverte, ont tôt fait de partager leurs réflexions, leurs

---

<sup>1</sup> Igor Babou, Joëlle Le Marec, « Les pratiques de communication professionnelle dans les institutions scientifiques », *Revue d'anthropologie des connaissances*, vl. 2, n° 1, 2008, p. 129.

<sup>2</sup> Dominique Wolton : « Abondance et gratuité : pour quoi faire et jusqu'où ? Entretien avec Joëlle Farchy, Pascal Froissart et Cécile Méadel », *op. cit.*, p. 31.

ambitions et leurs travaux que nous avons essayé de restituer tout au long de ce mémoire. Aujourd'hui, la DipSO joue en effet pleinement le jeu de la science ouverte, en rendant accessible une large gamme de documents qui permettent à toute personne intéressée de s'initier et d'approfondir les enjeux de la dynamique d'ouverture et de gestion des données de la recherche. A terme, ces efforts pourront nourrir une histoire des sciences agronomiques engagées dans le mouvement d'ouverture, portées par le numérique, et dont les traces alors documentées des produits de la recherche en enrichiront le récit.

Il importe toutefois de se défier de l'idéal du numérique qui se ferait « *substitut à la mémoire* »<sup>1</sup>, voulant amasser, mais non transmettre. En prenant une « *fonction constitutive dans notre rapport à la temporalité* »<sup>2</sup>, il construit les catégories d'un passé qui n'est qu'itérations répétées, d'un futur qu'une somme de prévisions calculées, et d'un présent qu'une succession d'immédiatetés, le tout à l'appui du présentisme. Dès lors, la science ouverte, portée par le numérique dont la propension synchronique « *connotée positive par l'échange et le partage* »<sup>3</sup> est toujours davantage valorisée, se déploie – et peut-être même se complaît – dans ce présent omniprésent qui « *cannibalise les catégories du passé et du futur* » et « *fabrique quotidiennement d'abord, puis à chaque instant et en continu, le passé et le futur dont il a besoin* »<sup>4</sup>.

Or les crises environnementales, climatiques, sanitaires, et alimentaires, qui s'affranchissent de toute frontière artificielle, mobilisent la recherche agronomique sur de nouvelles temporalités qui ne peuvent s'accommoder à notre régime d'historicité actuel. Les efforts de production des objets documentés en science ouverte, marqués du sceau du « partage » et de la « préservation », doivent donc s'engager sur la piste de la « valorisation » à la manière d'INRAE. Cette valorisation n'est pas celle qui, à l'ère mémorielle qui est la nôtre, pourrait être un signe supplémentaire de ce présent omniprésent qui s'en remet entièrement à la mémoire pour penser l'articulation des temporalités. Non, c'est une « valorisation » qui s'ouvre à la « réutilisation » ; elle n'est pas commémoration, mais nouveau champ des possibles, dont l'épaisseur diachronique nécessaire advient à l'appui d'un environnement numérique où ses dimensions éthiques et politiques<sup>5</sup> doivent être questionnées et la valeur de l'oubli réintroduite, afin de faire de la science ouverte une œuvre de transmission.

---

<sup>1</sup> Patrimoine et numérique. *Technique et politique de la mémoire*, op. cit., p. 219.

<sup>2</sup> *Ibid.*, p. 211.

<sup>3</sup> Pascal Robert, *Mnémotechnologies. Une théorie générale critique des technologies intellectuelles*, op. cit., p. 383.

<sup>4</sup> François Hartog, *Chronos. L'Occident aux prises avec le Temps*, Paris, Gallimard, 2020, p. 278.

<sup>5</sup> Pascal Robert, *Mnémotechnologies. Une théorie générale critique des technologies intellectuelles*, op. cit., p. 381.

# ANNEXES

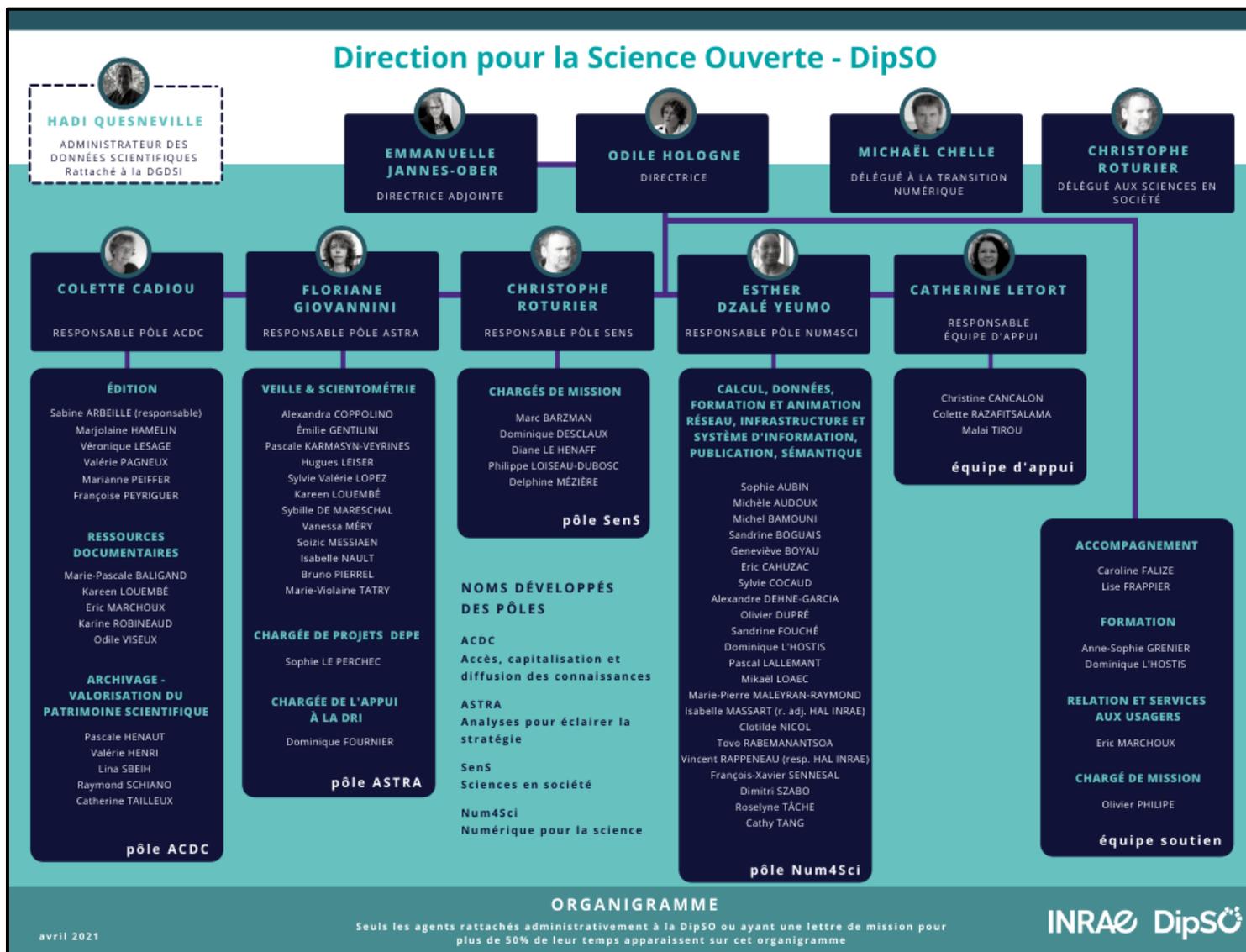
---

## **Table des annexes**

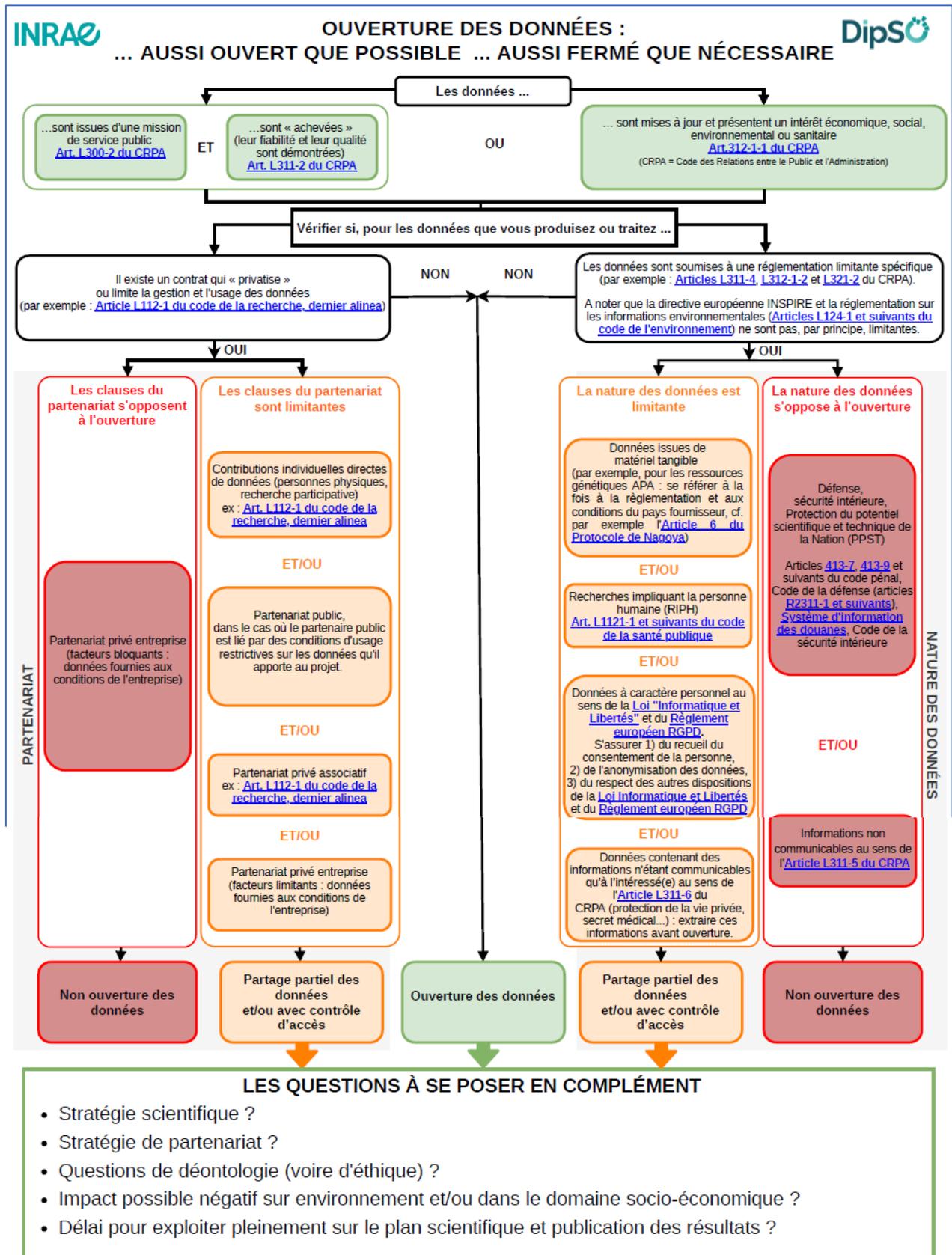
Annexe 1 : Organigramme de la Direction pour la Science Ouverte (DipSO)

Annexe 2 : Processus de diffusion des données de la recherche (DipSO)

# Annexe 1 : Organigramme de la Direction pour la Science Ouverte (DipSO)



## Annexe 2 : Processus de diffusion des données de la recherche



Poster conçu par Olivier Philippe (DipSO), Stephanie Rennes (DAJ), Dimitri Szabo (DipSO), Anne-Sophie Martel (UMR EPIA)  
 08/01/2020



## SOURCES

---

### Entretiens

#### Visioconférence

Pierre Cornu, entretien le 11 juin 2021.

Odile Hologne, entretien le 16 juin 2021.

Colette Cadiou, entretien le 1<sup>er</sup> juillet 2021.

Raymond Schiano, entretien le 1<sup>er</sup> juillet 2021.

Odile Maeght-Bournay, entretien le 19 juillet 2021.

### INRAE

Pierre Adenot, Stéphane Bansard, David Benaben, Veronique Brunaud, Christophe Caron, *et al.*, « AgroDataRing : une infrastructure partagée et mutualisée pour le stockage longue durée », *Cahier des Techniques de l'Inra*, Inra, 2018.

Comité consultatif commun d'éthique pour la recherche agronomique, *Avis 8 sur les enjeux éthiques et déontologique du partage et de la gestion des données issues de la recherche*, Inra, 2016, ffhal-02796585.

Fanny Dedet. « HAL et les archives ouvertes institutionnelles : l'exemple de ProdInra », *L'Open Access, questions économiques, éthiques et juridiques*, Université Clermont Auvergne, octobre 2015, Clermont-Ferrand, hal-02801873.

Fanny Dedet, Anne-Sophie Martel, Windpouire Esther Dzale Yeumo, Sylvie Cocard, « Le portail Data Inra et les services associés », *Cahier des Techniques de l'Inra*, Inra, 2018, ffhal-02622475f.

DipSO, « Une première année riche d'ouverture. Bilan 2020 », INRAE, juin 2021, <https://www6.inrae.fr/dipso/content/download/3649/35674/file/Bilan-DipSO-2020.pdf>.

Esther Dzalé Yeumo, Dominique L'Hostis, « Open Science. Gestion et partage des données de la recherche », Formation URFIST, 22 janvier 2015, Paris, <https://hal.inrae.fr/hal-02800107/document>.

Esther Dzalé Yeumo, Richard Fulss, Rosemary Shrestha, « Wheat Data Interoperability WG », <https://www.rd-alliance.org/group/working-and-interest-group-chairs-wheat-data-interoperability-wg/outcomes/wheat-data>.

Esther Dzalé Yeumo, « DOI & Portail Dataverse : Retour d'expérience de l'Inra », Séminaire « Données scientifiques », le 16 février 2017, Paris.

Christine Gaspin, Dominique Pontier, Laurence Colinet, Frédéric Dardel, Alain Franc, *et al.*, *Rapport du groupe de travail sur la gestion et le partage des données*, Avis du conseil scientifique, Inra, juin 2012.

Odile Hologne, Françoise Boudet-Bône, « FuturIST : un projet pour accompagner l'évolution des métiers de l'information scientifique et technique », *Cahier des Techniques de l'Inra*, Inra, 2012, ffhal-01191338f.

Odile Hologne, *Mise en perspective de l'évolution des métiers de l'IST dans un organisme de recherche*, Frédoc 2013, Octobre 2013, Aussois, France, hal-02810258.

Odile Hologne. « Données de la recherche : rôle des professionnels IST à l'Inra », *Congrès de l'ADBU*, septembre 2013, Le Havre, France, hal-02807047.

Odile Hologne. « Quels services IST dans un contexte d'Open Science », *Bibliothèques et Chercheurs : quelles relations, quelle coopération ?*, 2015.

Inra, « Charte pour le libre accès aux publications et aux données », 2016, Inra, [https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016\\_Charte-libre-access-Inra-Def\\_Inra-fran%C3%A7ais.pdf](https://ist.inrae.fr/wp-content/uploads/sites/21/2020/02/2016_Charte-libre-access-Inra-Def_Inra-fran%C3%A7ais.pdf).

Inra, « Charte de déontologie de l'Inra », 2018, <https://www6.inrae.fr/label-hr-excellence/content/download/3512/33926/version/2/file/charte%2Bde%2Bd%C3%A9ontologie.pdf>.

INRAE, « Données, partenariat et innovation dans un contexte de science ouverte », 2020, [https://datapartage.inrae.fr/content/download/3747/39730/version/1/file/Fiche\\_SO\\_PI\\_Parteneriats.pdf](https://datapartage.inrae.fr/content/download/3747/39730/version/1/file/Fiche_SO_PI_Parteneriats.pdf).

INRAE Forum DipSO, avril 2020, <https://forum.dipso.inrae.fr/t/bienvenue-sur-discourse/7>.

INRAE, « Cycle de vie des données : un outil pour améliorer la gestion, la mise en qualité et l'ouverture des données », DoRaNUM, publié le 25/06/2021, DOI : 10.13143/GDBG-CF63.

INRAE, « INRAE2030. Partageons la science et l'innovation pour un avenir durable », document d'orientation, décembre 2020.

INRAE, « Un avis bienvenu à l'heure du numérique et de la "science ouverte" », [inrae.fr, https://www.inrae.fr/actualites/avis-bienvenu-lheure-du-numerique-science-ouverte](https://www.inrae.fr/actualites/avis-bienvenu-lheure-du-numerique-science-ouverte).

INRAE, « ProdInra. L'archive ouverte de l'Inra », [ist.inrae.fr](https://ist.inrae.fr/le-libre-acces-a-inrae/hal-inrae/prodInra/), consulté le 24 juin 2021, <https://ist.inrae.fr/le-libre-acces-a-inrae/hal-inrae/prodInra/>.

INRAE, « Le libre accès à INRAE », [ist.inrae.fr, https://ist.inrae.fr/le-libre-acces-a-inrae/](https://ist.inrae.fr/le-libre-acces-a-inrae/).

INRAE, « Une direction pour la science ouverte », [inrae.fr, https://www6.inrae.fr/dipso/La-Direction-pour-la-science-ouverte](https://www6.inrae.fr/dipso/La-Direction-pour-la-science-ouverte).

Emmanuelle Jannès Ober, « Problématique de gestion, de diffusion des données », Séminaire « Données scientifiques », le 16 février 2017, Paris.

Mireille Jean, Denis Poupardin, « Les archives des unités de recherche : le point de vue des scientifiques de l'Institut national de la recherche agronomique (I.N.R.A.) », *La Gazette des archives*, n°176, 1997, p. 28 à 49.

Denis Poupardin, Communication, Colloque « Archorales » du 2 décembre 2008, Inra,  
[https://www6.Inrae.fr/comitedhistoire/content/download/3002/30151/version/1/file/Expose\\_Denis\\_Poupardin.pdf](https://www6.Inrae.fr/comitedhistoire/content/download/3002/30151/version/1/file/Expose_Denis_Poupardin.pdf).

Hadi Quesneville, Esther Dzalé, « Gouvernance des données », Café numérique 2, le 1<sup>er</sup> mars 2021.

Lina Sbeih, Fanny Dedet, Patrick Moreau, Esther Dzale, « L'archivage des données de la recherche à l'Inra. Eléments de réflexion, démarche et perspectives », *Cahier des Techniques de l'INRA*, INRA, 2020, ffhal-02861909f.

Odile Vilotte, Francois Houllier, Philippe Mauguin, Claude Ronceray, Christine Cherbut, *et al.*, « Document d'Orientation #Inra2025 », Inra, 2016.

## Divers

Australian National Data Service, “What is research data”, *ANDS Guide*, 2017,  
<https://www.ands.org.au/guides/what-is-research-data>.

Commission européenne, *Horizon 2020*, H2020 Program, Version 5.2, juin 2019, [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/amga/h2020-amga\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf).

« Deuxième plan national pour la science ouverte. Généraliser la science ouverte en France. 2021-2024 », Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, [https://www.ouvrirlascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte\\_2021-2024.pdf](https://www.ouvrirlascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte_2021-2024.pdf).

Direction des services et des réseaux, *Les partenaires de Gallica en 2017*, Bibliothèque nationale de France, février 2018,  
[https://www.bnf.fr/sites/default/files/2018-11/partenaires\\_gallica.pdf](https://www.bnf.fr/sites/default/files/2018-11/partenaires_gallica.pdf).

OCDE, *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, 2007.

Ouvrir la science, <https://www.ouvrirlascience.fr/>.

« Premier Plan national pour la science ouverte », Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 4 juillet 2018.

*La science ouverte : une révolution nécessaire*, 1ère journée pour la science ouverte au CNRS, Paris, 8 octobre 2019.

Anne Thessen, David Patterson, “Data issues in the life sciences”, *ZooKeys*, (150), 2011, <https://doi.org/10.3897/zookeys.150.1766>.

Urfistinfo, « “Données” de la recherche, les mal-nommées », <https://urfistinfo.hypotheses.org/2581>.

# BIBLIOGRAPHIE

---

## SCIENCES DE L'INFORMATION ET DE LA COMMUNICATION

### Enjeux du numérique

BACHIMONT Bruno, « Le numérique comme milieu : enjeux épistémologiques et phénoménologiques », *Interfaces numériques*, volume 4, n° 3, 2015, [En ligne], DOI : 10.25965/interfaces-numeriques.386.

BACHIMONT Bruno, *Patrimoine et numérique : technique et politique de la mémoire*, Paris, INA, 2017, 246 p.

BARZMAN Marco, GERPHAGNON Mélanie, MORA Olivier (coord.), *La transition numérique dans la recherche et l'enseignement supérieur à l'horizon 2040*, Versailles, Editions Quae, 2020, 149 p.

BOULLIER Dominique, *Sociologie du numérique*, Paris, Armand Colin, 2019, 384 p.

CORMERAIS Franck *et al.*, « Les SIC à l'épreuve du digital et des Humanités : des origines, des concepts, des méthodes et des outils », *Revue française des sciences de l'information et de la communication*, n° 8, 2016, DOI: 10.4000/rfsic.1820.

DENIS Jérôme, GOËTA Samuel, « La fabrique des données brutes : Le travail en coulisses de l'open data », MABI Clément, MONNOYER SMITH Laurence (dir.), *Ouvrir, partager, réutiliser : Regards critiques sur les données numériques*, Paris, Éditions de la Maison des sciences de l'homme, 2017, <https://doi.org/10.4000/books.editionsmsh.9050>.

GROUDIEV Stéphanie, « La bibliothèque scientifique numérique : un cadre politique de coordination des actions en faveur de l'IST », *Bulletin des bibliothèques de France*, t. 58, n° 1, 2013.

PROST Hélène, SCHÖPFEL Joachim, « Les entrepôts de données en sciences de l'information et de la communication (SIC). Une étude empirique », *Études de communication. Langages, information, médiations*, n° 52, Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication de l'Université Lille 3, 2019, p. 71 à 98.

ROBERT Pascal, *Mnémotechnologies. Une théorie générale critique des technologies intellectuelles*, 2010, Lavoisier, 400 p.

WOLTON Dominique, « Abondance et gratuité : pour quoi faire et jusqu'où ? Entretien avec Joëlle Farchy, Pascal Froissart et Cécile Méadel », SCHAFER Valérie (dir.), *Information et communication scientifiques à l'heure du numérique*, Paris, CNRS Editions, 2014.

## Les sciences saisies par les SIC

BOUKACEM-ZEGHMOURI Chérifa, BRAVO Blanca Rodríguez, « Une information scientifique, entre évaluation et médiatisation », *Les Enjeux de l'information et de la communication*, vol. 20/2, n° 2, GRESEC, 2019, p. 5 à 11.

CARLINO Vincent, MOLINATTI Grégoire, « Traces numériques et engagement du chercheur : contribution à une éthique de la communication en régime de controverse », *Les Enjeux de l'information et de la communication*, vol. 20, n° 2, GRESEC, 2019, p. 13 à 25.

CAZENAIVE Christine *et al.*, *Conservation et valorisation du patrimoine des organismes de recherche*, FRÉDoc 2006, 3e formation des réseaux de la documentation, Chaumont-sur-Tharonne, 10-11-12 octobre 2006, Saint-Etienne, Publications de l'Université de Saint-Etienne, 2007, 245 p.

CHARVOLIN Florian *et al.*, *Des sciences citoyennes ? La question de l'amateur dans les sciences naturalistes*, La Tour d'Aigues, Éditions de l'Aube, 2007, 253 p.

CLAVERT Frédéric, MULLER Caroline, *Le goût de l'archive à l'ère numérique*, 2018, co-écriture en ligne, <http://www.gout-numerique.net/>.

MABI Clément, PLANTIN Jean-Christophe, MONNOYER-SMITH Laurence, « Interroger les données en SHS à partir de leur écosystème », Valérie Schafer (dir.), *Information et communication scientifiques à l'heure du numérique*, Paris, CNRS Editions, 2014, p. 63 à 78.

SCHAFER Valérie, « Renouveau de la science et science du renouveau à l'heure numérique », SCHAFER Valérie (dir.), *Information et communication scientifiques à l'heure du numérique*, Paris, CNRS Editions, 2014, 9 à 30.

RYGIEL Philippe, *Historien à l'âge numérique*, Villeurbanne, Presses de l'ENSSIB, 2017, 2014 p.

## La science et ses archives

BERT Jean-François, *Qu'est-ce qu'une archive de chercheur ?*, Marseille, OpenEdition press, 2014, 82 p.

CHARMASSON Thérèse, « Archives scientifiques ou archives des sciences : des sources pour l'histoire », *La revue pour l'histoire du CNRS*, n° 14, CNRS Éditions, 2006, DOI: 10.4000/histoire-cnrs.1790.

CORNU Marie, FROMAGEAU Bertrand, MÜLLER Jérôme (dir.), *Archives de la recherche : problèmes et enjeux de la construction du savoir scientifique*, Paris, l'Harmattan, 2014, 205 p.

DELMAS Sophie, FROSSARD Odile, « Les archives du Commissariat à l'énergie atomique et aux énergies alternatives : organisation de la fonction et exploitation des fonds historiques d'un grand organisme public, acteur clé de la recherche », *Gazette des archives*, vol. 231, n° 3, 2013, p. 173-182.

Georges Margot, « Institutions de recherche, chercheurs et archives des savoirs : quelle(s) définition(s) par les producteurs ? », Bert Jean-François, Ratcliff Marc (dir.), *Frontière d'archives : recherches, mémoires, savoirs*, Paris, Editions des archives contemporaines, 2015.

JEAN Mireille, POUPARDIN Denis, « Les archives des unités de recherche : le point de vue des scientifiques de l'Institut national de la recherche agronomique (I.N.R.A.) », *Gazette des archives*, vol. 176, n° 1, 1997, p. 28-49.

LAURENT Sébastien-Yves, *Archives « secrètes », secrets d'archives ? L'historien et l'archiviste face aux archives sensibles*, Paris, CNRS Éditions, 2003, 288 p.

LEBLOND Corinne, *Archivage et stockage pérennes : enjeux et réalisations*, Paris, Lavoisier, 2009, 224 p.

LEFEBVRE Muriel, « L'infra-ordinaire de la recherche. Écritures scientifiques personnelles, archives et mémoire de la recherche », *Sciences de la société*, n° 89, Presses Universitaires du Mirail, 2013, p. 3-17, DOI: 10.4000/sds.203.

MAEGHT-BOURNAY Odile, *Les archives orales de l'Inra et la biographie historique : le cas Jacques Poly*, Mémoire de Master 2 Recherche Histoire, philosophie et didactique des sciences de l'Université Claude Bernard Lyon 1, 88 p.

POMART Julien, « AAF / Section Aurore : Un groupe de travail sur les données de la recherche », *Archives de la Fondation Maison des sciences de l'homme*, 18 juillet 2014, <https://archivesfmsh.hypotheses.org/1209>.

POUPARDIN Denis, « Les archives écrites et orales à l'INRA : des matériaux utiles à son histoire et à sa mémoire », *Bulletin de Liaison de l'Association pour l'Etude de l'Histoire de l'Agriculture au XXe siècle*, supplément au n° 5, 2000.

## ***Open science, open access : ouverture et données de la recherche***

ANDRO Mathieu, HOLOGNE Odile, ANNAÏG Mahé, « Estimation des dépenses de publication de l'Inra dans un modèle théorique "Gold Open Access" », *Documentaliste-Sciences de l'Information*, vol. 51, no. 4, 2014, p. 70-79.

AUMONT Gilles, CHELLE Michaël, DZALE-YEUMO Esther, HOLOGNE Odile, PHILIPPE Olivier, *et al.*, « Principles of data governance for research organizations - INRAE's approach », *Data for policy 2020*, Septembre 2020, hal-02955903.

BOSC Hélène, « Archives ouvertes : quinze ans d'histoire », C. AUBRY, J. JANIK, *Les Archives Ouvertes : enjeux et pratiques. Guide à l'usage des professionnels de l'information*, Paris, ADBS, 2005, p. 27-54.

BOUKACEM-ZEGHMOURI Chérifa, DILLAERTS Hans, « Information scientifique et diffusion des savoirs : entre fragmentations et intermédiaires », *Revue française des sciences de l'information et de la communication*, n° 15, Société Française de Sciences de l'Information et de la Communication, 2018,

<http://journals.openedition.org/rfsic/5522>.

CABRERA Francisca, *Les données de la recherche en Sciences humaines et sociales : enjeux et pratiques. Enquête exploratoire*, Mémoire de fin d'étude, Conservatoire national des arts et métiers, 2014, 238 p.

CHARTON Ghislaine, « L'Open science au prisme de la Commission européenne » dans *Education et sociétés*, vol. 1, n° 41, 2018.

DA SYLVA Lyne, « Les données et leurs impacts théoriques et pratique sur les professionnels de l'information », *Documentation et bibliothèque*, VI. 63, n° 4, octobre-décembre 2017, <https://doi.org/10.7202/1042308ar>.

GASPIN Christine *et al.*, *Rapport du groupe de travail sur la gestion et le partage des données*, 2012, <https://hal.inrae.fr/hal-02809520>.

GENOVA Françoise, *et al.*, « Building a Disciplinary, World-Wide Data Infrastructure », *Data Science Journal*, 16, 2017, DOI: <http://doi.org/10.5334/dsj-2017-016>.

JACQUEMIN Bernard, SCHÖPFEL Joachim, FABRE Renaud, « Libre accès et données de recherche. De l'utopie à l'idéal réaliste », *Études de communication. Langages, information, médiations*, n° 52, Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication de l'Université Lille 3, 2019, p. 11-26.

REBOUILLAT Violaine, *Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs*, thèse de doctorat en sciences de l'information et de la communication, Conservatoire national des arts et métiers, 2019, 493 p.

SCHÖPFEL Joachim (dir.), *La publication scientifique : analyses et perspectives*, Paris, Hermes science publications Lavoisier, 2008, 367 p.

STRASSER Bruno, « Data-driven Sciences : From Wonder Cabinets to Electronic Databases », *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 43, n° 1, mars 2012, p. 85 à 87.

VAN NOORDEN Richard, « Open Access: The True Cost of Science Publishing », *Nature*, 2013, n° 495, p. 426 à 429, <http://dx.doi.org/10.1038/495426a>.

WILKINSON Mark D., DUMONTIER Michel, *et al.*, « The FAIR Guiding Principles for scientific data management and stewardship », *Scientific Data*, 2016, doi: 10.1038/sdata.2016.18.

## SCIENCES EN ETUDE

### Généralités

LECOURT Dominique, *Dictionnaire d'histoire et philosophie des sciences*, Paris, PUF, 2006 [1999], 1195 p.

HARTOG François, *Régimes d'historicité. Présentisme et expériences du temps*, Paris, Editions du Seuil, 2012 [2003], 323 p.

PESSIS Céline, TOPÇU Sezin, BONNEUIL Christophe (dir.), *Une autre histoire des « Trente Glorieuses »*, Paris, La Découverte, 2013, 309 p.

PESTRE Dominique, DAHAN Amy (dir.), *Les sciences pour la guerre*, Paris, Éditions de l'EHESS, 2004, 401 p.

## **Sciences agronomiques**

BONNEUIL Christophe, DENIS GILLES, MAYAUD Jean-Luc, (dir.), *Sciences, chercheurs et agriculture : pour une histoire de la recherche agronomique*, Paris, l'Harmattan, 2008, 300 p.

CORNU Pierre, *La passion naturaliste. Trois études d'anthropologie historique de la « question agraire » à l'époque contemporaine*, Mémoire original pour l'habilitation à diriger des recherches en histoire contemporaine, Université Lyon 2, 2012, p. 340.

HERVIEU Bertrand, HUBERT Bernard, *Sciences en campagne : regards croisés, passés et à venir*, La Tour-d'Aigues, Éditions de l'Aube, 2009, 394 p.

JOLLIVET Marcel (dir.) *Sciences de la nature, sciences de la société : les passeurs de frontières*, Paris, CNRS éditions, 1992, 589 p.

## **Sciences et Etat**

BONNEUIL Christophe, « Engagement public des chercheurs. De la République des savants à la démocratie technique : conditions et transformations de l'engagement public des chercheurs », *Natures, sciences, sociétés*, vol. 14, 2006.

PESTRE Dominique, *Science, argent et politique : un essai d'interprétation*, Paris, France : Institut national de la recherche agronomique, 2003, 201 p.

## **Histoire de l'Inra**

CORNU Pierre, « La recherche agronomique française dans la crise de la rationalité des années soixante-dix », *Histoire de la recherche contemporaine*, t. III, n° 2, 2014, p. 154-166.

CORNU Pierre, VALCESCHINI Egizio, MAEGHT-BOURNAY Odile, *L'histoire de l'INRA, entre science et politique*, Versailles, Editions Quae, 2018, 463 p.

VALCESCHINI Egizio et HUBERT Bernard, « Un parcours dans les mondes de la recherche agronomique. L'Inra et le Cirad », *Histoire de la recherche contemporaine. La revue du Comité pour l'histoire du CNRS*, Tome III-N°2, Comité pour l'histoire du CNRS, 2014, p. 106-112.

## **Philosophie des sciences**

COUTELLEC Léo, *La science au pluriel. Essai d'épistémologie pour des*

*sciences impliquées*, Versailles, Éditions Quae, 2015, 83 p.

KUHN Thomas Samuel, *La structure des révolutions scientifiques*, Paris, Flammarion, 1972, 246 p.

LATOUR Bruno, *Nous n'avons jamais été modernes : essai d'anthropologie symétrique*, Paris, La Découverte, 1991, 210 p.

### **Sociologie des sciences**

BECK Ulrich, *La société du risque : sur la voie d'une autre modernité*, Paris, Flammarion, 2004, 521 p.

CALLON Michel, LASCOUMES Pierre, BARTHE Yannick, *Agir dans un monde incertain. Essai sur la démocratie technique*, Paris, Editions du Seuil, 2001, 437 p.

FLECK Ludwik, *Genèse et développement d'un fait scientifique*, Paris, Les Belles lettres, 2005 [1935], 280 p.

LATOUR Bruno, *La science en action : introduction à la sociologie des sciences*, Paris, Gallimard, 1995, 663 p.

### **Histoire, mémoire et patrimoine**

CAZENAVE Christine, GIRARD Françoise, *Conservation et valorisation du patrimoine des organismes de recherche*, Saint-Etienne, Publications de l'Université de Saint-Etienne, 2007, 212 p.

DESCAMPS Florence, *Archiver la mémoire : de l'histoire orale au patrimoine immatériel*, Paris, Editions de l'EHESS, 2019, 216 p.

DODEBEI Vera, TARDY Cécile (dir.), *Mémoire et nouveaux patrimoines*, Marseille, OpenEdition Press, 2015, 246 p.

HARTOG François, *Régimes d'historicité. Présentisme et expériences du temps*, Paris, Editions du Seuil, 2012 [2003], 323 p.

MUSIANI Francesca, SCHAFER Valérie, « Patrimoine et patrimonialisation numériques », *RESET*, n° 6, 2017, <http://journals.openedition.org/reset/803> ; DOI : 10.4000/reset.803.

### **Anthropologie : entre nature et culture**

LATOUR Bruno, *Le métier de chercheur : regard d'un anthropologue*, Paris, Institut national de la recherche agronomique, 1995, 95 p.

DESCOLA Philippe, *Par-delà nature et culture*, Paris, Editions Gallimard, 2005, 786 p.

## **TABLES DES ILLUSTRATIONS**

---

1. Arborescence des données de la recherche en SHS – Francisca Cabrera, 2014.....	p. 19.
2. Nombre de publications et de rétractions par année - Yaroslav Pigenet, 2014.....	p. 31.
3. Illustration "Principes pour la gouvernance des données" - INRAE, 2018.....	p. 74.
4. Schéma des rôles et responsabilités des acteurs de la donnée - INRAE, 2018 .....	p. 75.
5. Cycle de vie des données de la recherche – Esther Dzalé Yeumo, 2015.....	p. 85.
6. Distribution géographique des données à l’Inra - Esther Dzalé Yeumo, 2016.....	p. 92.
7. Distribution géographique de l’infrastructure AgroDataRing - INRAE, 2020.....	p. 93.
8. Flux des données - « Fiche technique AgroDataRing », INRAE, 2020.....	p. 94.
9. Schéma de Gallica marque blanche - Alix Lamé-Bergis, Isabelle Mangou, 2020.....	p. 111.

# TABLE DES MATIERES

---

<b>SIGLES ET ABREVIATIONS .....</b>	<b>5</b>
<b>INTRODUCTION .....</b>	<b>7</b>
<b>LA SCIENCE OUVERTE ET LE LIBRE ACCES : DE NOUVELLES PRATIQUES POUR UNE NOUVELLE SCIENCE .....</b>	<b>14</b>
<b>La fabrique du savoir.....</b>	<b>14</b>
Les données, matière première de la recherche .....	14
La production des données à l'ère numérique .....	22
<b>Les ambitions du partage des données.....</b>	<b>26</b>
Les sciences en crise .....	26
Les enjeux scientifiques et éthiques .....	31
Les enjeux politiques et économiques .....	36
<b>Une double dynamique internationale et nationale .....</b>	<b>41</b>
La science ouverte à l'international.....	41
Le mouvement en France .....	45
Le cadre juridique .....	49
<b>INRAE ET LA GESTION DES DONNEES DE LA RECHERCHE.....</b>	<b>55</b>
<b>L'engagement précurseur des instituts français de recherche agronomique .....</b>	<b>55</b>
La recherche agronomique française à l'heure de l'ouverture .....	55
La gestion des données de la recherche agronomique à l'ère du village mondial .....	62
Les premières pierres : réinventer l'édition scientifique.....	65
<b>La gouvernance des données à INRAE .....</b>	<b>71</b>
Les principes de la gouvernance .....	71
L'information scientifique et technique : une reconfiguration plurielle ...	76
Les producteurs de données : une sensibilisation de tous les instants .....	81
<b>La gestion des données de la recherche : de l'information scientifique à la technique.....</b>	<b>84</b>
Le cycle de vie des données à INRAE.....	84
L'environnement technique de la gouvernance des données à INRAE.....	88

**LES SCIENCES AGRONOMIQUES EN PERSPECTIVE : HISTOIRE  
ET EPISTEMOLOGIE EN ENVIRONNEMENT NUMERIQUE ..... 95**

**Faire archive : intentionnalité et trace de la donnée patrimoniale  
scientifique..... 95**

Archives des sciences, archives de chercheur, archives de la recherche... 95

Des archives orales au Comité d'Histoire : INRAE et ses archives ..... 99

INRAE, une nouvelle dynamique transverse .....105

Les enjeux de la patrimonialisation des archives de la recherche .....107

**Ecrire l'histoire de la recherche agronomique en environnement  
numérique.....114**

L'histoire en environnement numérique et ouvert.....114

L'histoire numérique des sciences.....119

**CONCLUSION.....123**

**ANNEXES .....125**

Table des annexes .....125

Annexe 1 : Organigramme de la Direction pour la Science Ouverte .....126

Annexe 2 : Processus de diffusion des données de la recherche .....127

**SOURCES .....128**

**BIBLIOGRAPHIE .....131**

**TABLES DES ILLUSTRATIONS .....137**

**TABLE DES MATIERES .....138**