

Diplôme national de master

Domaine – sciences humaines et sociales

Mention – sciences de l'information et des bibliothèques

Parcours – sciences de l'information et des bibliothèques et
information scientifique et technique

MEMOIRE

Science ouverte et gestion des données : Défis et stratégies dans les laboratoires océanographiques français

Marine FOURRIER

Sous la direction d'Agnieszka TONA
Maîtresse de conférences – ENSSIB

Remerciements

Je souhaite tout d'abord exprimer ma reconnaissance aux personnes interrogées dans le cadre de ce mémoire, dont les contributions et le partage d'expérience ont constitué une ressource essentielle à l'élaboration de ce mémoire.

J'adresse également mes remerciements à ma tutrice, pour l'accompagnement dont elle a fait preuve, la confiance accordée dans le choix d'un sujet atypique, ainsi que pour ses conseils aux moments clés de l'avancement de ce travail.

Enfin, je remercie l'ENSSIB pour la qualité de son enseignement et de son encadrement, qui m'ont permis d'acquérir les connaissances et les compétences nécessaires à la réalisation de ce mémoire.

Enfin merci Chris pour ton soutien au cours de cette année pas toujours évidente.

Résumé :

Ce mémoire examine comment les acteurs des laboratoires océanographiques français s'approprient les politiques de science ouverte plus précisément sur les données de recherche. Les résultats révèlent un écart entre les injonctions institutionnelles et les pratiques réelles, façonnées par des normes disciplinaires, des différences générationnelles et des contraintes de ressources. Alors que certains adoptent pleinement le partage des données, d'autres adoptent des approches sélectives ou stratégiques. L'étude met en évidence les impacts organisationnels et propose des recommandations pratiques pour améliorer la formation, le soutien et la reconnaissance du travail de gestion des données.

Descripteurs :

Science ouverte ; Données de recherche ; Océanographie ; Principes FAIR ; Pratiques scientifiques.

Abstract :

This thesis examines how those involved in French oceanographic laboratories are adopting open science policies, specifically with regard to research data. Findings reveal a gap between institutional mandates and actual practices, shaped by disciplinary norms, generational differences, and resource constraints. While some fully embrace data sharing, others adopt selective or strategic approaches. The study highlights the organizational impacts and offers practical recommendations to improve training, support, and recognition for data management work.

Keywords :

Open science; Research data; Oceanography; FAIR principles; Scientific practices.

Droits d'auteurs

Droits d'auteur réservés.

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles et respectant la mention de confidentialité ci-dessus est prohibée.

Sommaire

SIGLES ET ABREVIATIONS	7
INTRODUCTION.....	9
I. COMPRENDRE LE PAYSAGE : CONCEPTS, OUTILS ET REFERENTIELS	11
1. Définir et caractériser les données océanographiques	11
2. Stocker et diffuser : infrastructures, portail et diversité des pratiques	16
3. Les principes FAIR comme cadre du partage des données	19
II. CADRES DE REFERENCE ET OBSTACLES AU PARTAGE DES	
DONNEES DE RECHERCHE	21
1. Environnement institutionnel et juridique de la science ouverte	21
1. Orientation des acteurs institutionnels	21
2. Règles, obligations et zones grises du cadre juridique	22
2. Les pratiques et tensions au cœur du travail scientifique.....	23
1. Entre ouverture et rétention : une lecture socio-anthropologique	23
2. Science ouverte et carrière académique : concilier injonctions et	
intérêts individuels	25
III. METHODOLOGIE ET DEMARCHE D'ENQUETE	27
1. Une approche qualitative fondée sur un corpus mixte	27
1. Le corpus textuel : sources et critères de sélection	27
2. Le corpus d'entretiens : choix des enquêtés et grille d'entretien	31
2. Méthodes d'analyse : de la grille d'entretien à la confrontation avec	
la littérature	33
IV. ENTRE PRESCRIPTIONS ET REALITES : APPROPRIATION	
CONTRASTEE DES INJONCTIONS.....	34
1. Ce que disent les textes : un cadrage institutionnel de la gestion des	
données	34
2. Ce que font les acteurs : des conceptions situées et contextualisées des	
données	36
3. Les infrastructures au milieu des pratiques de gestion	39
4. Science ouverte en pratique : entre incitation, obligation et stratégies	
d'évitement.....	42
5. Ouvrir... mais sous conditions : arbitrages, tensions et rapports de	
force	45

6. Ressources humaines, organisation et leviers d'action pour renforcer la gestion des données.....	51
CONCLUSION.....	57
CORPUS.....	61
BIBLIOGRAPHIE	63
ANNEXES	68
Annexe 1 : Grille d'entretien	69
Annexe 2 : Grille d'analyse du corpus textuel	71
TABLE DES ILLUSTRATIONS	77
TABLE DES MATIÈRES	79

Sigles et abréviations

ANR : Agence Nationale de la Recherche.
BAP : Branche d'Activité Professionnelle.
CDS : Centre de Données et de Services.
CNRS : Centre National de la Recherche Scientifique.
CNU : Conseil National des Universités.
COPIL : Comité de Pilotage.
CSV : Comma-Separated Values.
CTD : Conductivity Temperature Depth.
DDOR : Direction des Données Ouvertes de la Recherche.
DMP : Data Management Plan.
DOI : Digital Object Identifier.
EMODnet : European Marine Observation and Data Network.
ENA : European Nucleotide Archive.
FAIR : Principes Findable, Accessible, Interoperable, Reusable (données ouvertes).
GBIF : Global Biodiversity Information Facility.
GOSHIP : Global Ocean Ship-based Hydrographic Investigations Program.
IRD : Institut de Recherche pour le Développement.
IST : Information Scientifique et Technique.
MESRI : Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation.
NetCDF : Network Common Data Form.
NOC : National Oceanography Centre.
NSF : National Science Foundation.
OBIS : Ocean Biogeographic Information System.
ODATIS : Pôle de données et de services pour l'océan.
OI : Ocean Informatics.
OSU : Observatoire des Sciences de l'Univers.
PANGAEA : Base de données multidisciplinaire pour les sciences de la Terre.
PGD : Plan de Gestion des Données.
PPR : Programme Prioritaire de Recherche.
SeaDataNet : Infrastructure européenne de données océanographiques.
SEANOE : Archive ouverte pour la recherche océanographique.
Sextant : Portail géographique donnant accès aux données marines et littorales.
SHS : Sciences Humaines et Sociales.
SISMER : Systèmes d'informations scientifiques de la Mer.
SNO : Service National d'Observation.
WoRMS : World Register of Marine Species.
ZENODO : Plateforme d'archivage sciences ouvertes (généraliste).

INTRODUCTION

Depuis une quinzaine d'années, la science ouverte s'impose comme un horizon stratégique majeur pour les politiques de recherche. Elle repose sur le postulat que la diffusion large et rapide des résultats, en particulier des données, permet d'accélérer la production de connaissances, de favoriser la collaboration interdisciplinaire et de renforcer la confiance entre science et société. Dans le domaine océanographique, où l'acquisition de données est au cœur de l'activité scientifique, la question de leur stockage, de leur partage et de leur réutilisation prend une dimension particulière. Les données collectées, qu'il s'agisse de paramètres physiques, chimiques, biologiques ou génomiques, sont précieuses pour comprendre les dynamiques océaniques et pour répondre aux enjeux environnementaux globaux. Leur mise à disposition peut bénéficier à un large éventail d'utilisateurs, des chercheurs académiques aux acteurs industriels, en passant par les institutions publiques et les décideurs politiques (Caruso et al. 2022). Pourtant, de nombreuses données produites ne sont pas partagées, ce qui freine le développement de modèles prédictifs fiables sur l'état des océans et limite la portée collective des efforts de recherche (Pendleton et al. 2019).

C'est dans ce contexte que s'inscrit ce mémoire, nourri par un double ancrage personnel : une carrière initiale dans la recherche académique française en océanographie (thèse puis post-doctorats) et une reconversion professionnelle dans la documentation scientifique et technique via le suivi du master SIBIST. Ce parcours m'a offert une connaissance fine des environnements scientifiques océanographiques, des formats et standards de données, ainsi que des contraintes techniques propres à leur gestion. L'idée de ce mémoire est née d'un étonnement persistant : celui du décalage entre le discours institutionnel ambitieux sur la science ouverte et les pratiques observées par le passé dans les laboratoires de recherche. Cet étonnement m'a conduit à interroger la manière dont ces injonctions, qui me semblaient parfois perçues comme des obligations, parfois comme de simples recommandations, sont effectivement connues, comprises, intégrées ou contournées par les acteurs de la recherche océanographique française.

La problématique de ce mémoire peut se formuler ainsi :

Est-ce que, et de quelle manière, les acteurs des laboratoires de recherche océanographique en France s'approprient les injonctions institutionnelles de stockage, de partage et de gestion des données, et quelles en sont les éventuelles implications sur l'organisation du travail de recherche ?

L'hypothèse principale est que les politiques institutionnelles en faveur de la science ouverte se heurtent aux logiques de compétition académique et aux incertitudes juridiques, conduisant à une adoption sélective, parfois stratégique, des pratiques de partage des données.

L'objectif de ce travail est double : d'une part, analyser la diversité des représentations et des pratiques relatives à la gestion et au partage des données dans les laboratoires océanographiques ; d'autre part, mettre en lumière les tensions, arbitrages et ajustements organisationnels que suscitent les politiques de science ouverte. La démarche repose sur l'étude conjointe de deux corpus : un corpus textuel issu de documents institutionnels et scientifiques, et un corpus d'entretiens semi-directifs menés auprès d'acteurs variés du monde océanographique.

Ce mémoire est structuré en quatre grandes parties. La première définit les données océanographiques, leurs lieux de stockage et leurs cadres techniques et normatifs, en particulier les principes FAIR. La deuxième examine les contextes institutionnel et juridique de la science ouverte ainsi que ce que l'état de l'art montre des tensions qu'ils suscitent dans la pratique. La troisième présente la méthodologie de recherche de ce mémoire, fondée sur un corpus mixte et une analyse qualitative. Enfin, la quatrième développe l'analyse des deux corpus, confronte les prescriptions institutionnelles aux pratiques observées, met en évidence les stratégies d'appropriation ou d'évitement et identifie des leviers pour améliorer la gestion des données.

I. COMPRENDRE LE PAYSAGE : CONCEPTS, OUTILS ET REFERENTIELS

La gestion des données dans les laboratoires océanographiques français ne peut être correctement appréhendée sans une clarification de certains concepts. Cette première partie vise à poser les définitions et les enjeux fondamentaux liés à la nature de données scientifiques (plus particulièrement en océanographie), à leur stockage (via des infrastructures ou entrepôts dédiés ou non), et aux principes FAIR¹, devenus un pilier du mouvement de science ouverte.

1. DEFINIR ET CARACTERISER LES DONNEES OCEANOGRAPHIQUES

La notion de données de recherche est polysémique et variable selon les disciplines. Dans un sens large, les données scientifiques désignent tout contenu structuré ou structurable issu d'une activité de recherche, utilisé comme preuve ou matériau d'analyse. L'Inist-CNRS (2022) propose la définition suivante : « Les données de la recherche sont [...] tous les éléments factuels qu'un chercheur peut produire ou collecter et dont il a besoin pour valider les résultats de sa recherche. Il est important de noter qu'il existe plusieurs définitions des données de la recherche. Elles restent une notion difficile à définir car elles englobent un panel diversifié de réalités. D'une discipline à une autre, ou d'un point de vue institutionnel à un autre, la définition des 'données de la recherche' est plus ou moins restrictive. »

Dans le domaine de l'océanographie, cette définition se décline en fonction de spécificités propres aux sciences marines, qui mêlent des observations naturalistes, des mesures physiques, chimiques, biologiques ou géologiques, et un fort ancrage technologique. Plus précisément, les mesures océaniques recouvrent des thématiques disciplinaires couvrant la biologie marine (étude d'organismes marins et de toxines), la biogéochimie marine (étude d'éléments chimiques dissous ou en suspension dans l'eau de mer comme le CO₂, les nutriments, etc.), la physique de l'océan (courants, température, etc.), la géologie (bathymétrie, débits fluviaux en mer), la météorologie (précipitation, aérosols, etc.), pour ne citer que les principales. Cette « science de l'océan » représente en réalité « des sciences de l'océan » impliquant des typologies de données variées. Il est ainsi difficile de généraliser sur ce que sont les données de la recherche en océanographie. Selon Zhu et al. (2024), les données océanographiques peuvent être définies comme : « des informations quantitatives ou qualitatives collectées dans des environnements marins, à l'aide d'outils instrumentaux, de campagnes de terrain, ou de simulations

¹ FAIR : Findable, Accessible, Interoperable, Reusable. Principes d'ouverture des données.

numériques, visant à comprendre les dynamiques physiques, biologiques et chimiques des océans. »

L'étude de l'océan nécessite l'utilisation de nombreuses plateformes complémentaires. Les techniques de mesure océanographiques ont grandement évolué avec les développements technologiques. De nos jours, l'océan est étudié via des mesures *in situ* effectuées depuis des navires océanographiques, via des plateformes autonomes (ou semi-autonomes) déployées dans les océans ou encore via des techniques d'observation de l'océan depuis l'espace, parmi d'autres.

Les types de données rencontrées en océanographie incluent donc des mesures *in situ* issues de campagnes en mer, des prélèvements d'échantillons biologiques ou géochimiques, des données expérimentales issues de laboratoires ou de cultures biologiques ou encore des données issues de capteurs autonomes ou de satellites. Pour autant, une réalité commune des données océanographiques, contrairement aux données issues de certaines autres disciplines, consiste en la difficulté d'acquisition de ces données. La complexité de ces données s'accompagne de coûts humains et logistiques élevés : campagnes en mer longues et coûteuses (humainement et financièrement), équipements sensibles, risques de contamination, incertitudes sur les résultats (notamment pour les travaux impliquant du vivant). Ces contraintes ont un impact direct sur la gestion et le partage des données. En effet, l'investissement humain et financier nécessaire pour obtenir des données peut être conséquent.

Il me paraît important d'introduire ici quelques éléments de contexte plus personnels. Au fil de mes années dans le monde académique de l'océanographie, j'ai eu l'opportunité de participer à diverses campagnes en mer à bord de différents navires des flottes océanographiques française et belge. Bien que ce type d'expérience reste rare et précieuse, je souhaite insister sur ce qu'elle implique concrètement en termes de coût humain et de charge de travail.

Selon la taille de l'équipage et les spécialités représentées à bord, une journée de travail peut facilement dépasser les 14 heures, entre la collecte des échantillons et leur traitement immédiat. Ce rythme intense se maintient parfois sur plusieurs semaines, sans interruption. À bord, il n'existe ni week-end ni jours fériés ; le travail de nuit est courant, et les conditions de mer (houle, tempêtes) peuvent rendre les journées encore plus éprouvantes. Chaque jour de mer est compté. Le temps alloué à une campagne est limité, et tous les scientifiques à bord sont pleinement conscients de la valeur des données produites non seulement parce qu'elles sont difficiles à obtenir, mais aussi parce qu'elles sont financées par de l'argent public et les campagnes océanographiques sont extrêmement coûteuses. Cette conscience collective conduit à une véritable volonté d'optimisation de chaque opération.

À terre, il faut parfois plusieurs semaines pour retrouver un rythme de vie normal, notamment après avoir travaillé sur des horaires nocturnes. Et le travail est loin d'être terminé. En plus du traitement des données, qui peut mobiliser les chercheurs pendant des semaines, voire des mois selon les spécialités : d'un contrôle qualité relativement rapide pour les données physiques, à de longues analyses en laboratoire pour les

paramètres chimiques, jusqu'à des observations prolongées au microscope ou à la loupe binoculaire pour identifier des espèces biologiques, toutes les autres responsabilités, missions et tâches reprennent également. Enseignement, encadrement, rédaction de publications, réponses à appels à projets, gestion administrative : il faut jongler avec l'ensemble de ces obligations en parallèle du suivi post-campagne, ce qui ajoute une pression supplémentaire.

Par ailleurs, malgré toute l'anticipation possible, certaines manipulations ou prélèvements échouent en mer : capteurs défaillants, contaminations d'échantillons, erreurs humaines... la fatigue cumulée et le rythme soutenu peuvent y contribuer, et ces aléas, parfois liés à de simples coups de malchance, rappellent la fragilité inhérente à la production de données océanographiques.

Ce témoignage n'a pas vocation à ajouter une dimension émotionnelle au propos, mais plutôt à donner un aperçu concret de la réalité du travail scientifique en océanographie, une réalité qui reste souvent méconnue en dehors du milieu et à contextualiser certains verbatims qui seront issus des entretiens.

En complément de l'échantillonnage traditionnel via les campagnes en mer, dans les années 2000 a été lancé le programme Argo (Roemmich et al. 2019, 2022) dédié à l'observation des propriétés physiques des océans. Les flotteurs profileurs sont des plateformes autonomes lagrangiennes qui dérivent au grès des courants. Elles contrôlent leur flottabilité en modifiant leur volume grâce à un système de vessie hydraulique, permettant au flotteur de se déplacer verticalement sur la colonne d'eau à bas coût énergétique. Le programme Argo est un programme international impliquant des scientifiques issus de dizaines de pays et comprenant un réseau mondial de plus de 4000 flotteurs profileurs actifs de nos jours. Les flotteurs profileurs sont le plus souvent déployés depuis des navires océanographiques ou des navires d'opportunité. Les flotteurs fonctionnent selon des cycles de 10 jours (Figure 1) impliquant une phase de descente, une phase de dérive en profondeur et une phase de remontée. Durant cette phase de mouvement vertical (et notamment lors de l'ascension depuis plus de 2000 mètres de profondeur), les capteurs du flotteur sont activés et acquièrent des données, ce qui permet d'obtenir de profils verticaux. Une fois en surface, les données sont transmises par satellite à des centres de données à terre. Le flotteur replonge ensuite et entame un nouveau cycle de mesures, et ce jusqu'à ce que sa batterie se vide au bout de plusieurs années (en moyenne > 5 ans) ou que le flotteur soit récupéré dans certains cas.

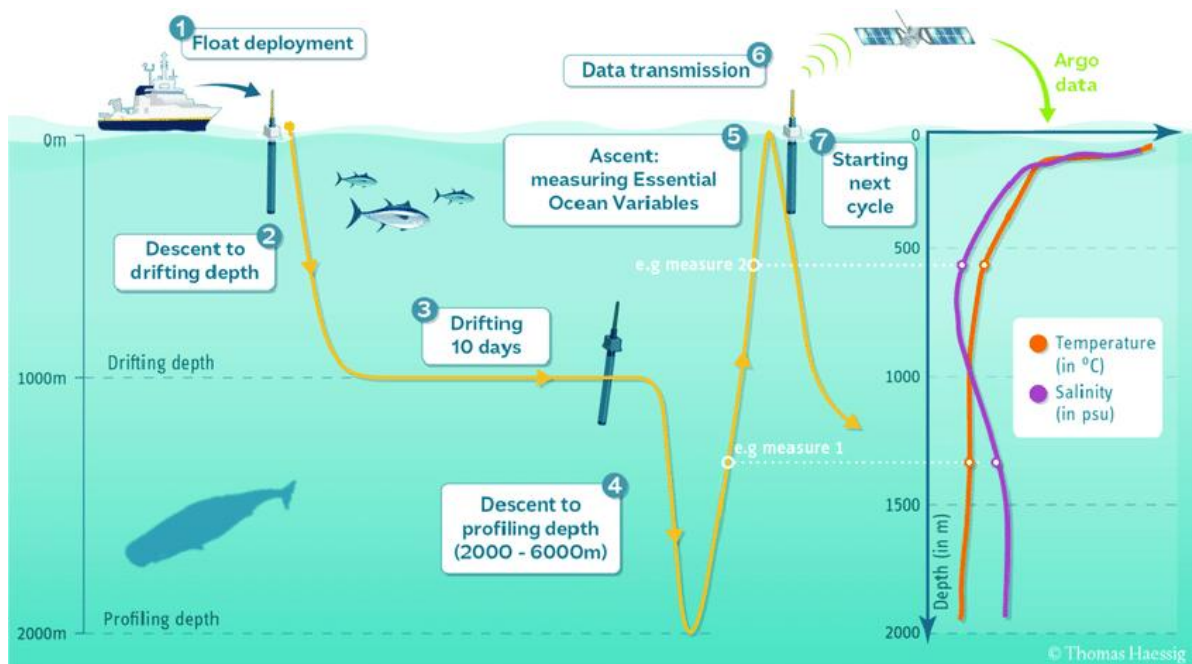


Figure 1 : Cycle de vie d'un flotteur profileur Argo. ©Thomas Haessig.

Les flotteurs profileurs permettent de couvrir de larges échelles de temps (phénomènes journaliers, saisonniers à pluriannuels) et d'espace (de quelques kilomètres à l'échelle d'un bassin océanique) complétant ainsi les dispositifs plus classiques d'échantillonnage du milieu marin principalement basés sur les campagnes à bord de navires océanographiques.

Ces dispositifs ont révolutionné certaines disciplines océanographiques. Il est essentiel de préciser qu'une spécificité du programme Argo est que l'ensemble des données mesurées par ces plateformes autonomes est en accès libre à n'importe qui dans le monde entier. Les données, accessibles librement et gratuitement, offrent des avantages tangibles à la société. Elles permettent de prendre des décisions éclairées et fondées dans un large éventail de domaines : sécurité maritime, prévisions météorologiques et climatiques, gestion durable des ressources marines, résilience des écosystèmes et adaptation aux événements extrêmes. Ces données sont utilisées également par des scientifiques de pays ayant des moyens limités et qui n'auraient jamais pu collecter de telles données par des moyens classiques (échantillonnage en mer). Ces notions de non-exclusivité et de partage sont inscrites dans les textes fondateurs du programme Argo et ont été mises en place dès son début.

A l'origine, les flotteurs profileurs étaient équipés uniquement de capteurs pour des variables physiques simples : pression, température et salinité. Au vu du succès du programme Argo et des développements technologiques (notamment la miniaturisation des capteurs), le programme BGC-Argo (pour BioGeoChemical-Argo) a officiellement vu le jour en 2016 (Claustre, Johnson, et Takeshita 2020; Roemmich et al. 2019) et de nouvelles variables océanographiques plus complexes (oxygène dissous, nutriments, mesures bio-optiques ou encore mesure de pH) ont pu être mesurées via ces plateformes.

Des travaux comme ceux de Sabina Leonelli et de Joachim Schöpfel ont souligné le caractère fondamentalement contextuel des données de recherche. Pour Leonelli

(2015), une donnée n'a pas d'existence indépendante ou universelle : ce qui est considéré comme une « donnée » dépend toujours d'un cadre de recherche spécifique. Autrement dit, les données ne se définissent pas par leurs seules propriétés intrinsèques, mais par leur rôle dans un processus de recherche. Répondre à la question « qu'est-ce qu'une donnée ? » implique donc de s'ancrer dans des situations concrètes de recherche. Dans le même esprit, Schöpfel, Kergosien, et Prost (2017) identifient quatre dimensions clés qui influencent la nature d'une donnée scientifique : sa nature factuelle, son enregistrement, la communauté qui la génère ou l'utilise, et enfin, sa finalité. Par ailleurs, Heaton et Millerand (2013) soulignent l'importance croissante des métadonnées dans ces domaines, qui deviennent indispensables à la compréhension et à la réutilisation des jeux de données.

Les données produites en océanographie et en biologie marine présentent une grande diversité de formats et de structurations. Certaines sont hautement standardisées, notamment dans le domaine des sciences physiques de l'océan, où l'on retrouve fréquemment des fichiers au format NetCDF². Ce format, conçu pour le stockage de données scientifiques multidimensionnelles, est bien adapté à des jeux de données en « 4D » (longitude, latitude, profondeur, temps), usuels en océanographie opérationnelle ou dans les observations longues séries. Ces fichiers sont généralement associés à des variables qui, dans le meilleur des cas, utilisent des référentiels ou des vocabulaires contrôlés (unités, etc.), facilitant leur interopérabilité.

Cependant, cette normalisation coexiste avec des formats plus hétérogènes, notamment en chimie ou en biologie marine, où les types de données (taxinomie, traits fonctionnels, variables biogéochimiques, etc.) et leurs modalités de collecte diffèrent largement d'un projet à l'autre. Il n'est pas rare de trouver des données stockées dans des fichiers Excel aux formatages variés, des bases Access ou même sous forme manuscrite dans des cahiers de laboratoire avec une retranscription purement manuelle. Ces formats, parfois encore très utilisés dans certaines équipes, posent des difficultés en termes de conservation, de partage ou de réutilisation et peuvent être source d'erreurs. De plus, l'absence de référentiels partagés ou la variation locale des conventions de nommage compliquent l'intégration des données dans des portails ou entrepôts communs. Cette hétérogénéité, qui reflète aussi la richesse des approches et des objets d'étude, constitue un défi majeur pour l'application concrète des principes de science ouverte et pour toute politique de gestion unifiée des données à l'échelle des institutions ou nationale.

² Network Common Data Form. Format standard pour données scientifiques multidimensionnelles.

2. STOCKER ET DIFFUSER : INFRASTRUCTURES, PORTAIL ET DIVERSITE DES PRATIQUES

La question du stockage des données est centrale dans les dynamiques de science ouverte. Les chercheurs et les institutions doivent faire des choix techniques (formats, plateformes, accès) mais aussi politiques (centralisation, souveraineté, interopérabilité). Le stockage des données scientifiques repose sur un écosystème de plateformes, entrepôts, portails aux logiques variées : certains sont nationaux (SEANOE³, Recherche Data Gouv), d'autres européens (EMODnet⁴, SeaDataNet⁵), ou encore portés par des communautés disciplinaires spécifiques.

La promotion de la transparence et de l'accessibilité de la science est particulièrement importante compte tenu de sa nature technique, qui la rend difficile à comprendre pour la grande majorité des citoyens. Une expertise et des infrastructures adéquates sont nécessaires pour évaluer la fiabilité et l'importance potentielle des données, garantissant ainsi que les parcours de données peuvent contribuer au développement de connaissances perçues comme pertinentes et utiles dans une variété de domaines (Leonelli 2013).

Les entrepôts de données de recherche remplissent quatre fonctions essentielles : décrire, archiver, identifier et rendre consultables les jeux de données. Ils assurent la pérennité des données tout en facilitant leur découverte et leur réutilisation, notamment grâce à l'attribution d'identifiants pérennes (comme des DOI⁶) et à l'ajout de métadonnées normalisées. Ces fonctions sont au cœur des politiques de science ouverte. Dans ce contexte, émerge la notion d'entrepôt de données de confiance, en réponse à la nécessité de garantir la fiabilité, la pérennité et la transparence des services proposés. Un entrepôt est considéré comme « de confiance » s'il respecte les critères définis dans le guide *Criteria for the Selection of Trustworthy Repositories*, élaboré par un groupe d'organisations internationales telles que CoreTrustSeal et DataCite. Ces critères incluent notamment la clarté de la gouvernance, la transparence des politiques d'accès, la sécurité des données et l'engagement à long terme en matière de préservation.

En France, le Collège des Données de la Recherche du Comité pour la Science Ouverte s'est emparé de cette question et propose une méthodologie d'identification des entrepôts thématiques de confiance, accompagnée d'une première liste d'entrepôts reconnus. Cette démarche vise à orienter les chercheurs vers des infrastructures fiables et compatibles avec les exigences institutionnelles, tout en valorisant les initiatives disciplinaires déjà en place.

³ SEANOE : archive ouverte pour la recherche océanographique. Plateforme de dépôt et de diffusion des jeux de données marines.

⁴ EMODnet : European Marine Observation and Data Network. Initiative de l'Union européenne pour centraliser et diffuser les données marines.

⁵ SeaDataNet : infrastructure européenne de gestion et d'accès aux données océanographiques.

⁶ DOI : Digital Object Identifier. Identifiant pérenne pour les publications ou données.

En océanographie, plusieurs entrepôts illustrent concrètement la dynamique de structuration et de diffusion des données scientifiques. SEANOE, porté par l’Ifremer, est un entrepôt français spécialisé dans les données marines, reconnu au niveau international (Merceur, Petit De La Villeon, et Van Iseghem 2023). Il permet la mise à disposition de jeux de données issus de la recherche marine avec un accompagnement éditorial et répond aux critères d’un entrepôt de confiance. À l’échelle européenne, d’autres infrastructures comme EMODnet (European Marine Observation and Data Network ; Martín Míguez et al. 2019), SeaDataNet (Schaap et Lowry 2010) ou encore PANGAEA⁷ assurent également l’archivage et la diffusion de données environnementales marines. Ces plateformes visent à centraliser l’accès à des données hétérogènes, à en garantir l’interopérabilité et à promouvoir la standardisation des métadonnées entre disciplines et pays. Dans cette perspective, Richou & Schöpfel (2023) défendent l’idée d’un entrepôt national fédéré, garantissant à la fois la souveraineté des données produites par la recherche publique française et leur large diffusion.

La coexistence de ces plateformes engendre une multiplicité des pratiques de stockage : certains chercheurs déposent leurs données dans plusieurs plateformes selon les exigences des financeurs ou des revues, d’autres conservent localement les jeux de données, sans dépôt officiel. Certains de ces choix sont faits selon la visibilité, la facilité d’usage, ou la compatibilité avec les outils de traitement. Selon Schöpfel (2023), cette hétérogénéité révèle une tension entre centralisation et autonomie, entre normes de visibilité et logiques internes aux laboratoires. Cette diversité est aussi mise en évidence par Millerand et Baker (2020), qui décrivent un paysage fragmenté des infrastructures de données, souvent peu lisible pour les chercheurs eux-mêmes.

Dans le cadre de ce travail, nous nous inscrivons dans l’approche de Larroche (2018) en considérant les entrepôts de données comme des dispositifs numériques, c’est-à-dire comme des outils permettant la médiation de l’information scientifique et ainsi d’étudier les entrepôts dans leur dimension sociale et non seulement dans leur dimension technique.

Il existe différents types d’entrepôts selon les typologies de Rücknagel et collaborateurs (2015) :

- Les entrepôts institutionnels liés à une institution en particulier et couvrant généralement plusieurs disciplines de recherche.
- Les entrepôts gouvernementaux gérés par des institutions gouvernementales.
- Les entrepôts disciplinaires qui couvrent une discipline générale et reçoivent des contributions issues de différentes institutions.
- Les entrepôts multidisciplinaires.
- Les entrepôts relatifs à un projet de recherche particulier.

⁷ PANGAEA : Base de données multidisciplinaire pour les sciences de la Terre.

Pour mieux comprendre les usages, contraintes et besoins associés à ces infrastructures, il est nécessaire de replacer les entrepôts dans une vision systémique : celle du cycle de vie des données. Le cycle de vie des données en science ouverte (Figure 2) décrit l'ensemble des étapes que traverse une donnée, depuis sa conception jusqu'à sa réutilisation. Il constitue un cadre de référence pour structurer les pratiques de gestion, favoriser leur standardisation, et accompagner les chercheurs dans des démarches plus rigoureuses, durables et interopérables.

Les différentes phases de ce cycle peuvent être décrites comme suit : planification, collecte, analyse, documentation, stockage, conservation, exposition et réutilisation. La planification est une étape d'anticipation, de réflexion, de formation si besoin et d'organisation, en amont du projet de recherche (c'est à cette étape notamment qu'est rédigée la première version du plan de gestion de données). La collecte est l'étape au cours de laquelle le chercheur acquiert/génère/rassemble des données à des fins d'analyse. L'analyse est comme son nom l'indique une étape d'analyse des données, pouvant requérir un traitement informatique spécifique (logiciel scientifique, calcul intensif...). La documentation correspond à une étape de description des données à l'aide de métadonnées et d'informations complémentaires si nécessaire. Le stockage est l'étape consistant à sauvegarder les données et leurs métadonnées sur un support de stockage (disque dur, cloud...) jusqu'au terme du projet de recherche. La conservation constitue l'étape au cours de laquelle le chercheur confie à une infrastructure dédiée l'archivage d'une sélection de données qu'il souhaite conserver sur le long terme. L'exposition est l'étape de diffusion des données (le chercheur peut les déposer dans un entrepôt et/ou les décrire dans un *data paper*). Enfin la réutilisation correspond à l'étape au cours de laquelle un chercheur souhaite collecter des données pour les exploiter, les intégrer à d'autres données. Cette étape sera d'autant facilitée que les données seront faciles à trouver, accessibles, intelligibles et exploitables par des machines.



Figure 2 : Cycle de vie des données de recherche (d'après Cat-OPIDoR).

Comme montré par Rebouillat et Schöpfel (2023), les entrepôts ne couvrent qu'une portion réduite du cycle de vie des données de recherche : conservation, exposition et réutilisation (certains entrepôts permettant également le stockage et la documentation). En effet, ils ne permettent pas l'étape d'analyse car ne proposent pas de plateforme de calcul.

Austin et collaborateurs (2017) rappellent que la publication des données a lieu soit via des entrepôts de données, soit par des *data papers* via des *data journals* dédiés, qui garantissent que les objets de recherche puissent être documentés, conservés et archivés, mais aussi respectent les principes FAIR permettant la réutilisation future des données par d'autres utilisateurs.

3. LES PRINCIPES FAIR COMME CADRE DU PARTAGE DES DONNEES

Dans le respect des dispositions légales régissant l'accessibilité aux données de la recherche (Open Data) et des sciences ouvertes (Open Science), les principes FAIR, formalisés par Wilkinson et al. (2016), visent à améliorer la gestion et la réutilisation des données scientifiques. Ils constituent aujourd'hui une norme internationale dans les politiques de science ouverte. Selon Mons et al. (2017), ces principes ne sont pas des règles rigides mais des orientations destinées à guider les chercheurs dans la mise en accessibilité raisonnée de leurs données.

Pour développer l'interopérabilité des dispositifs de données de recherche, un comité international a rédigé un catalogue de principes : les principes FAIR. Regroupés en dix principes (Figure 3), ils ont pour objectif « d'améliorer les infrastructures qui soutiennent la réutilisation des données de recherche, en mettant l'accent sur l'amélioration de la capacité des machines à trouver et à utiliser automatiquement les données, en complément à la réutilisation des chercheurs ». Ainsi, ces principes ne sont pas uniquement orientés par et pour les chercheurs (Rebouillat et Schöpfel 2023).

Findability	Accessibility	Interoperability	Reusability
• F1 : Les (méta)données se voient attribuer un identifiant unique et pérenne au niveau mondial. • F2 : Les (méta)données sont décrites avec des métadonnées riches. • F3 : Les métadonnées incluent clairement et explicitement l'identifiant des données qu'elles décrivent. • F4 : Les (méta)données sont enregistrées ou indexées dans un dispositif permettant de les rechercher.	• A1 : Les (méta)données peuvent être récupérées par leur identifiant en utilisant un protocole de communication standardisé. • A2 : Les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles.	• I1 : Les (méta)données utilisent un langage formel; accessible partagé et largement applicable pour la présentation des connaissances. • I2 : Les (méta)données utilisent des vocabulaires qui suivent les principes FAIR.	• R1 : Les (méta)données sont richement décrites avec une pluralité d'attributs précis et pertinents.

Figure 3 : Les principes FAIR.

Les principes FAIR se déclinent comme suit : Findable (F) : les données doivent être repérables via des identifiants persistants et des métadonnées descriptives ; Accessible (A) : elles doivent être consultables via des protocoles ouverts, même si certaines données sensibles nécessitent des restrictions ; Interoperable (I) : elles doivent être lisibles par des machines, avec des formats standards ; Reusable (R) : elles doivent être bien documentées et associées à des licences claires permettant leur réutilisation.

L'application de ces principes en océanographie pose plusieurs défis concrets : hétérogénéité des formats, manque de ressources humaines pour rédiger des métadonnées complètes, infrastructures non interoperables, etc. Pourtant, dans les disciplines marines, où les données peuvent être coûteuses à collecter et difficiles à reproduire, l'enjeu de leur réutilisation est particulièrement stratégique.

Comme rappelé par Rebouillat et Schöpfel (2023), les principes FAIR ne visent pas à assurer la qualité des données de recherche, ni à contribuer à l'interopérabilité légale des dispositifs et de leurs utilisations.

II. CADRES DE REFERENCE ET OBSTACLES AU PARTAGE DES DONNEES DE RECHERCHE

1. ENVIRONNEMENT INSTITUTIONNEL ET JURIDIQUE DE LA SCIENCE OUVERTE

Dans les sciences de l'information et de la communication, la donnée n'est pas considérée comme une évidence, mais comme un objet construit. Denis et Goëta (2013, 2016) montrent que les « données brutes » n'existent pas en soi : elles résultent d'opérations de sélection, de formatage et de « brutification » qui visent à les présenter comme naturelles et immédiatement exploitables. Goëta (2016) souligne par ailleurs que produire des données, c'est aussi instaurer des publics, c'est-à-dire créer les conditions sociales et techniques de leur usage. Autrement dit, les données ne sont jamais neutres : elles sont le produit de pratiques situées. Dans le cadre de ce mémoire, ces débats théoriques ne seront pas décrits en détail. Pour autant, il semble important de rappeler que les données océanographiques abordées ici ne sont pas de simples traces objectives de l'océan : elles sont le résultat d'un travail scientifique, organisationnel et technique qui conditionne leur statut et leur circulation.

Dans ce travail, nous considérerons les données de recherche selon la définition d'Agnès Robin, maîtresse de conférences en droit privé et sciences criminelles et dont les travaux portent, entre autres, sur le droit de la recherche et de l'innovation. Ainsi, dans son ouvrage *Droit des données de la recherche*, Agnès Robin (2022) définit les données de la recherche comme l'ensemble des informations recueillies, produites ou transformées dans le cadre d'une activité scientifique. Il s'agit ainsi des éléments concrets (données brutes, résultats d'expériences, observations, etc.) qui constituent le fondement théorique et empirique à la base de la construction des connaissances scientifiques. Ces données, recueillies ou générées au cours de projets scientifiques, ne sont pas régies par un cadre légal et réglementaire unique dans le droit européen ou français. Cette définition coïncide avec celle du CNRS⁸ (2017) qui regroupe les expressions « données de la recherche », « données scientifiques » et « données issues de la recherche » couvrant une réalité commune.

1. Orientation des acteurs institutionnels

Le cadre juridique entourant les données de recherche oscille entre plusieurs logiques parfois contradictoires : celle de la science ouverte, qui prône l'accessibilité et la transparence des connaissances (Robin, 2022), et celle d'un droit universel à la connaissance, qui pose la question de la propriété des données scientifiques et de leur exploitation commerciale (Blanquart et al., 2017). De plus, Blanquart et al., (2017) rappellent que la modalité privilégiée de régulation est le droit souple, centrée autour de

⁸ Centre National de la Recherche Scientifique.

la publication de livres blancs ou de recommandations non contraignantes. Ces dernières arrivent souvent *a posteriori* et encadrent des pratiques qui ont déjà émergé et se sont largement répandues hors réglementation.

Cette ambiguïté se retrouve dans la définition même des acteurs impliqués : tandis que certains chercheurs sont considérés comme auteurs (à l'origine d'une avancée scientifique), d'autres sont davantage perçus comme valorisateurs des données (cherchant à rendre utilisables ou commercialisables les résultats de la recherche) (Collin, 2017). Ce flou conceptuel influence directement la manière dont les scientifiques s'approprient et/ou résistent à ces injonctions.

Par ailleurs, la *Loi pour une République numérique* (CNRS, 2017) a introduit le principe de libre réutilisation des données de la recherche. Cependant, son application demeure floue, notamment en raison d'une absence de régulation claire sur les droits des chercheurs sur leurs propres données brutes (Blanquart et al., 2017). En pratique, beaucoup de scientifiques ignorent les cadres juridiques régissant les données qu'ils produisent ou utilisent, ce qui accentue les disparités entre disciplines et institutions.

Enfin, la complexité du statut juridique des données ajoute une difficulté supplémentaire. La structure de base des données est protégée par le droit d'auteur, le contenu de base relève du droit *sui generis* sur les bases de données et les éléments constitutifs du contenu peuvent être soumis à des protections spécifiques (Blanquart et al., 2017).

Ces différentes couches juridiques rendent difficile l'adoption d'un cadre harmonisé pour la gestion des données scientifiques. Ainsi, bien que la science ouverte soit un objectif largement promu, sa mise en œuvre est freinée par des enjeux de gouvernance, de protection des droits et d'appropriation⁹ des régulations par les chercheurs eux-mêmes.

2. Règles, obligations et zones grises du cadre juridique

Le partage des données scientifiques s'inscrit dans un cadre institutionnel qui oscille entre incitation à l'ouverture et contraintes structurelles. L'initiative de la « Décennie des sciences océaniques pour le développement durable » des Nations unies (Pendleton et al., 2023) illustre cette volonté de promouvoir une approche collaborative et accessible des connaissances marines. De même, en France, le *Deuxième Plan National pour la Science Ouverte* (Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, 2021) cherche à renforcer l'accessibilité des données en instaurant des

⁹ La notion d'appropriation désigne le processus par lequel des individus ou des collectifs s'emparent d'outils, de normes ou de dispositifs, en les adaptant, les réinterprétant et parfois en les transformant. Elle ne se réduit pas à l'usage, mais suppose une activité créative et située. De Certeau (1980) la décrit comme un « art de faire » à travers lequel les acteurs détournent et réinventent les prescriptions qui leur sont imposées. Dans le champ des technologies et de l'innovation, Akrich (1998) a montré que l'appropriation implique des opérations de traduction et de reconfiguration des objets techniques par leurs usagers. Dans la continuité, Jouët (2000) souligne qu'elle suppose un travail d'adaptation des dispositifs aux pratiques et représentations des individus, alors que Proulx (2005) propose de l'envisager comme un processus à la fois cognitif (comprendre), pratique (savoir utiliser) et symbolique (donner du sens).

Dans le cadre de ce mémoire, la notion d'appropriation est entendue dans un sens large : il ne s'agit pas uniquement d'évaluer une conformité stricte aux prescriptions institutionnelles, mais de saisir la variété des manières dont les chercheurs en océanographie s'ajustent aux injonctions de science ouverte qu'il s'agisse de réinterprétation, de résistance ou de contournement.

obligations de diffusion pour les recherches financées sur fonds publics et en mettant en place des infrastructures dédiées comme Recherche Data Gouv.

Ce plan repose sur trois mesures phares : rendre obligatoire la diffusion des données issues de financements publics, créer une plateforme nationale fédérée pour l'hébergement des données et promouvoir leur gestion selon les principes FAIR.

Toutefois, cette injonction à l'ouverture se heurte parfois à des cadres législatifs et à des habitudes institutionnelles plus anciennes. Une analyse systémique de la *Loi pour une République numérique* (CNRS - Direction de l'Information Scientifique et Technique 2016) met en évidence un paradoxe : bien que la législation encourage l'ouverture des données, son application reste complexe en raison d'un manque d'accompagnement concret pour les chercheurs. Le *Livre blanc sur la science ouverte* (CNRS - Direction de l'Information Scientifique et Technique 2016) révèle que plus de la moitié des chercheurs interrogés ignorent si les données qu'ils utilisent sont libres de droits, ou les exploitent malgré les restrictions. Ce constat souligne un besoin de formation et de sensibilisation aux enjeux juridiques et techniques du partage des données.

Par ailleurs, la transition numérique transforme profondément les pratiques de recherche. Le rapport du CNRS (2017) sur la science et le numérique montre que la multiplication des outils technologiques (capteurs, simulations, instruments de mesure numériques) a conduit à une explosion du volume de données disponibles. Cette évolution nécessite non seulement des infrastructures adaptées pour stocker et diffuser ces données, mais aussi une adaptation des pratiques scientifiques et des politiques institutionnelles.

2. LES PRATIQUES ET TENSIONS AU CŒUR DU TRAVAIL SCIENTIFIQUE

1. Entre ouverture et rétention : une lecture socio-anthropologique

Le partage des données ne se limite pas à des considérations techniques ou légales, mais reflète également des dynamiques sociales. Dans les laboratoires, les rapports aux données varient selon les communautés scientifiques et les rôles professionnels. Millerand (2012) met en lumière l'émergence d'une nouvelle catégorie de travailleurs invisibilisés, dédiés à la gestion des données, mais rarement reconnus comme des contributeurs à part entière, malgré leur rôle clé dans la circulation et l'interopérabilité des données (Millerand, 2012). Cette mise en invisibilité, qualifiée de processus actif de dévalorisation, soulève des interrogations sur la reconnaissance institutionnelle de ces acteurs et sur leur impact sur les dynamiques de recherche.

L'essor des grandes bases de données scientifiques, en juxtaposant des disciplines auparavant isolées, modifie également les pratiques de travail et les attentes envers ces nouveaux intermédiaires de la recherche (Millerand 2012). Ces derniers, souvent assimilés à des « techniciens invisibles » se retrouvent au cœur des infrastructures de

stockage et de partage des données, sans pour autant bénéficier d'une pleine reconnaissance institutionnelle (Millerand 2012). Cette tension illustre la dichotomie entre le travail de gestion des données et celui des chercheurs, dont le principal objectif reste la production de connaissances scientifiques (Millerand et Baker 2011).

Dans cette dynamique, les rapports aux données restent hétérogènes. Millerand (2011) illustre comment des disciplines et des communautés de recherche peuvent adopter des postures très différentes face au stockage et au partage des données, renforçant l'idée d'une appropriation différenciée des injonctions institutionnelles.

Comme le soulignent Millerand et Baker (2011), les scientifiques ne considèrent pas toujours la gestion des données comme une priorité, celle-ci étant reléguée à une tâche de support plutôt qu'à une fonction centrale de l'activité scientifique. L'exemple de l'initiative Ocean Informatics (OI) menée à la Scripps Institution of Oceanography entre 2002 et 2006 présenté dans leurs travaux met en lumière la complexité de ces transformations. OI est pensée comme une infrastructure informationnelle permettant à la fois de structurer la gestion des données scientifiques et de favoriser l'interopérabilité entre disciplines (Millerand et Baker 2011). Cependant, l'intégration de ces nouvelles logiques informationnelles dans la recherche océanographique n'est pas sans poser de défis : les pratiques scientifiques doivent s'adapter à des processus de normalisation et de mise en base de données souvent perçus comme contraignants (Jaton et Vinck, 2016).

Par ailleurs, le processus même de mise en base de données peut être source de frictions comme montré par Jaton et Vinck (2016). Loin d'être neutre, cette étape implique des choix stratégiques, influencés par des critères institutionnels ou personnels. Collin (2017) rappelle à ce titre que la définition de l'auteur et du valorisateur reste floue, et que les contributions des différents acteurs impliqués dans la production des données ne sont pas toujours reconnues de manière équitable. Ce manque de clarté peut exacerber les stratégies de rétention des données, notamment lorsque celles-ci sont perçues comme des ressources stratégiques pour la carrière des chercheurs. De plus, comme le montre Anichini (2017), les chercheurs peuvent être tentés d'écarter certains résultats jugés insuffisants ou non valorisants pour leur carrière, contribuant ainsi à une fabrique de l'ignorance où certaines informations, bien que scientifiquement pertinentes, ne sont pas diffusées. Cette tension entre ouverture et contrôle des données met en évidence des enjeux stratégiques liés à la publication et à la reconnaissance académique, en particulier dans un contexte où la science ouverte devient un critère d'évaluation institutionnelle.

Enfin, les rapports différenciés aux données selon les disciplines et les institutions renforcent l'idée d'une appropriation inégale des pratiques de stockage et de partage (Millerand, 2011). Certains domaines scientifiques, plus habitués à la gestion collaborative des données, développent des infrastructures adaptées pour faciliter l'interopérabilité, tandis que d'autres peinent à intégrer ces nouvelles exigences dans leur mode de fonctionnement. Cette diversité illustre la nécessité d'une réflexion plus large sur l'accompagnement des chercheurs dans cette transition vers une science plus ouverte et sur la place des gestionnaires de données dans ce processus.

Ainsi, la diversité des rapports aux données (Millerand, 2011) reflète une appropriation différenciée des normes institutionnelles et des injonctions à l'ouverture, oscillant entre adhésion, contournement et résistance

2. Science ouverte et carrière académique : concilier injonctions et intérêts individuels

Si les politiques de science ouverte encouragent le partage des données dans un contexte de recherche scientifique au 21^{ème} siècle plus riche en données et plus collaborative que par le passé, elles entrent parfois en contradiction avec les logiques de compétition académique. D'un côté, les politiques de science ouverte favorisent le partage des données en mettant en avant des principes d'accessibilité et de transparence. De l'autre, Zhu et al. (2024) mettent en évidence comment les chercheurs naviguent entre la nécessité de partager leurs résultats et la pression de publier pour progresser dans leur carrière. Ce paradoxe est particulièrement marqué dans des disciplines où les données sont le principal capital scientifique et peuvent devenir des leviers de différenciation et de pouvoir.

Ces tensions se retrouvent à l'échelle internationale. Les études de Tenopir et al. (2011, 2020) basées sur un questionnaire auprès de plus de 1300 scientifiques à des stades de carrière et dans des champs disciplinaires variés montrent que les pratiques de partage, stockage et réutilisation des données varient considérablement selon les contextes disciplinaires et nationaux. Certains États ont mis en place des politiques incitatives, voire obligatoires, de partage des données, particulièrement lorsque la recherche est financée par des fonds publics. En Europe et aux États-Unis, par exemple, des organismes comme la National Science Foundation (NSF) ou la Commission européenne ont instauré des obligations en matière de gestion des données, incluant l'élaboration de plans de gestion de données pour garantir leur accessibilité à long terme. Ainsi, les répondants ont indiqué qu'un financement public était la raison la plus importante et incitative de préserver et stocker correctement les données. En effet, si la recherche est issue de fonds publics, les résultats devraient, selon eux, devenir propriété publique et ainsi être préservés correctement. Cependant, ces politiques rencontrent des résistances de la part des chercheurs, notamment en raison du manque de reconnaissance institutionnelle du travail lié à la gestion et au partage des données.

Cette ambivalence se manifeste également dans la temporalité du partage. Comme le montrent Tenopir et al. (2011), de nombreux chercheurs hésitent à diffuser leurs données avant la publication de leurs résultats, craignant de perdre un avantage concurrentiel ou de compromettre leurs futures publications. Si une majorité reconnaît l'importance du partage après publication, la situation devient plus complexe lorsqu'il s'agit de collaborer avec l'industrie ou avec des chercheurs extérieurs à leur propre réseau académique. Ce phénomène contribue à façonner des dynamiques de rétention sélective, où certaines données sont stratégiquement conservées et d'autres mises à disposition selon des logiques d'intérêt.

Ainsi, la transition d'un modèle fondé sur la possession des données à un modèle basé sur leur mise en commun se heurte à des résistances structurelles et culturelles. Les

conclusions de Zhu et al. (2024) vont dans le sens de celles de Tenopir et al. (2011, 2020) à savoir que les tensions entre l'intérêt personnel et l'intérêt public ont toujours été une question centrale dans le processus de partage des ressources et de circulation de l'information. Le partage des données n'est pas seulement une question technique ou éthique, mais un enjeu profondément lié aux modes d'évaluation de la recherche et aux incitations académiques. Si la reconnaissance institutionnelle du travail lié à la gestion des données reste marginale et que les modèles d'évaluation de la recherche privilégient la production d'articles plutôt que la valorisation des données elles-mêmes, ces tensions ont de fortes probabilités de persister.

III. METHODOLOGIE ET DEMARCHE D'ENQUETE

1. UNE APPROCHE QUALITATIVE FONDEE SUR UN CORPUS MIXTE

Afin de mener à bien la recherche autour de ce questionnaire, une approche avec un corpus mixte a été choisie. D'une part, un corpus textuel constitué des textes institutionnels et juridiques de référence. D'autre part, un corpus d'entretiens semi-directifs avec des professionnels issus de différents laboratoires de recherche en océanographie.

Initialement, le corpus textuel avait pour objectif d'être constitué de documents au plus près du quotidien des chercheurs comme des guides déclinés par les institutions, des vade-mecum, des posters ou d'autres supports accessibles et orientés pour les personnels des laboratoires de recherche océanographiques. Cependant, suite aux entretiens et à la faible connaissance des personnes interrogées de ces documents au sein de leurs institutions, le choix a été fait de réorienter le corpus textuel sur les documents fondateurs autour du droit des données de la recherche, des pratiques de science ouverte à l'attention des chercheurs, des gestionnaires des données. Ces documents prennent plutôt la forme d'ouvrage de référence, de livres blancs. L'objectif de ce corpus textuel est d'étudier le discours institutionnel afin de définir les éléments saillants pour mieux analyser les réponses issues des entretiens et de mettre en évidence les éventuelles ambiguïtés de ce discours.

Par ailleurs, le choix a été fait de privilégier quelques entretiens ciblés sur des gestionnaires de données / professionnels de recherche produisant des données. Ce choix par rapport à celui d'un questionnaire large et moins ciblé est afin d'obtenir des réponses moins nombreuses mais plus qualitatives. Les personnes répondant aux questionnaires sont souvent biaisées et la profondeur des réponses apportée reste souvent assez superficielle par rapport à ce qui pourrait être obtenu via des entretiens. L'objectif de ces entretiens est donc de comprendre les pratiques et usages de ces professionnels de la recherche autour de leur gestion des données, ainsi que leur éventuelle connaissance et/ou appropriation des discours institutionnels ayant trait à ce sujet.

1. Le corpus textuel : sources et critères de sélection

Dans le cadre de ce mémoire, un corpus institutionnel ciblé a été constitué afin d'éclairer le contexte normatif dans lequel s'inscrit la gestion et le partage des données de recherche. Ce corpus comprend cinq documents produits entre 2019 et 2022 par deux acteurs majeurs de la science ouverte en France : le CNRS (Centre national de la recherche scientifique) et le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI). L'ensemble de ces textes constitue le socle du discours structurant les politiques actuelles autour des données scientifiques. Le choix a été fait de sélectionner des documents issus de ces 2 acteurs en raison de leur statut « reconnu » par les employés du monde de la recherche océanographique indépendamment de leurs

tutelles propres. De plus, au vu du caractère mouvant et complexe des thématiques d'ouverture des données de recherche, ces documents bien que couvrant une période temporelle restreinte montrent des évolutions dans le discours et les recommandations.

L'objectif de ce corpus est double : d'une part, il permet une lecture analytique des injonctions institutionnelles en matière de science ouverte et plus particulièrement sur les données de recherche ; d'autre part, il constitue une base de réflexion critique pour interroger leur réception et leur mise en œuvre concrète par les chercheurs, gestionnaires de données et autres acteurs impliqués au sein des laboratoires océanographiques. L'étude de ces textes vise notamment à repérer les éventuelles ambiguïtés, injonctions contradictoires ou flous conceptuels susceptibles d'influencer les pratiques quotidiennes et les appropriations par les professionnels de la recherche.

Feuille de route du CNRS pour la science ouverte (CNRS, 2019)

Datée du 18 novembre 2019, la *Feuille de route du CNRS pour la science ouverte* constitue un tournant décisif dans la politique de l'organisme. Premier document du corpus, elle affirme l'engagement du CNRS dans une transformation structurelle de la recherche, fondée sur l'ouverture des publications scientifiques, des codes sources, et surtout, des données de recherche.

Elle introduit une stratégie nationale articulée autour de sept axes prioritaires : les publications, les données de la recherche, la fouille et l'analyse des textes, l'évaluation des chercheurs, la refondation de l'information scientifique et technique (IST), la formation et les compétences, et le positionnement international. Cette stratégie fixe des objectifs ambitieux : atteindre 100 % de publications en accès ouvert, promouvoir le partage des données selon les principes FAIR, soutenir les infrastructures d'analyse, repenser l'évaluation scientifique et renforcer la formation aux pratiques de science ouverte.

Au-delà de ces priorités, le document affirme les grands principes de cette transition (accessibilité, transparence, reproductibilité) et souligne le rôle central des données dans la production scientifique contemporaine. Il insiste sur la nécessité d'un accompagnement à grande échelle, à travers la mise en place d'un réseau de soutien à la gestion des données et la désignation de référents dans les laboratoires.

La feuille de route introduit également une dimension éthique et politique, en appelant à une souveraineté scientifique sur les infrastructures de dépôt et de diffusion, dans un contexte de dépendance croissante aux plateformes commerciales. Elle se positionne ainsi comme un texte fondateur, porteur d'un horizon de transformation à la fois organisationnelle et culturelle.

Plan Données de la Recherche du CNRS (CNRS, 2020)

Le *Plan Données de la Recherche* du CNRS, publié en 2020, prolonge la feuille de route de 2019 en traduisant ses orientations stratégiques en actions concrètes. Il vise à structurer une politique cohérente, ambitieuse et transversale autour des données de la

recherche, en cohérence avec les principes de la science ouverte et les principes FAIR. Ce texte marque le passage d'un discours stratégique à un cadrage opérationnel.

Le document insiste sur la nécessité de faire émerger une véritable culture de la donnée au sein du CNRS. Pour cela, il propose un plan d'actions fondé sur une gouvernance dédiée : la Direction des Données Ouvertes de la Recherche (DDOR) et sur une articulation fine entre outils, infrastructures, standards et pratiques. Il aborde la diversité des données scientifiques, les étapes de leur cycle de vie, ainsi que les modalités de stockage, d'archivage, de partage et les rôles des différents acteurs, tant scientifiques que techniques.

Le plan identifie plusieurs axes prioritaires que sont le développement de compétences internes par la formation ; la création d'une documentation adaptée aux profils variés de chercheurs ; la structuration d'un réseau d'appui à la gestion des données ; et l'intégration systématique des principes FAIR dans l'ensemble du cycle de vie des données.

Ce texte est particulièrement utile pour comprendre comment les institutions entendent opérer un changement de culture. Il privilégie la régulation douce, l'accompagnement pédagogique et l'incitation plutôt que des contraintes strictement juridiques, tout en précisant les obligations émergentes auxquelles les chercheurs devront progressivement se conformer.

Deuxième Plan national pour la science ouverte (MESRI, 2021)

Le *Deuxième Plan national pour la science ouverte* (2021–2024) étend le périmètre stratégique de la science ouverte en France. Dans la continuité du premier plan de 2018, il en renforce les exigences, notamment en matière de diffusion des données issues de la recherche publique. L'ouverture des données devient désormais la norme, sauf exception dûment justifiée (secret industriel, données sensibles, etc.).

Ce plan vise à généraliser la science ouverte à l'échelle nationale à travers quatre axes stratégiques que sont l'accès ouvert aux publications ; l'ouverture des données de la recherche ; la diffusion des codes sources produits par la recherche publique ; et la transformation des pratiques d'évaluation scientifique.

Il s'appuie sur des obligations légales renforcées, telles que la *Loi de programmation de la recherche* (2020), qui fixe l'objectif de 100 % de publications en accès ouvert d'ici 2030. Dans ce cadre, plusieurs mesures structurantes sont mises en place, parmi lesquelles la mise en œuvre de l'obligation de diffusion des données issues de financements publics ; la création de Recherche Data Gouv, une infrastructure nationale, fédérée et pluridisciplinaire pour le partage des données ; la promotion de politiques de gestion des données fondées sur les principes FAIR.

Le plan privilégie un modèle d'édition scientifique « diamant » (sans frais pour les auteurs), interdit le financement des revues hybrides, et promeut une évaluation scientifique fondée sur des critères qualitatifs, en rupture avec la logique des indicateurs bibliométriques.

Enfin, il positionne clairement l'ouverture des données comme un enjeu non seulement scientifique, mais aussi démocratique et stratégique, en lien avec la transparence, la collaboration internationale et la souveraineté numérique. Il réaffirme ainsi le rôle actif de la France dans la construction de l'Espace européen de la recherche.

Guide pour les chercheurs : Partager les données liées aux publications scientifiques (MESRI, 2022)

Le *Guide pour les chercheurs : Partager les données liées aux publications scientifiques* (MESRI, 2022) est un document technique et pragmatique conçu pour accompagner directement les chercheurs dans le partage de leurs données de recherche. Il met l'accent sur l'importance de l'ouverture des données pour renforcer la transparence, la reproductibilité et la réutilisation des résultats scientifiques.

Destiné à un public de praticiens, ce guide propose un parcours clair et structuré, articulé autour de plusieurs étapes essentielles que sont la rédaction d'un plan de gestion des données (Data Management Plan, DMP) ; le choix d'un entrepôt de dépôt, de préférence institutionnel ou disciplinaire plutôt que commercial ; l'usage de formats ouverts et de standards reconnus dans chaque discipline ; et la citation correcte des jeux de données, via l'attribution d'identifiants pérennes (DOI) permettant de lier publication et données.

Le guide prend en compte la diversité des pratiques selon les disciplines, tout en promouvant une harmonisation progressive autour des principes FAIR. Il reflète ainsi un double objectif des politiques publiques : normaliser les pratiques pour améliorer la qualité et l'interopérabilité des données, tout en respectant les spécificités scientifiques et les rythmes d'appropriation.

Passeport pour la science ouverte (Ouvrir la science, 2024)

Enfin, le *Passeport pour la science ouverte* (Ouvrir la science, 2024) est un guide ministériel à visée pédagogique, principalement destiné aux jeunes chercheurs et doctorants. Synthétique et accessible, il vise à les familiariser dès le début de leur parcours avec les principes, les obligations et les bonnes pratiques de la science ouverte.

Le guide propose une entrée vulgarisée dans les principaux enjeux liés à la gestion et à la diffusion des données de la recherche. Il aborde notamment la distinction entre données brutes et données traitées ; l'importance de la documentation des données à l'aide de métadonnées ; la temporalité du partage (quand et comment diffuser les données) ; et les droits d'usage, de réutilisation et les éventuelles restrictions (données personnelles, secret industriel, etc.). Il s'appuie sur le cadre juridique français, en rappelant le principe selon lequel les données doivent être « aussi ouvertes que possible et aussi fermées que nécessaire », conformément au *Code de la recherche* et à la *Loi pour une République numérique*.

Le guide détaille les étapes du cycle de vie des données, promeut les principes FAIR et recommande l'utilisation d'entrepôts de confiance pour leur diffusion. Il met également en lumière les enjeux émergents, tels que la sobriété numérique, et propose un écosystème

d'accompagnement institutionnel pour aider les chercheurs à s'approprier ces pratiques. En mettant l'accent sur la construction d'une culture commune de la science ouverte, ce document vise à inscrire ces pratiques dans les usages professionnels dès la formation doctorale.

2. Le corpus d'entretiens : choix des enquêtés et grille d'entretien

En complément du corpus textuel institutionnel, un second corpus a été constitué à partir d'entretiens semi-directifs menés avec des professionnels impliqués dans la recherche océanographique en France. L'objectif de ces entretiens est de documenter les perceptions, les pratiques et les éventuelles résistances des acteurs de terrain face aux injonctions institutionnelles en matière de gestion, de partage et de valorisation des données de recherche.

L'étude s'inscrit dans une perspective qualitative, et a pour ambition de répondre à la problématique centrale du mémoire rappelée ici : « Est-ce que, et de quelle manière, les acteurs des laboratoires de recherche océanographique en France s'approprient les injonctions institutionnelles de stockage, de partage et de gestion des données, et quelles en sont les éventuelles implications sur l'organisation du travail de recherche ? »

Le choix d'entretiens semi-directifs a été fait afin de permettre une certaine liberté dans l'expression des personnes interrogées, tout en garantissant d'aborder les mêmes thématiques au cours des différents entretiens. Les sujets étaient l'introduction et le parcours, les données en océanographie : définitions et pratiques, les infrastructures et outils de stockage / partage, la science ouverte et les principes FAIR, les cadres juridiques et institutionnels, les freins, tensions et arbitrages, les dimensions sociales et culturelles, les perspectives et évolutions, pour finir avec une clôture. Le détail des questions abordées est présenté en [annexe 1](#). Ces entretiens permettent de mettre en évidence les pratiques effectives des professionnels interrogés, des résistances, des craintes, mais aussi de la connaissance ou méconnaissance des injonctions institutionnelles autour de la science ouverte des données de recherche.

Le corpus est constitué de 4 entretiens réalisés entre mai 2025 et juillet 2025, d'une durée moyenne de 45 minutes à 1 heure. Ils ont été enregistrés avec l'autorisation des participants pour permettre une écoute plus active lors de l'entretien en complément de la prise de notes initiale. Les participants ont également pu revenir sur des propos qu'ils avaient pu énoncer dans les entretiens afin de retirer tout élément qu'ils pouvaient considérer comme préjudiciables. Ces entretiens ont suivi une trame commune, tout en s'adaptant aux réponses de chaque professionnel interviewé pour fluidifier les échanges. Les verbatims issus des entretiens n'ont volontairement pas été reformulés pour respecter au plus proche les formulations des personnes interrogées, formulations parfois crûes retranscrivant ainsi de manière authentique leurs propos. Pour autant, les répétitions verbales simples (tics de langage ou hésitations) lorsque les interrogés réfléchissaient et s'exprimaient simultanément ont été retirées pour rendre la lecture plus fluide.

Les participants proviennent de laboratoires et instituts français spécialisés dans l'océanographie et l'étude des sciences marines. Les profils sont diversifiés : un

professeur, un chercheur, un enseignant chercheur et un gestionnaire de données. L'anonymat des répondants est rigoureusement respecté ; les noms des laboratoires ont été retirés, de même que nombreuses informations permettant une identification indirecte (noms de projets, de structures, genre des participants mis au masculin pour tous, etc.).

Le nombre d'entretiens est restreint. En croisant les entretiens, l'analyse permettra de mettre en évidence des tensions communes, mais aussi des variations selon les disciplines, les générations, ou les postes occupés. Ainsi, l'analyse qui est faite des résultats de ces entretiens n'a pas pour but de généraliser à outrance mais au contraire de faire des constats au sein d'un échantillon de personnes. Certains éléments retrouvés chez l'ensemble des professionnels ne sont peut-être pas représentatifs mais probablement symptomatiques de quelque chose de plus large. L'objectif est de mettre en lumière différentes notions telles que les écarts entre les discours institutionnels et les pratiques réelles, la diversité des logiques professionnelles mobilisées autour des données (valorisation, protection, partage), la manière dont les acteurs connaissent, adaptent ou contournent les normes prescrites ; et les résistances, freins identifiés, mais aussi les initiatives positives individuelles ou locales.

Ce corpus présente pour autant certaines limites, notamment la taille restreinte de l'échantillon. Plus de 35 professionnels de différentes structures et à différents stades d'avancement de carrière ont été contactés, dont 11 ont apportés des réponses positives pour participer à des entretiens en visioconférence. Au final, seuls 4 entretiens ont eu lieu. En effet, de nombreuses annulations ont eu lieu face à la tenue de l'UNOC (Conférence des Nations Unies sur les Océans) à Nice en juin qui a énormément mobilisé la communauté océanographique française et de la charge de travail conséquente parfois mal envisagée par les personnes sollicitées. De plus, suite à des soucis de santé puis à un événement familial urgent, deux entretiens ont été déprogrammés de mon fait pour lesquels nous n'avons jamais réussi à retrouver un créneau d'entretien.

2. METHODES D'ANALYSE : DE LA GRILLE D'ENTRETIEN A LA CONFRONTATION AVEC LA LITTERATURE

L'approche adoptée pour l'analyse des entretiens est volontairement pragmatique et orientée vers la compréhension des pratiques et perceptions des acteurs interrogés. Dans un premier temps, chaque entretien a fait l'objet d'une lecture intégrale de sa transcription, complétée par une réécoute attentive de l'enregistrement vidéo. Cette double démarche permet non seulement de vérifier l'exactitude et corriger les transcriptions, mais aussi de saisir les nuances qui peuvent se perdre à l'écrit : hésitations, silences ou changements de ton, autant d'éléments susceptibles de révéler une émotion sous-jacente.

La première grille d'analyse mobilisée correspond directement à la grille d'entretien. Celle-ci constitue un cadre initial structuré autour des grands axes thématiques abordés : définitions et pratiques liées aux données, infrastructures et outils de stockage/partage, principes FAIR, cadres juridiques et institutionnels, freins et tensions, ainsi que perspectives et évolutions. L'objectif de cette étape est d'extraire et de classer les réponses dans les catégories correspondant aux questions posées, afin de faciliter leur comparaison ultérieure.

Dans un second temps, les réponses à une même question ont été mises en regard entre les différents participants. Cette confrontation permet d'identifier à la fois des convergences, témoignant de points de vue partagés ou de pratiques communes, et des divergences, qui peuvent révéler des spécificités disciplinaires, des différences générationnelles ou des positionnements personnels face aux injonctions institutionnelles. Ce travail de mise en parallèle favorise également la mise en évidence d'éléments inattendus, non pressentis en amont des entretiens mais qui apparaissent comme des thématiques saillantes au fil des discours, ainsi que des éléments attendus, et confirmés par les propos des personnes interrogées.

En parallèle, le corpus textuel a fait l'objet d'un traitement similaire, bien qu'adapté à sa nature. Chaque document a été lu intégralement et codé selon une grille d'analyse spécifique disponible en [annexe 2](#), permettant d'identifier les messages communs véhiculés par l'ensemble des textes, ainsi que les divergences de ton, de priorité ou de cadrage entre institutions ou entre périodes de publication. Cette analyse a également porté une attention particulière aux formulations d'injonctions (obligations, incitations, recommandations) et aux ambiguïtés possibles, afin de mieux comprendre les interprétations que l'on peut retrouver auprès des professionnels de la recherche.

Enfin, les résultats issus de ces deux parties (corpus d'entretiens et corpus textuel) ont été mis en perspective avec la littérature scientifique existante. Les éléments saillants issus des verbatims sont confrontés aux discours institutionnels et aux travaux académiques, permettant d'identifier des zones de recoupement, mais aussi des tensions ou des écarts. Ainsi, la mise en regard de tous ces éléments permet d'établir une synthèse dans le cadre de ce mémoire qui tend à prendre en compte à la fois le vécu des professionnels de la recherche océanographique et le cadre normatif dans lequel ils évoluent.

IV. ENTRE PRESCRIPTIONS ET REALITES : APPROPRIATION CONTRASTEE DES INJONCTIONS

1. CE QUE DISENT LES TEXTES : UN CADRAGE INSTITUTIONNEL DE LA GESTION DES DONNEES

Les différents textes du corpus se rejoignent sur des notions clés. Comme cela sera montré dans les parties suivantes, les personnes interrogées n'ayant qu'une connaissance plus que partielle des textes institutionnels, cette partie n'a pas pour objectif de synthétiser les recommandations issues de ces textes mais s'intéresse plutôt aux messages communs véhiculés ainsi qu'aux éventuelles divergences pouvant être à l'origine de difficultés d'appropriations par les professionnels de la recherche. La grille d'analyse des documents est disponible en [annexe 2](#).

Tous les documents convergent le principe directeur « aussi ouvert que possible, fermé autant que nécessaire » (ou avec la nuance « pas plus fermées que nécessaire »). L'ensemble des documents promeut les principes FAIR comme standard de référence pour la gestion des données de recherche. De plus, l'ensemble des documents reconnaissent l'importance et l'obligation croissante des DMP, notamment dans les financements européens et ANR¹⁰. Une convergence claire se dessine contre l'utilisation des entrepôts commerciaux d'éditeurs, au profit d'entrepôts institutionnels, généralistes ou disciplinaires. Par ailleurs, tous les documents insistent sur le besoin de développer une « culture de la donnée » au sein de la communauté scientifique.

Pour autant, des divergences émergent entre les documents sur la force des injonctions. En effet, les documents de 2019 et 2020 contiennent principalement des recommandations et incitations institutionnelles. Le document de 2021 introduit des obligations légales et le document de 2024 mentionne un ancrage dans le *Code de la recherche* avec des obligations légales strictes.

Le niveau de prescription diffère aussi avec un gradient d'exigence selon le document. Les documents CNRS semblent mettre l'accent sur l'accompagnement et l'incitation, les plans nationaux mentionnent plus une vision stratégique avec des obligations progressives, là où document de 2024 présente un cadre légal strict avec justifications obligatoires.

Les infrastructures sont présentées différemment dans les documents où semble ressortir une montée en puissance des infrastructures nationales, suivant une dynamique temporelle : les documents de 2019 se concentrent sur les réservoirs thématiques, le document de 2021 mentionne la création de Recherche Data Gouv et le guide de 2024 présente un écosystème d'accompagnement qui serait structuré avec un accompagnement personnalisé par des experts, des critères techniques précis pour le choix des entrepôts et

¹⁰ ANR : Agence Nationale de la Recherche.

une invitation à réfléchir à l'impact environnemental et à la sobriété numérique associés à ces enjeux.

La dimension temporelle des documents du corpus, mentionnée précédemment peut être résumé comme suit. Une première phase (2019-2020) axée autour de la sensibilisation et de la structuration avec le développement des bonnes pratiques, la mise en place des outils et infrastructures et une approche plutôt incitative. Une deuxième phase (2021) sur l'institutionnalisation présentant les obligations légales introduites, l'infrastructure nationale créée et la vision stratégique affirmée. Une troisième phase (2024) mettant en avant la maturité et l'opérationnalisation affichant un cadre légal stabilisé, les outils pratiques développés et la prise en compte d'enjeux plus transversaux (environnement, éthique).

Les tensions identifiées au sein de ce corpus peuvent être résumées autour de quatre grands axes. Premièrement, la nuance entre autonomie et contrainte. En effet, le corpus oscille entre incitations souples et obligations formelles, entre liberté académique et obligations réglementaires. Les chercheurs sont invités à partager leurs données (« obligation de diffusion » des données financées sur fonds publics), mais le discours ambivalent, mêlant accompagnement et prescription ce qui peut être source de confusion. Deuxièmement, le paradoxe entre diversité disciplinaire et standardisation. Les principes FAIR et les infrastructures recommandées supposent une homogénéisation des pratiques, difficilement compatible avec la diversité des disciplines et des types de données. Les documents reconnaissent cette diversité sans toujours proposer des solutions adaptées à la complexité des contextes scientifiques spécifiques. Troisièmement, la tension entre l'innovation et la « souveraineté » des outils. Le discours institutionnel appelle à l'usage d'outils performants, mais insiste également sur l'importance de la souveraineté numérique. Cette tension se retrouve par la coexistence d'outils privés largement adoptés comme Google Drive et de plateformes publiques comme Recherche Data Gouv. Dernièrement, le décalage entre les ambitions affichées et la réalité opérationnelle. En effet, le discours sur la transformation des habitudes repose sur des conditions rarement réunies dans les laboratoires (manque de formation, infrastructures incomplètes, absence de référents), ce qui renforce l'écart entre le prescriptif et la mise en œuvre.

2. CE QUE FONT LES ACTEURS : DES CONCEPTIONS SITUEES ET CONTEXTUALISEES DES DONNEES

Les quatre personnes interrogées occupent des fonctions variées dans le domaine océanographique. Le premier chercheur (Chercheur1) est professeur en écologie marine et directeur d'un OSU¹¹. Le deuxième chercheur (Chercheur2) est maître de conférences en biogéochimie marine. Le troisième chercheur (Chercheur3) est chercheur en océanographie physique, impliqué dans l'infrastructure de recherche Argo France. Chercheur3 a aussi été chef de mission d'une campagne océanographique et a donc « un point de vue situé par rapport à ça ». Enfin, le gestionnaire de données (Gestionnaire1) a un parcours orienté ingénierie de logiciel. Au cours des cinq dernières années, une part croissante de son activité a été « dédiée à la gestion des données ». Gestionnaire1 est en charge ou en cocharge de la gestion des données pour des réseaux d'observation nationaux par exemple, ou des projets du type PPR¹².

Concernant leur formation à la gestion des données, les réponses révèlent une absence généralisée de formation initiale spécifique. Chercheur1 témoigne : « on m'a jamais rien imposé, j'ai été progressivement sensibilisé parce que [...] je me suis préoccupé de l'observation [...] *de facto* faire de l'observation ça voulait dire rendre les données open, ça ne pouvait pas être autrement. J'y ai été sensibilisé de par les responsabilités que j'ai prises ». Gestionnaire1 explique pour sa part être « rentré par la porte avec ma casquette responsable de la base de données. Mais inévitablement, on a embrayé sur les discussions, sur quel standard utiliser, comment diffuser la donnée et comment l'enrichir en métadonnées. Et c'est par ce biais là que j'ai commencé à m'intéresser beaucoup plus au cycle de vie de la donnée ». Cette acquisition de compétences « sur le tas » est confirmée par Chercheur2 : « Non clairement, j'ai eu aucune formation, c'est on va dire formation sur le tas. Enfin si tant est qu'on puisse parler de formation, disons que y a des gens qui m'ont expliqué un peu comment ça marchait, les idées qu'il y avait derrière et ainsi de suite ». Chercheur3 avoue également ne pas avoir été formé ni s'y être particulièrement intéressé : « Non. J'avoue que peut-être qu'il y avait la possibilité, mais je me suis pas dit super j'y vais ouais. Mais ça nous intéresse peut-être pas trop non plus si je peux être honnête ».

Les données traitées varient selon les spécialités. Gestionnaire1 travaille sur des données biologiques et hydrologiques classiques. Chercheur2 effectue des mesures d'échantillons discrètes sur des rosettes. Chercheur3 se concentre sur des données de campagne (hydrologiques) et des données de flotteurs Argo. Enfin Chercheur1 est plus spécialisé sur des données biologiques marines.

L'utilisation des standards apparaît inégale selon les domaines. Chercheur3 mentionne : « NetCDF pour les flotteurs. [Les campagnes] NetCDF aussi en fait, ou

¹¹ OSU : Observatoire des Sciences de l'Univers. Structure fédérative regroupant laboratoires et services pour l'observation et l'étude de la Terre et de l'Univers.

¹² PPR : Programme Prioritaire de Recherche. Dispositif national de financement ciblé sur des thématiques scientifiques stratégiques.

éventuellement format Excel, CSV¹³, tableur ». Chercheur3 ajoute : « le labo, ça fait des années et des années qu'il travaille sur ce genre de données, donc finalement on a déjà développé des chaînes de traitement qui génèrent les données dans un format standard on va dire qui suit des protocoles internationaux ». Chercheur1 souligne l'importance de la standardisation : « c'est pour ça qu'on va vers des vocabulaires contrôlés, c'est le NOC¹⁴, c'est surtout les anglais hein qui ont développé du vocabulaire. [...] si tu standardises pas les choses, c'est inutilisable. Si chacun y va de sa propre unité, de sa propre façon de nommer les choses, là ça pose souci il est sûr que plus on va quitter la physique et plus on va aller vers la bio, plus c'est compliqué à standardiser ».

Gestionnaire1 détaille les standards utilisés dans le cadre d'un projet qu'il a monté : « on avait des outils de bancarisation et ensuite on a discuté avec les collègues chercheurs. [...] Pour les SNO¹⁵, ils ont des bases de données dans lesquelles ils insèrent au jour le jour [les données] pour lequel ils ont des gabarits et nous, une fois par an, on extrait ce qu'il faut des bases pour générer le jeu de données annuel. Et là on utilise des standards de métadonnées pour les décrire correctement et de manière complète. Ça s'est surtout fait dans le cadre [de ce projet], parce qu'on était parti un peu de zéro, [...] dans le groupe bancarisation, on a sensibilisé les acteurs au fait qu'il fallait qu'on utilise des standards qui existent. Alors il y a des standards qui sont naturels, hein ? Pour tout ce qui est description de la biodiversité, ben comme c'est du marin utilise WoRMS¹⁶, point barre. Et tout ce qui est physicochimique, on utilise plutôt les vocabulaires SeaDataNet ».

Cependant, l'application reste problématique selon Chercheur2 : « ce formalisme [des noms de variables], il existe [...] par contre on l'applique assez mal. » Pour autant, il indique essayer de « s'y tenir » mais remarque lors d'interactions avec des collègues qu'« il n'y a aucune rigueur là-dedans [...] ». Donc clairement, je pense que ça fait partie des points sur lesquels il faut qu'on fasse des progrès parce que je pense que ça entraîne, vraiment des données qui sont mal interprétées par la suite [...] ça peut avoir des conséquences assez fortes ».

La gestion des données suit généralement le processus des Plans de Gestion des Données (DMP). Chercheur1 explique : « Dans la plupart des projets, entre autres les projets ANR, il faut rédiger un data management plan au début, actualiser le data management plan à mi-parcours et puis faire un data management plan à la fin du projet. » Pour autant, il indique avoir parfois le sentiment qu'il s'agit d'une « obligation légale, c'est à dire, on regarde ce qu'on a bien fait. [...] On contrôle pas dans la pratique après si les choses elles sont réellement faites. C'est un premier pas, mais c'est pas suffisant je pense. Et puis moi, j'ai vu des collègues dans des projets râler parce qu'on leur demandait

¹³ Comma-Separated Values. Format de fichier texte pour données tabulaires (séparées par des virgules).

¹⁴ NOC : National Oceanography Centre. Centre de recherche britannique spécialisé en sciences et technologies marines, très impliqué dans l'établissement de vocabulaires contrôlés en océanographie.

¹⁵ SNO : Service National d'Observation. Dispositif labellisé pour collecter et suivre des données scientifiques à long terme dans un domaine spécifique.

¹⁶ WoRMS : World Register of Marine Species. Base de données mondiale recensant et validant les noms scientifiques des espèces marines.

un DMP, puis qui le faisaient bon gré, mal gré, puis [...] se préoccupaient pas forcément de la manière dont les données seraient bancarisées ».

Chercheur2 partage un point de vue similaire, même s'il ne s'est « jamais particulièrement investi dans ces parties là, mais j'ai toujours eu quand même le sentiment que c'était un peu une partie qui était traitée souvent un peu à la légère » dans le cadre des projets qu'il a pu voir. Pour autant, il note une amélioration ces dernières années avec « une prise de conscience de l'importance de ça ». Chercheur3 témoigne de son expérience : « Pour le DMP, j'ai eu à le faire dans un projet, mais comme c'était sur Argo en fait c'était déjà tout fait, simple ».

Gestionnaire1 pour sa part regrette les évolutions du DMP au niveau de l'ANR : « l'ANR demandait une mise à jour tous les 2 ans pour les projets qui dépassent les 5. Et malheureusement, ils ont changé leur fusil d'épaule et ils ne s'attendent plus qu'à une version finale en fin de projet. Et je dis malheureusement parce qu'effectivement, c'est très compliqué de mobiliser les scientifiques à remplir ce plan de gestion de données. Ils le voient effectivement comme une corvée. Alors que nous, on essaie de promouvoir l'idée que c'est un document vivant et c'est une trace de ce qui se fait en matière de gestion de données dans le projet et qui est super utile [...] Pour moi, c'est paradoxal pour des chercheurs qui ont l'habitude de tenir des cahiers de labo, de pas faire le clic en disant 'mais c'est pareil qu'un cahier de labo' ».

Gestionnaire1 donne un exemple concret pour lequel il a été sollicité pour accompagner le dépôt d'un jeu de données historiques bien précis « ils avaient plein de fichiers qui avaient été acquis au cours du temps, donc là on parle des données de 1952 à 1999 avant que ça intègre le SNO et que ce soit donc normalisé. [Une personne] m'a contacté pour qu'on regarde ensemble comment mettre en forme les données et comment et où les déposer, et ce que ça implique en termes de métadonnées. Donc là c'est un petit projet qu'on a fait sur 2-3 mois, et qui a abouti effectivement à ce que on dépose un joli fichier CSV sur l'entrepôt SEANOE avec toutes les bonnes métadonnées ».

Chercheur3 qui a été chef de mission océanographique poursuit avec son expérience personnelle : « Moi perso c'est compliqué. J'étais épuisé après les campagnes et quelque part, Oh, me replonger là-dedans. J'arrivais pas à m'intéresser à me motiver, à faire ça, [...] en plus, des fois il y a plein de jeux de données différents. [...] des fois on peut perdre un peu le fil de qu'est ce qui a été fait. À nouveau, ça demande aussi du temps et de l'investissement et quelque part, une fois que on a fait le travail de la campagne, qu'on a fait le travail de la valorisation et ben on passe à autre chose parce qu'on nous demande aussi de passer à autre chose et on a plus l'énergie, le temps, l'envie de se pencher sur ce travail-là, qui est un peu pénible. En fait, c'est le 'not so glamorous work' comme disent certains et qui est pourtant nécessaire ».

Ainsi, l'analyse des entretiens montre un ensemble de points communs marquants dans la manière dont les enquêtés conçoivent et manipulent les données en océanographie. Aucun d'entre eux n'a bénéficié d'une formation initiale spécifique à la gestion ou au partage de données ; l'apprentissage s'est fait « sur le tas », par échanges avec des collègues ou au fil de la prise de responsabilités impliquant la manipulation de données.

Tous travaillent sur des données liées à leurs spécialités, qu'il s'agisse de données biologiques, hydrologiques, biogéochimiques, échantillons discrets sur rosettes, ou données de flotteurs Argo, et utilisent des formats standards comme NetCDF, CSV ou Excel, parfois complétés de vocabulaires contrôlés (ex. WoRMS pour la biodiversité, SeaDataNet pour la physicochimie). Les plans de gestion de données, requis notamment dans les projets ANR, sont connus et produits, mais souvent perçus comme une formalité administrative dont l'application réelle reste peu contrôlée.

Des divergences apparaissent dans l'appropriation et l'application concrète des standards. Certains, comme Chercheur1 ou le gestionnaire de données, soulignent l'importance de la standardisation et veillent à l'appliquer rigoureusement, tandis que d'autres, comme Chercheur2, reconnaissent un manque de rigueur et pointent des difficultés d'harmonisation entre collègues. La perception du DMP reflète aussi cette variabilité : pour le gestionnaire, il s'agit d'un « document vivant » utile au suivi et à la traçabilité, alors que pour plusieurs chercheurs, il reste perçu comme une « corvée » ou une exigence formelle, rarement intégrée pleinement au processus scientifique. Enfin, la motivation et l'énergie à consacrer à la gestion post-campagne varient fortement : certains intègrent cette étape dans leur routine, d'autres peinent à s'y replonger après des semaines de travail intense en mer.

3. LES INFRASTRUCTURES AU MILIEU DES PRATIQUES DE GESTION

Les principaux entrepôts mentionnés sont SEANOE et SISMER¹⁷. Gestionnaire1 explique que l'entrepôt de référence pour les données est SEANOE, l'entrepôt opéré par Ifremer pour ODATIS¹⁸. Pour les données de biodiversité (données en occurrences taxons), l'objectif est qu'elles soient visibles dans le portail GBIF¹⁹ et/ou OBIS²⁰ au vu des passerelles entre les deux. Enfin, pour les données de séquençage génomique, comme l'indique Gestionnaire1 : « il n'y a pas à tortiller, c'est l'ENA²¹. En fait y a un consortium mondial avec un nœud européen, un nœud japonais, un nœud américain. [...] Et c'est incontournable parce que quand on publie un papier qui fait référence à une séquence, il faut qu'on donne un numéro d'accession qui est dans l'ENA ».

Chercheur3 décrit le processus pour les campagnes : « qui sont faites sur les navires de la flotte océanographique française [...] on acquiert les données, on les met dans un format qui nous semble le mieux, on l'envoie au SISMER et normalement il y a un échange avec le SISMER qui dit 'bon alors est ce que ça correspond bien à telle variable

¹⁷ SISMER : Systèmes d'informations scientifiques de la Mer. Service de gestion et de diffusion des données océanographiques.

¹⁸ ODATIS : Pôle de données et de services pour l'océan.

¹⁹ Global Biodiversity Information Facility. Infrastructure mondiale de données sur la biodiversité.

²⁰ OBIS : Ocean Biogeographic Information System. Base de données mondiales sur la répartition des espèces marines.

²¹ ENA : European Nucleotide Archive. Base de données européenne regroupant les séquences d'ADN et d'ARN produites par la recherche.

SeaDataNet, ainsi de suite' et donc ensuite elles sont archivées au niveau du SISMER. Et donc ça, ça nous permet d'avoir un DOI, d'avoir toutes ces choses-là donc ça c'est vraiment on va dire, voilà le schéma ultra classique. » Pour les données qui ne relèveraient pas directement du SISMER « y a cette plateforme qui s'appelle SEANOE qui permet de manière assez libre de déposer les données, c'est à dire y a pas de contraintes sur le format. Et donc, et ça permet d'attribuer un DOI à nos jeux de données. [...] soit ce qui rentre vraiment dans les prérogatives du SISMER et on suit exactement ce qu'ils disent, et ce qu'on arrive pas à mettre dans le SISMER, ça va chez SEANOE ». Chercheur3 ajoute une nuance : « Il y a un truc que je maîtrise pas très bien : il y a le SISMER et il y a SEANOE. Et c'est un petit peu nébuleux. Je sais que finalement, on les met à 2 endroits différents ».

Chercheur1 complète sur ces entrepôts et positionne la démarche française : « data.gouv.fr, il est valable que si t'as pas de choses dédiées c'est à dire c'est la poubelle où tu mets (enfin poubelle [...] y a rien de péjoratif dans mes propos hein) [...] nous on a un pôle de données océan qui s'appelle ODATIS. C'est une structure nationale, hein, publique donc y a aucune raison d'aller mettre des choses dans un fourre-tout. » Chercheur1 insiste sur le fait d'utiliser les entrepôts thématiques lorsqu'ils existent, ce qui est le cas dans les disciplines océanographiques et de biologie marine. Chercheur1 mentionne également SEANOE mais avec plus de réticence : « C'est pas parfait [...] Ils sont partis du principe que si tu veux récupérer beaucoup de données, il ne faut pas mettre beaucoup de contraintes. Du coup, ça marche bien. [...] Mais le fait qu'il y ait pas trop de contraintes, ça fait que parfois les métadonnées, elles, sont pas très bien renseignées et ou aussi elles sont renseignées mais pas selon un vocabulaire standardisé. ». Pour autant, il insiste sur la nécessité « qu'elles soient sous SEANOE plutôt que sur une machine, donc ça dépend si tu vois le verre à moitié vide ou à moitié plein. Mais en tout cas moi je pense que SEANOE, pour nous en France c'est mieux que tout un tas de plateformes tierces ».

Gestionnaire1 mentionne une autre plateforme qui est utilisée mais qu'il déconseille fortement : « Je connais évidemment ZENODO²² et je sais qu'il est utilisé par des personnes pour aller vite pour avoir un DOI là encore et c'est facile parce qu'il y a tellement peu de métadonnées à renseigner. » Il mentionne également les autres entrepôts disciplinaires qu'il a pu rencontrer notamment en SHS²³ « il y a des entrepôts qui sont construits, notamment l'entrepôt Nakala qui fait partie de l'infrastructure Huma-Num si je ne me trompe pas, qui a l'impression d'être un peu un fourre-tout aussi quand même »

L'importance des catalogues est soulignée par Gestionnaire1 : « Là, j'ai cité principalement des entrepôts. Mais on se sert aussi de catalogues de type Sextant²⁴. Alors Sextant au départ, c'est un outil Ifremer qui fait à la fois entrepôt et catalogue pour des données spatialisées. Et donc ce qui est intéressant pour nous, c'est que on peut décrire dans Sextant à l'aide d'une fiche de métadonnées, des ressources qui sont disponibles sur SEANOE et sur d'autres entrepôts, ce qui veut dire que un point d'entrée quand on cherche

²² ZENODO : Plateforme d'archivage sciences ouvertes (généraliste).

²³ SHS : Sciences Humaines et Sociales.

²⁴ Sextant : portail géographique donnant accès aux données marines et littorales.

quelque chose et qu'on sait pas dans quel dépôt il se trouve, quand c'est du marin, on va tenter de rechercher ça dans Sextant. » Ainsi, Sextant permet de construire des catalogues thématiques, notamment dans le cadre de projets spécifiques. Pour autant seul Gestionnaire¹ a connaissance de cette nuance au sein des personnes interrogées. Il voit également un intérêt à Recherche Data Gouv différent de celui de Chercheur¹ « Il me semble que Recherche Data Gouv peut aussi moissonner d'autres entrepôts, donc est ce qu'il y a pas un intérêt en tant que catalogue extrêmement large pour quelqu'un qui connaîtrait pas les entrepôts thématiques comme point d'entrée ».

L'émergence des *data papers* est décrite par Chercheur² : « ce qu'on fait beaucoup et les gens aiment bien faire ça, c'est ce qui s'appelle les *data papers*. [...] je pense que ça y a une certaine motivation pour les gens, parce qu'ils peuvent le mettre dans leur CV et ça apparaît comme une véritable publication, donc ça c'est devenu relativement commun on va dire de faire ça ». Pour autant il en met en avant les limites : « dans certains cas, on est tombé dans l'autre extrême, c'est à dire qu'il y a des papiers qui sont faits pour des mises à jour de base de données qui sont complètement marginales et à mon avis, qui méritent même pas une nouvelle publication. Le papier d'origine il suffisait largement et là il y a le côté un peu pervers du truc, c'est que tout le monde veut avoir le plus de publis possible ».

Ainsi, les propos recueillis montrent que les dispositifs et infrastructures de stockage et de diffusion des données occupent une place centrale, mais que leur usage reste hétérogène et parfois confus. SEANOE et SISMER apparaissent comme les deux principales références pour les données océanographiques. SISMER est perçu comme le canal officiel pour les données issues des campagnes de la flotte océanographique française, avec un processus structuré menant à l'attribution d'un DOI. SEANOE, plus souple dans ses contraintes de format, est utilisé pour les données qui ne relèvent pas du SISMER, mais cette flexibilité entraîne parfois une qualité inégale des métadonnées et un manque de standardisation. D'autres plateformes comme GBIF, OBIS ou ENA sont incontournables pour certaines typologies de données spécialisées, notamment la biodiversité ou le séquençage génomique.

Au-delà des entrepôts, les catalogues de métadonnées, comme Sextant, jouent un rôle d'interconnexion en centralisant la description de ressources présentes sur différents dépôts, facilitant ainsi leur repérage. Toutefois, cette distinction entre entrepôt et catalogue reste peu maîtrisée par les chercheurs, seuls les spécialistes en gestion des données en ont pleinement conscience. L'usage de plateformes plus généralistes comme ZENODO est évoqué, mais souvent déconseillé en raison de l'appauvrissement des métadonnées.

Parallèlement, les *data papers* s'imposent comme un outil incitatif au partage, car ils offrent une reconnaissance académique en apparaissant comme des publications. Toutefois, certains interrogés pointent des dérives, comme la multiplication de publications peu pertinentes motivées davantage par la recherche d'un nombre important de publications que par un réel apport scientifique. Ces constats traduisent une tension

persistante entre accessibilité des données, qualité de leur documentation et les logiques de valorisation académique.

4. SCIENCE OUVERTE EN PRATIQUE : ENTRE INCITATION, OBLIGATION ET STRATEGIES D'EVITEMENT

La familiarité avec les principes FAIR varie selon les enquêtés. Chercheur3 témoigne en avoir entendu parler via Argo : « Je vois, mais au quotidien, sans plus ». De plus, Chercheur3 explique son approche : à la suite des campagnes en mer « l'idée c'était que justement j'envoie au SISMER, donc c'est censé être à eux de gérer ça. Et puis nous on fait en sorte que les fichiers des données, des campagnes, on suit les protocoles GOSHIP²⁵ donc c'est censé être FAIR ».

Gestionnaire1 explique avoir intégré le pôle de données et de services qui s'occupe des données marines : ODATIS. Dans ce cadre via des ateliers techniques thématiques ouvert à tous, il a pu monter en compétences sur « les bonnes pratiques, [qui ainsi] percolent à l'ensemble des responsables des CDS²⁶ d'ODATIS. [...] C'est vraiment un moyen très utile d'acquérir au moins des références vers les bonnes technos qu'il faut creuser par la suite ». Par ailleurs, Gestionnaire1 poursuit en indiquant que c'est dans le cadre d'une ANR qu'il a pris « conscience qu'il fallait qu'on s'y mette. [...] En tant que responsable de certains jeux de données, on a été amené à évaluer leur degré de fairness et donc là, il a bien fallu un s'approprier les principes et deux regarder dans quelle mesure on était F, A, I, R et c'est super intéressant ».

Chercheur2 enseigne auprès d'étudiants de master et indique qu'il (avec un collègue enseignant) mentionne ces principes au cours de leurs unités d'enseignement : « Donc je pense pas non plus qu'on puisse dire que c'est un cours complet là-dessus. Et je pense pas que ni lui ni moi [...] on est des véritables spécialistes de tout ce que y a derrière ce principe là, mais en tout cas [les étudiants] ont connaissance de l'existence de ce principe ».

L'intégration des principes FAIR se fait progressivement mais Chercheur1 explique voir encore des réticences et notamment une application douteuse des principes FAIR via une mise à disposition de données volontairement incomplètes : « j'en connais qui font ça, qui mettent des données de séquences de l'ADN sur genbank par exemple, mais qui donnent pas les coordonnées géographiques précises du point de l'échantillonnage qui donnent le strict minimum. Donc si y a pas de métadonnées propres et standardisées, ça sert à rien. Donc y a quand même beaucoup, beaucoup, beaucoup d'efforts à faire pour que les données soient vraiment utiles ».

Gestionnaire1 via ses activités a pu organiser une journée science ouverte dans son laboratoire de travail avec une partie dédiée à la littérature scientifique, ainsi qu'une

²⁵ GOSHIP : Global Ocean Ship-based Hydrographic Investigations Program. Programme océanographique.

²⁶ CDS : Centre de Données et de Services. Structure fournissant l'accès, la gestion et la diffusion d'informations spécialisées.

partie dédiée à la gestion des données. Il y est intervenu « pour tenter d'expliquer à [s]es collègues la démarche de fairisation d'un jeu de données [biologiques] pour que concrètement ils voient ce que ça implique ». Dans le cadre d'un PPR où il est coporteur du groupe de travail sur la gestion FAIR des données il indique ne pas arrêter « aux différents COPIL²⁷ de rappeler ce que sont les principes FAIR et ce que ça implique pour le partage des données, ce qui est parfois pas facile ». Dans ce contexte, il propose également un webinaire : « alors on l'a pas intitulé FAIR, mais on dit à la fin que c'est du FAIR, mais le webinaire va s'intituler 'comment réutiliser et partager les données marines ?' Et ça finit par 'le FAIR sans le dire' ».

Les formations et l'accompagnement restent limités. Seul Gestionnaire1 a proposé des éléments de réponse sur ce sujet. Notamment, Gestionnaire1 décrit les efforts d'ODATIS : des consortiums d'expertises scientifiques se réunissent autour d'une thématique quelques fois par an et permettent de « mettre en place aussi des recommandations [...] qui relèvent donc de la thématique, comment les enrichir en métadonnées comment les partager, sur quelle plateforme, et cetera et cetera. Et donc, ça, c'est aussi, c'est un moyen de coconstruire avec la communauté les recommandations qu'on essaie de diffuser par la suite ». Le CDS ODATIS est aussi souligné et décrit par Chercheur1. Gestionnaire1 souligne l'importance de l'intégration dans la formation des personnels : « c'est quelque chose qu'on fait volontiers, hein, promouvoir l'utilisation des entrepôts ».

Par ailleurs, Gestionnaire1 raconte l'expérience concrète issue d'un projet auquel il a participé activement. Dans ce cadre ont été organisées des sessions avec tous les porteurs de groupe de travail afin de leur expliquer ce qu'est un plan de gestion de données, son utilité ainsi qu'un accompagnement à la prise en main de l'outil via des sessions individuelles. Malgré cela, Gestionnaire1 déplore que « ça a été un peu la déshérence [...]. On se rend compte que c'est un peu un calvaire pour les gestionnaires de données d'aller relancer régulièrement les personnes en charge de jeux de données pour qu'ils [...] fournissent l'information permettant de remplir le DMP ».

Les connaissances juridiques apparaissent floues. Chercheur2 témoigne : « J'ai énormément d'interrogations là-dessus et j'ai eu jusqu'à maintenant que des informations plutôt contradictoires. Enfin, je parlais du principe en tout cas que toutes les données qui étaient acquises dans le cadre d'une activité, on va dire publique ou de services publics, elles devaient être diffusées librement et le plus rapidement possible. Alors après il y a la règle d'un pseudo embargo de 2 ans dont je connais pas la valeur légale ni quoi que ce soit. Donc [une personne proche spécialisée dans le droit], il m'a donné des réponses justement là-dessus, qui sont beaucoup moins claires que ce que je pensais. C'est à dire que, selon lui, il y a pas vraiment de base légale à l'obligation que les données soient rendues immédiatement publiques, c'est plutôt une démarche qui est personnelle. À titre personnel, je considère que sur des données publiques, l'idée c'est de faire plus vite

²⁷ COPIL : Comité de Pilotage.

possible que ces données soient disponibles. En tout cas c'est pas clair pour moi qu'est ce que légalement on doit faire ou pas ? ». Chercheur3 exprime une vision simple mais confuse : « Les données sont censées être libres, dès leur acquisition. Non, c'est pas ça ? Et ben je pense que c'est les organismes de recherche qui sont propriétaires ». Pour Gestionnaire1, « tout projet qui a un financement public avec au moins 50 % se doit de mettre ses données en open data ».

Cependant, pour certains domaines, les obligations sont claires. Chercheur1 précise une « obligation légale quand tu fais du suivi sur les politiques publiques en environnement que les données soient bancarisées et sauvegardées. » Par exemple dans le contexte de la « Directive-cadre 'Stratégie pour le milieu marin'²⁸, ils ont obligation de bancariser les données sinon ils ne sont pas payés ».

Sur le sujet et la perception des injonctions institutionnelles, Chercheur1 souligne les différences entre formations obligatoires : « tu rentres dans un labo, t'as une obligation de suivre une formation hygiène et sécurité. Tu l'as pas, c'est le code professionnel, on peut te virer. A aucun moment, t'as une obligation légale de suivre une formation sur la donnée. Donc y a des efforts de sensibilisation, accompagnés par nos tutelles, y a aucun doute là-dessus. Il y a un plan de science ouverte à l'université, [...] Il y a vraiment beaucoup de choses de faites par nos tutelles, mais rien à ma connaissance, n'a de caractère obligatoire et contraignant. [...] Si t'as pas envie, tu peux continuer à faire comme avant quoi. Bah il va rien t'arriver ».

Chercheur2 note les différences entre tutelles : « j'ai jamais vu de message commun sur ces questions-là [issu des tutelles du laboratoire]. J'ai le sentiment quand même que le CNRS a une véritable réflexion et communique et est assez moteur là-dessus. Alors à l'université, ben comme bien souvent là-dessus, je pense qu'ils sont complètement largués. Ils ont dû envoyer une fois un message sur la science ouverte avec une sorte de blabla qui a aucun sens. [...] Ceci dit, à mon avis l'IRD²⁹ je pense qu'ils ont une véritable réflexion là-dessus, en particulier dans leur mission vis-à-vis du Sud, [mais] je la connais pas en tout cas ». Pour un éventuel accompagnement, Chercheur3 se « tournerai vers des collègues de bureau ou des collègues du SISMER. C'est acquis ou acté que c'est la science ouverte qu'il faut, c'est dit » par ses tutelles.

Gestionnaire1 évoque les ressources disponibles : « ODATIS fait quand même un effort d'aller dans différents laboratoires pour dire 'Regardez si vous avez des besoins, il y a tel et tel contact pour vous accompagner, notamment dans la gestion de données' », tout en insistant sur le fait que les ressources ne sont « pas infinies dans ODATIS ». Chercheur2 mentionne aussi l'existence du pôle ODATIS qui ont « toujours et systématiquement répété qu'ils étaient à la disposition des personnes pour aider dans la gestion de ces plans de données. Et, je pense qu'ils le font avec beaucoup de bonne volonté

²⁸ Directive-cadre « Stratégie pour le milieu marin » (DCSMM) : cadre européen visant à atteindre ou maintenir le bon état écologique des mers et océans.

²⁹ IRD : Institut de Recherche pour le Développement. Organisme public français dédié à la recherche et à la coopération scientifique internationale.

donc par contre je pense qu'on peut pas dire qu'on a aucune aide vis-à-vis de ça, c'est à dire on a quand même des gens qui sont spécialistes de ça qui sont prêts à aider pour ça ».

Ainsi, l'analyse des entretiens révèle une familiarité très variable avec les principes FAIR, oscillant entre connaissance superficielle, application indirecte et véritable appropriation. Certains, comme Chercheur3, les approchent de façon implicite via le respect de protocoles internationaux (ex. GOSHIP) ou par délégation aux infrastructures comme SISMER, tandis que Gestionnaire1 a développé une expertise active, notamment à travers son implication dans ODATIS et ses évaluations de la « fairness » de données dans le cadre de projets ANR. Les initiatives pédagogiques restent inégales : Chercheur2 les mentionne dans un cadre d'enseignement de master, mais de manière non exhaustive, et Chercheur1 souligne l'existence de pratiques contournant l'esprit du FAIR en fournissant des données volontairement incomplètes.

Les formations et dispositifs d'accompagnement apparaissent concentrés autour de quelques acteurs moteurs. En particulier Gestionnaire1 décrit des actions concrètes dans le cadre de son poste : organisation de journées thématiques, interventions internes pour expliquer la « fairisation », animation de webinaires et participation à la co-construction de recommandations au sein de la communauté ODATIS. Malgré ces efforts, l'appropriation reste inégale : la mise en œuvre des plans de gestion de données se heurte à un manque d'engagement durable, nécessitant des relances répétées et traduisant la perception du DMP comme une contrainte administrative.

Les connaissances juridiques sont marquées par des incertitudes et des divergences d'interprétation. Certains chercheurs se réfèrent à un principe général de mise à disposition rapide des données publiques, sans en connaître précisément les bases légales, tandis que d'autres rappellent des obligations claires dans des contextes spécifiques (par exemple dans le cadre de politiques publiques). Enfin, la perception des injonctions institutionnelles oscille entre incitations fortes mais non contraignantes et disparités entre tutelles : le CNRS et l'IRD sont jugés plus actifs que les universités, et les ressources offertes par ODATIS sont reconnues, tout en restant utilisées de manière plus que limitée. Cette diversité de postures semble traduire un environnement où l'incitation prédomine sur l'obligation, laissant une large place aux pratiques mimétiques entre collègues et à l'initiative individuelle, sans démarche de standardisation.

5. OUVRIR... MAIS SOUS CONDITIONS : ARBITRAGES, TENSIONS ET RAPPORTS DE FORCE

Les résistances générationnelles sont identifiées par Chercheur1 : « j'ai bien vu l'évolution maintenant, les gens sont quand même plus enclins à diffuser leurs données qu'ils l'étaient dans le passé, mais il reste encore des vieux, on peut dire des vieux cons, même des jeunes cons hein, qui pensent que la donnée, elle est à eux quoi, et que, 'Ben après eux le déluge'. Donc ça, ça existe encore hein, c'est indéniable y a des gens qui sont pas contre, mais qui pensent que c'est pas noble et c'est pas leur travail, eux ils sont là

pour penser et qu'ils ont pas du temps à perdre à bancariser leurs données ». Gestionnaire1 a une perception similaire à celle présentée par Chercheur1 et évoque l'effort technique nécessaire : « Ligne par ligne, colonne par colonne, [...] il faut bien que celui qui a produit la donnée qu'il soit chercheur, qu'il soit ingénieur, qu'il soit technique, hein, peu importe, il faut bien qu'à un moment il y ait quelqu'un qui remplisse un tableau. En suivant des standards, sans se dire 'Oh, je m'en fous des colonnes, des lignes de ceci, de cela'. Le format standard qu'on lui a distribué, qu'il le remplisse, et puis qu'il appuie sur 3 boutons pour que le fichier il soit téléchargé, contrôlé. [...] Donc il peut y avoir de l'aide technique à différents niveaux, mais à un moment il faut qu'il y ait une personne qui accepte que c'est son travail ».

Chercheur1 illustre avec un exemple dramatique : « Les premiers suivis [biologiques] et hydro à [laboratoire], c'est dans les années 80, par [M. X] ». Celui-ci est parti à la retraite en emportant son ordinateur avec les données en question dessus. Chercheur1 avait pu récupérer les données et possède une des deux copies existantes. Pour autant, il avait reçu l'instruction de les « garder, surtout pas les diffuser ». Mais convaincu de l'importance de les rendre disponibles, il va « essayer de faire en sorte qu'elles soient bancarisées, même si c'était pas ce qu'il souhaitait. Parce qu'on peut pas faire autrement ». Il insiste également sur le fait que bancariser n'est pas une fin en soit, « si c'est bancarisé dans un truc qui ensuite est fermé à double tour », ce pourquoi il souhaite que la bancarisation respecte les principes FAIR.

La question de la valorisation du travail est centrale. Gestionnaire1 explique : « je pense que les scientifiques, ils considèrent que ça prend du temps sur d'autres activités qui pour eux sont plus valorisables ». Chercheur1 développe : « c'est humain hein, de regarder comment [on] va être évalué par les autres et pourquoi on s'embêterait à mettre en Open Access des données qu'on a acquis après beaucoup d'heures de travail et parfois de souffrance pour que d'autres en tirent les marrons du feu, fassent des gros papiers, puis aient une carrière qui avance 2 fois ou 3 fois plus vite que toi. Donc si t'as pas la reconnaissance du travail effectué, ça peut pas fonctionner, donc il faut pas juste imaginer que les gens qui gardent les données c'est des horribles égoïstes qui n'ont rien compris et qui seraient une vieille garde arriéré. Il peut y en avoir, hein mais c'est pas que ça ».

Chercheur3 témoigne ses pratiques suite à une campagne en mer. Lors de la campagne, certaines données ont été envoyées « quasi-temps réel au SISMER ». Pour le reste, « on attend ». En effet, l'objectif est « de publier avant d'envoyer des données ». Ceci implique de faire des traitements ce qui implique que le délai moyen est de 2 ans « parce que nous, il nous faut un an en gros pour faire la calibration des données, pour finaliser le rapport pour les données campagne. Et puis après on essaie de publier et une fois que c'est publié, on met à disposition ».

Plusieurs craintes émergent. Gestionnaire1 identifie « la crainte que des équipes plus instrumentées valorisent les données avant les producteurs, avec des arguments qui se tiennent parce que, par exemple, aller échantillonner en mer avec un bateau, ça coûte des sous, ça coûte des ressources humaines et donc on va pas tout de suite mettre en accès

les données qu'on a acquises avant que nous l'équipe qui avons collecté on puisse publier ».

Gestionnaire1 mentionne aussi une autre motivation « de pas le faire de manière propre, c'est d'espérer que les gens qui s'intéressent à leurs données les contactent, au plus on dépose des données propres, bien documentées, bien décrites, au moins on a besoin de revenir vers le producteur. Et ça, certains producteurs disent, 'mais moi, les gens ils peuvent pas utiliser mes données sans me contacter' ».

Chercheur1 évoque les fantasmes autour de la peur de ne pas avoir la primeur sur ses propres données et qu'elles soient exploitées par d'autres. Il donne un exemple concret de données de suivis effectués dans un de ses lieux d'étude qu'il a mis à disposition au niveau européen. Pour autant « ça fait 3 ans que j'ai un papier prêt à soumettre et que j'arrive pas à me trouver une semaine tranquille pour le finaliser donc ça traîne, ça traîne, et personne n'a piqué les données, a écrit le papier avant moi. Donc je pense que là-dedans y a un peu de fantasmes ».

Par ailleurs, la question de la qualité des données préoccupe Chercheur3 : « c'est vraiment une crainte en fait que les données soient mal utilisées ». Il explique que cela a déjà pu avoir lieu. Dans son exemple, des capteurs de flotteurs Argo présentaient des dérives de salinité connues, les données ont donc été corrigées. Cependant, des données temps réel ont été utilisées dans des analyses océanographiques, pour créer des produits océanographiques³⁰. Ainsi, malgré l'identification et la correction de ces données, les produits n'ont pas été mis à jour ce qui « génère des publications qui montrent qu'il y a un problème dans les données Argo, alors qu'en fait il y a pas de problème parce qu'on les a corrigées, c'est juste qu'elles ont pas été mises à jour. Donc oui c'est un problème. » Par ailleurs, Chercheur3 insiste sur le fait que les chercheurs « endossent une certaine responsabilité sur la qualité des données qui sont mises à disposition et donc je comprends en fait le frein de te dire 'Ben attention moi je veux être sûr que la donnée que je mets à disposition elle est de bonne qualité parce que sinon j'ai peur' [...] donc c'est une crainte et en tout cas c'est quelque chose auquel on veille ».

Chercheur3 développe longuement les enjeux liés aux campagnes en commençant par les comparer à ce qui existe au niveau des flotteurs Argo. Chercheur3 indique qu'« en physique, ça a été quand même plutôt des précurseurs de ce partage des données. Argo a énormément joué dans cette idée de partage, mais même au tout début du projet Argo, quand il y avait quelqu'un qui mettait un flotteur à l'eau, finalement, il disait un peu, 'c'est mes données, c'est à moi'. Voilà donc petit sentiment de possession, qui peut s'expliquer ou se justifier ». De son point de vue, cela n'est plus le cas au sein du programme Argo au vu de son expansion et de la densité de flotteurs disponibles. Cependant, cela se retrouve pour les données issues de campagnes océanographiques. « En fait, c'est énormément de travail, de nuits blanches, de stress, d'argent. [...] Ben quelque part par rapport à quelqu'un qui reste dans son bureau et qui fait 'clic clic', qui télécharge, ben la personne qui s'est épuisée à faire sa campagne en mer [...] des fois elle

³⁰ Les produits océanographiques : données grillées.

a pas le temps de publier et il y a d'autres gens qui ont déjà récupéré des données qui ont publié avec. Enfin je pense que c'est un peu le risque ou c'est un peu la crainte ou c'est un peu une des raisons de faire ça. [...] ». Il considère ainsi comme « frein légitime » de considérer que beaucoup de temps et d'énergie ont été investis lors de campagnes en mer et qu'il semble juste d'avoir « du temps pour valoriser les données pour lesquelles [...] j'ai dépensé de l'énergie ».

A l'opposé, Chercheur3 prend le temps d'expliquer pourquoi selon lui Argo fonctionne et décrit aussi les « non freins ». Essentiellement, cela est lié à la nécessité de la communauté. En effet, « c'est pas les 3 flotteurs ou les 20 flotteurs qu'on met à l'eau qui vont nous permettre d'avoir la vision globale. En fait, il y a aucun pays qui peut à lui seul avoir suffisamment de flotteurs pour avoir la vision globale qui est nécessaire. Donc en fait dans Argo vu ce qu'on observe c'est l'océan à l'échelle globale, on a besoin de tous les autres. » Cet aspect collectif implique que « le partage des données et donc la mise à disposition des données est intrinsèque au besoin et au projet ». De plus, Chercheur3 rappelle que les données sont également utilisées par des services opérationnels ce qui implique qu'elles soient disponibles sous 24h pour ceux-ci. Il synthétise le succès du partage des données dans Argo : « C'est inscrit dans le projet au départ et consciemment ou inconsciemment tout le monde sait que ça n'a d'intérêt que si c'est mis en commun ».

Chercheur3 poursuit en extrapolant « je pense qu'un frein à la mise à disposition des données, c'est que il y a pas cet intérêt collectif, c'est à dire que finalement les gens ils vont mettre à disposition leurs données si ils y voient un intérêt propre, on va dire. Et s'ils n'ont pas intérêt ? Et Ben voilà. Ils vont traîner des pieds, ça va pas être la motivation première. En fait ça risque de ne pas être dans les priorités ».

Les différences entre disciplines sont marquées. Chercheur1 observe que le rapport à la donnée varie entre « celui qui acquiert des données de température ou des données avec des capteurs automatisés, et celui qui acquiert de la donnée biologique ». Chercheur1 détaille ces différences avec le temps passé par chaque type d'acteur à l'issue d'une campagne océanographique « celui qui fait des mesures de paramètres physiques, globalement, il revient, il a déchargé ses profils CTD³¹, il a tout, tout est prêt, il peut travailler de suite. Le chimiste, il va lui falloir quelques semaines hein pour ben acquérir ces données d'oxygène, d'alcalinité, etc. Le biologiste, alors maintenant on a inventé l'imagerie, mais l'imagerie, ça reste grossier d'un point de vue taxonomique. Et si tu imagines [quelqu'un] qui passerait des semaines et des mois derrière, à la bino [loupe binoculaire] à compter à identifier chacun des organismes, Ben lui il va peut-être pas avoir le même rapport à la donnée parce qu'il a passé un an de sa vie à dépouiller les échantillons d'une campagne ». De plus, il insiste sur les temporalités différentes « quand il a fini de dépouiller, on lui dit 'Ah Ben tiens, c'est la nouvelle campagne qui démarre'. Il n'a pas vraiment le temps d'exploiter ce qu'il a acquis l'année dernière. [...] Entre temps le physicien, il a déjà sorti 2 papiers car lui quand il a débarqué, toutes ses données était

³¹ CTD : Conductivity Temperature Depth. Instrument océanographique mesurant la conductivité (salinité), la température et la profondeur (pression) dans la colonne d'eau.

là. » Ainsi pour Chercheur1 le rapport à la donnée est lié à l'« investissement en temps, en énergie que tu dois consacrer à l'acquisition de la donnée élémentaire ».

Chercheur2 analyse cette question et répond de manière plus nuancée. Pour lui, il subsiste des « mauvaises habitudes qui existaient à l'époque, parce que y avait pas d'infrastructure pour le faire [ni] véritablement de compréhension des enjeux. » Il a le sentiment que la question de la gestion des données est reléguée au rang de « question subalterne un peu relou [...] pour laquelle il nous faut du personnel pour nous aider à faire ça, donc le sous-entendu là-dedans, c'est que ça nous saoule de faire ça. Et donc en gros, j'envoie mon tableau Excel dégueulasse et la personne elle se démerde pour le mettre dans le bon format. » Pour autant, ce détachement et ce désintérêt ne sont pour lui pas justifiables, « c'est une véritable erreur ça parce que je pense que la gestion des données [...] ça demande une véritable expertise sur les données », notamment pour les formats et les unités.

Chercheur2 poursuit avec un exemple concret qu'il a pu rencontrer au cours de son parcours professionnel. Le pH qui est une mesure d'acidité et donc d'acidification peut être mesuré de diverses manières (souvent en référence à une température d'analyse) impliquant « énormément d'unités différentes [...] et] valeurs numériques qui sont différentes ». Il a connaissance d'un lieu d'étude sur lequel le pH a été mesuré durant plusieurs années de manière partielle, sans relevé de température lors de l'analyse. Pour autant il ne « jette pas la pierre, c'était simplement la pratique qui était faite ». Par la suite, la mesure et l'analyse du pH dans ce lieu est devenue beaucoup plus rigoureuse suivant des méthodes plus adaptées et détaillées. Ainsi, « la qualité des données s'est très fortement améliorée ». Le problème qu'il souhaite mettre en avant avec cet exemple est l'absence d'indication de ce changement « radical » de méthode d'analyse lors de l'accession aux données. Dans un congrès, il a assisté à une présentation de ces jeux de données pour lesquels une tendance globale avait été calculée « qui était une tendance à l'acidification qui n'avait aucun sens, [...] qui était 10 à 100 fois plus élevé que l'acidification réelle. » Ceci n'a pas eu lieu dans une réunion de néophytes mais « à la réunion du pôle de données. Qui font du mieux qu'ils peuvent, mais ça demande quand même une certaine connaissance rigoureuse des paramètres que tu mesures et donc je pense que c'est une erreur que nous on fait de pas s'approprier ces problématiques parce que finalement Ben ça veut dire que alors non seulement les données sont mal archivées, mais en plus ça peut conduire à des interprétations qui sont totalement débiles ».

Des facteurs générationnels sont également mis en cause. Tout d'abord, Chercheur1 observe que « la jeune génération, elle a compris qu'il fallait la mettre [la donnée] à disposition. Elle le fait, elle se pose pas la question. [...] c'est les plus anciens qui sont restés les plus réfractaires pendant longtemps. Si tu prends une seule discipline, où le labeur pour obtenir la donnée est le même que t'es 70 ans, 50 ou 30 ans, ouais, les anciens ils sont plus réticents ».

Chercheur2 indique que pour lui « y a pas tous les jeunes chercheurs qui sont des personnes irréprochables là-dessus » même s'il distingue un « gros biais générationnel ». Pour autant, il indique trouver cela « beaucoup plus problématique » de la part d'un jeune

chercheur. « Pour un jeune chercheur effectivement qui arrive, je pense que la situation est complètement différente. Ils ont été formés complètement différemment et donc là, je pense que c'est de la pure fainéantise ». Dans le cas de ses collègues de la génération « qui arrivent en fin de carrière là, qui ont 60 ans », les pratiques étaient différentes. A l'époque, à l'issue d'une campagne en mer, le jeu de données allait donner lieu à une publication scientifique parce qu'il y avait moins de données. Ainsi les chercheurs gardaient les données « consciencieusement » post-campagne puis une fois la sortie de la publication scientifique, ils considéraient que « le travail était fait, et ce jeu de données, il y avait pas nécessairement besoin de l'archiver ». De plus, il mentionne la différence au niveau des publications qui intégraient les données dans le corps même du papier du fait de leur nombre beaucoup plus limité « il y avait des tableaux dans les papiers des années 80. »

Ainsi, les entretiens révèlent que les arbitrages autour de la mise à disposition des données marines révèlent une tension entre ouverture et réticences, souvent nourrie par des enjeux générationnels, des considérations pratiques et des logiques de valorisation scientifique. Si une évolution vers une plus grande acceptation est reconnue, des résistances persistent, motivées par la perception que la bancarisation n'est pas le cœur du métier du chercheur ou qu'elle détourne du temps de travail jugé plus valorisant. Les freins vont du sentiment de propriété sur les données (parfois jusqu'à leur rétention) à la volonté de maîtriser la qualité et l'usage qui en sera fait, en passant par la crainte de se voir devancer dans l'exploitation scientifique de données coûteuses et longues à acquérir. Ces réticences s'expriment particulièrement dans le contexte des campagnes en mer, où l'investissement humain et matériel est perçu comme justifiant un délai pour la valorisation prioritaire par l'équipe productrice. À l'inverse, des initiatives collectives comme le programme Argo montrent qu'un partage inscrit dès la conception du projet, associé à un besoin collectif fort, peut lever ces freins.

Les craintes liées à la réutilisation des données recouvrent plusieurs dimensions : peur d'une exploitation prématurée par des équipes mieux outillées, volonté de conserver un contact obligatoire avec le producteur, ou encore préoccupations concernant la qualité et la bonne interprétation des données. Des exemples concrets, tels que des erreurs de traitement ou l'absence de documentation sur des changements méthodologiques majeurs, illustrent les risques d'une mise à disposition insuffisamment encadrée. Ces enjeux mettent en lumière l'importance d'une expertise technique chez les gestionnaires de données pour garantir la cohérence et la fiabilité des données, et éviter des interprétations erronées, y compris dans des cercles scientifiques spécialisés.

Enfin, des différences structurelles existent entre disciplines et générations. Les temporalités de traitement des données varient fortement entre physique, chimie et biologie marines, influençant le rapport des chercheurs à la diffusion. Les mesures automatisées peuvent être valorisées quasi immédiatement, tandis que certaines données biologiques nécessitent des mois, voire un an, de traitement minutieux, renforçant la réticence à partager rapidement. Sur le plan générationnel, si les jeunes chercheurs sont globalement plus enclins à l'ouverture, des exceptions existent, et la résistance des plus

anciens est souvent liée à des habitudes ancrées dans un contexte scientifique où la bancarisation n'était ni une norme ni une obligation. Les motivations, qu'elles soient d'ordre stratégique, culturel ou liées à la charge de travail, montrent que l'ouverture des données reste une pratique sous conditions, parfois contrainte par des facteurs individuels plutôt que des bénéfices collectifs.

6. RESSOURCES HUMAINES, ORGANISATION ET LEVIERS D'ACTION POUR RENFORCER LA GESTION DES DONNEES

Les projets collaboratifs révèlent des différences notables de pratiques. Gestionnaire¹ observe que son témoignage est orienté recherche « on n'est pas du tout politique publique, et je sais que les pratiques sont différentes. Notamment parce que politique publique dépend pas mal du ministère de l'environnement pour ce qui nous concerne et avec des pratiques complètement différentes en matière de gestion de données, ils sont beaucoup plus franco-français et ça se voit dans les référentiels qu'ils utilisent ». Il poursuit en comparant avec d'autres domaines plus éloignés « le plus exotique avec qui j'ai discuté sans avoir aucune connaissance du tout c'est la génomique marine. [...] j'ai l'impression qu'ils sont quand même organisés de manière beaucoup plus rigoureuse parce qu'ils gèrent un flux de données qui est beaucoup plus important finalement que ce que nous on peut générer lors d'une campagne en mer. J'imagine qu'il y a aussi des imperfections certainement, dans ces données là, mais je suis pas assez spécialiste pour pouvoir en voir ». Chercheur² va dans le même sens en mentionnant également la génomique marine comme plus structurée et plus au point sur la mise à disposition des données.

Chercheur³ présente le projet collaboratif qu'il connaît bien : le programme Argo qui illustre des bonnes pratiques collaboratives. Dans ce cadre, il met en avant le lien entre une partie « où il y a des gens, [la] gestion des données c'est leur boulot, qui est entremêlée avec des scientifiques qui s'intéressent aussi ». Ces personnes se réunissent régulièrement entre gestionnaires de données et scientifiques experts pour prendre des décisions sur les variables scientifiques mesurées, leur traitement et leur partage. « C'est un groupe qui marche bien. Les gens sont motivés et il y a vraiment cet entrelacement entre des gestionnaires de données et des scientifiques. [...] Tout est bien structuré, organisé et du coup, on n'a pas de questions à se poser ».

Une évolution positive des pratiques est observée par Chercheur¹. Il indique que les tutelles évoluent progressivement « aujourd'hui, on évalue peut-être plus la carrière et le dossier d'un chercheur ou d'un enseignant chercheur, comme on le faisait il y a 10 ou 15 ans et que aujourd'hui publier des *data papers*, mettre en Open Access des datasets c'est intégré dans ton dossier ». Il comprend la focalisation de certaines personnes sur les publications et les dossiers d'avancement de carrière « c'est humain hein, de regarder comment va être évalué par les autres et pourquoi on s'embêterait à mettre en Open Access des données qu'on a acquis après beaucoup d'heures de travail et parfois de souffrance

pour que d'autres en tirent les marrons du feu, fassent des gros papiers, puis aient une carrière qui avance 2 fois ou 3 fois plus vite que toi ». Il mentionne également les comportements de rétention des données sur lequel il souhaite nuancer le propos « si t'as pas la reconnaissance du travail effectué, ça peut pas fonctionner, donc il faut pas juste imaginer que les gens qui gardent les données c'est des horribles égoïstes qui n'ont rien compris et qui seraient une vieille garde arriéré. Il peut y en avoir, hein mais c'est pas que ça ».

Chercheur2 reconnaît l'amélioration du traitement des dossiers de carrière « de l'expérience que j'ai eue au CNU³² ou dans différentes évaluations, c'est plus reconnu, c'est plus compris comme un critère ».

Interrogés sur les améliorations nécessaires, plusieurs pistes sont évoquées. Au niveau financier, Gestionnaire1 propose « le fait d'identifier dès le démarrage d'un projet, les ressources nécessaires pour assurer une bonne gestion des données ». Il donne l'exemple de soumission d'un projet ANR dans lequel des frais de gestion doivent être budgétisés au dépôt du dossier. « Et ben pourquoi on dirait pas, quand on dépose un projet où y a de la donnée, Eh bien, 10% ou 15% doivent être réservés pour la gestion de la donnée ». Il insiste sur le recrutement « de manière pérenne » de spécialistes de la gestion des données comme second levier. « Je comprends les chercheurs qui disent c'est pas mon domaine, c'est pas mes compétences. C'est un paysage mouvant, donc il faut vraiment des personnes qui suivent ça de très près et qui peuvent accompagner les chercheurs ». Il poursuit sur une difficulté au niveau du recrutement lié à l'ouverture d'un poste de gestion de données « quand on fait un profil, on hésite parfois entre l'emploi type au niveau de la fonction publique, y a différents BAP³³ et c'est compliqué parce que parfois on se dit que les compétences les plus importantes, c'est les compétences disciplinaires, mais parfois on se dit que les compétences les plus importantes, c'est les compétences en documentation, en IST et là aussi c'est un petit problème d'essayer de faire un profil qui concilie les 2 ». Pour ces profils techniques et scientifiques de gestionnaire de données, il insiste sur le fait que selon lui l'absence « dans le référentiel des métiers de la fonction publique, [d']un emploi type qui correspond à la gestion des données » complexifie le recrutement.

Par ailleurs, Chercheur2 insiste sur le temps et les ressources humaines idéales « du temps de personnes », ainsi que « le temps d'organiser des réunions pour en discuter. [...] C'est vraiment une question de temps qu'on est prêt à y consacrer parce que ça occupe du temps ». Pour lui, cette question du temps est centrale et justifie le recours au *data papers* parfois à outrance « Si tu arrives à valoriser ce temps que t'occupes, je pense que c'est pour ça que les gens se jettent dans ces histoires de *data papers*. C'est que effectivement, si t'as passé du temps tu te dis 'bah il faut que ça soit reconnu quelque part' et malheureusement la seule manière de reconnaître notre travail, c'est d'avoir un papier qui

³² CNU : Conseil National des Universités. Instance de recrutement académique.

³³ BAP : Branche d'Activité Professionnelle. Classification des métiers dans la fonction publique de la recherche et de l'enseignement supérieur.

montre ce que t'as fait et c'est la raison pour laquelle y a ces *data papers* qui sortent ». En effet, les *data papers* de mise à jour de base de données sont justifiés par « une personne qui s'est fait chier 3 semaines, 4 semaines, un mois à le faire. En réalité, ce papier, il a pas d'intérêt, c'est juste en gros une forme, une manière de reconnaître le temps qu'a passé cette personne ». Un second levier selon lui serait d'« avoir des personnes compétentes » au sein de chaque laboratoire pour pouvoir les solliciter au quotidien et être formé et accompagné dans les différents sujets ayant trait à la gestion de données, même s'il a conscience de l'effort de ressources humaines que cela représente. Dernièrement il indique que « des injonctions, des instructions un petit peu plus claires ça aiderait aussi ».

Chercheur3 rejoint la nécessité de ressources humaines plus fournies et évoque des besoins d'intermédiation « quelqu'un qui fait le lien entre les standards internationaux, les demandes des centres de données et puis les scientifiques ». Il reconnaît parfois avoir eu des soucis à ce sujet lors de la mise en base des données « je sais qu'en fait il y a un problème au niveau du centre de données, c'est à dire que des fois on leur envoie des données. Et quelque part, ils savent pas mettre au bon format où ils reçoivent les données. Mais ouais je pense qu'il manque quelqu'un qui fasse l'intermédiaire ». Chercheur3 considère aussi cela dû à un problème de déconnexion « je pense que les centres de données ils sont un tout petit peu déconnectés aussi de ce qui se passe dans les labos. Alors après voilà, je sais pas comment faire ».

Une clarification des discours institutionnels est souhaitée par Chercheur2 qui partage « l'impression qu'il y a même le travers inverse, c'est à dire que il y a tellement des beaux supports [des tutelles] généralement avec voilà des schémas, là avec des couleurs pour les différentes lettres et tout ça. [...] tu comprends rien finalement tu sais pas complètement qu'est-ce qu'il faut que je fasse pour fairiser mes données. »

La nécessité de contraintes plus structurantes est relevée par Chercheur1 « je pense que la bonne volonté ça suffit pas. Et de temps en temps, faut donner des coups de bâton et parce que t'auras toujours des petits malins qui tricheront qui diront que de toute façon ils s'en foutent ». Plus spécialement lorsque cela implique des enveloppes conséquentes de financement via de l'argent public « si t'as un gros projet ANR, si on te donne 600 kilos euros pour mener de la recherche pendant 2 ou 3 ans, la moindre des choses c'est que tu passes un peu de temps à mettre tes données à disposition. Et puis Ben si c'est pas fait, ben t'es plus habilité à déposer de nouveaux dossiers. C'est aussi facile que ça ». Il poursuit son propos avec un bémol sur les modalités de contrôle de ces actions mais aussi de leurs qualités qui seront difficiles à évaluer et modérer « mal fait et versus bien fait ».

Chercheur1 développe l'idée d'obligations graduées indiquant que « l'aspect sensibilisation est important mais à un moment ça suffit pas, à un moment il faut imposer. » Il nomme comme exemple l'observation pour laquelle les comportements sont plus contraints. Pour faire partie d'un SNO, ce qui apporte des avantages « on a une reconnaissance, on a plus de chances d'être financé pour pérenniser nos suivis, mais la contrepartie, c'est que les données sont distribuées de manière ouverte et interopérables.

Il faut que chacun y trouve son compte, mais moi je ne crois pas que la seule bonne volonté ou la seule sensibilisation, ça va vraiment faire bouger les lignes. Parce qu'à un moment, il faut manier la carotte et le bâton ».

Chercheur2 évoque l'effort de sensibilisation auprès des étudiants qu'il a pu encadrer et avoir en cours « je pense avoir systématiquement en tout cas insisté là-dessus, expliqué, montré comment ça marchait ». Il mentionne le master dans lequel il enseigne (dans lequel il a insisté pour que soient enseignés les principes FAIR). Il y a inclut un cours sur l'archivage et la réutilisation des données via des travaux dirigés et des oraux présentant des projets pratiques de réutilisation de données. Il insiste sur l'autonomie que ces pratiques peuvent leur apporter « on leur montre en plus l'énorme force du truc, c'est à dire que on leur dit, 'bah là vous allez en stage de M2 par exemple. Eh Ben sur votre sujet, vous pouvez vous sans même avoir besoin de votre responsable de stage, aller chercher des données complémentaires', leur dire où est ce qu'il faut aller chercher et ce genre de choses ».

Chercheur 1 complète avec la nécessité, pour obtenir « des métadonnées qui soient propres, qui suivent des vocabulaires standardisés, et cetera [... d'] une éducation à faire [...] auprès des collègues ».

Chercheur3 réfléchit sur ses propres besoins « Au fond tu vois avec tes questions, je me dis, c'est un sujet auquel je me suis pas hyper intéressé et donc peut être qu'effectivement une formation ou une sensibilisation ça pourrait être intéressant ».

Ainsi, les entretiens mettent en avant que les pratiques de gestion des données varient fortement selon les domaines et les cadres d'action, qu'il s'agisse de recherche académique ou de politique publique. Les projets collaboratifs illustrent ces contrastes : la génomique marine, par exemple, est perçue comme plus structurée, gérant des flux de données massifs avec des standards plus rigoureux. À l'inverse, certaines disciplines liées aux campagnes en mer adoptent des approches plus hétérogènes. Des initiatives comme le programme Argo se distinguent par leur gouvernance intégrée, où gestionnaires de données et scientifiques travaillent étroitement, se réunissent régulièrement et prennent collectivement des décisions sur le traitement et le partage des données. Ce fonctionnement, bien organisé et motivant, limite les zones d'incertitude et fluidifie le travail collaboratif.

Une évolution positive se dessine également dans la reconnaissance institutionnelle. Les tutelles intègrent désormais davantage la mise à disposition des données et la publication de *data papers* dans l'évaluation des carrières, ce qui contribue à valoriser un travail longtemps invisible. Toutefois, cette reconnaissance reste insuffisante pour compenser le temps et les efforts investis. Plusieurs chercheurs insistent sur la nécessité de prévoir, dès la conception des projets, un budget dédié à la gestion des données (par exemple un pourcentage fixe des financements) et de recruter de manière pérenne des spécialistes en mesure d'accompagner la communauté. Ces profils hybrides, à la fois

techniques et scientifiques, restent difficiles à recruter, faute d'un emploi type clairement identifié dans la fonction publique. Parallèlement, la disponibilité de ressources humaines qualifiées dans les laboratoires, capables de jouer un rôle d'intermédiation entre standards internationaux, centres de données et chercheurs, apparaît comme un levier clé.

L'amélioration des pratiques passe aussi par une clarification des messages institutionnels, jugés parfois trop abstraits ou déconnectés des besoins concrets. Pour certains, la bonne volonté et la sensibilisation doivent s'accompagner d'obligations plus structurantes, avec un système mêlant incitations et sanctions, notamment lorsque des financements publics conséquents sont en jeu. Des dispositifs comme la labellisation dans un SNO montrent que ce type de contrainte peut être accepté s'il s'accompagne d'avantages tangibles. La formation, enfin, constitue un autre pilier : intégrer l'archivage, la réutilisation et les principes FAIR dans l'enseignement, comme le font déjà certains masters, permet de familiariser les nouvelles générations avec la gestion des données dès leur entrée dans la carrière scientifique. Sensibiliser les chercheurs en activité, via des formations ou ateliers pratiques, pourrait également renforcer l'appropriation des standards et l'amélioration de la qualité des métadonnées en insistant sur des exemples concrets proches du quotidien des chercheurs.

CONCLUSION

Cette recherche s'est attachée à examiner comment les acteurs des laboratoires océanographiques français s'approprient les injonctions institutionnelles en matière de gestion, de partage et de stockage des données de recherche, dans un contexte de transition vers la science ouverte. L'analyse, fondée sur un corpus mixte associant textes institutionnels, ainsi qu'une série d'entretiens semi-directifs, a permis de croiser discours normatifs et pratiques réelles.

Les résultats confirment et nuancent l'hypothèse initiale. Oui, les politiques institutionnelles en faveur de la science ouverte se heurtent à des logiques de compétition académique, de hiérarchisation des priorités et à des incertitudes juridiques persistantes, comme le documente également la littérature sur d'autres domaines scientifiques (Borgman 2017). Les acteurs développent ainsi des stratégies sélectives : partage différé après publication, limitation volontaire des métadonnées, choix d'entrepôts moins exigeants en termes de standardisation. Ce constat rejoint les analyses de Tenopir et al. (2020), qui montrent que la culture du partage ne s'impose pas uniquement par des moyens techniques ou juridiques, mais par des changements organisationnels profonds et par la reconnaissance symbolique et matérielle de cette activité.

Dans le domaine océanographique, plusieurs spécificités expliquent l'hétérogénéité observée. La nature des données, leur complexité technique, les temporalités de traitement, et l'investissement humain nécessaire varient fortement selon qu'il s'agisse de paramètres physiques automatisés, de mesures chimiques nécessitant plusieurs semaines de traitement, ou de données biologiques. Cette diversité, également relevée dans les travaux de Caruso et al. (2022) et Pendleton et al. (2019), crée des écarts dans la perception de l'effort de mise à disposition et de l'intérêt du partage. Les programmes où l'interdépendance scientifique est forte, comme Argo, tendent à produire une culture plus collaborative et un partage quasi immédiat, confirmant l'importance du bénéfice collectif identifié comme moteur de l'ouverture des données.

Les résultats montrent également que les dispositifs techniques (SISMER, SEANOE, Sextant, etc.) bien qu'indispensables, ne suffisent pas à garantir un partage conforme aux principes FAIR. Ce point rejoint l'état de l'art (Wilkinson et al. 2016) qui souligne que la simple existence d'infrastructures ne produit pas l'interopérabilité ni la réutilisation si les données ne sont pas correctement décrites. L'absence d'emploi-type clairement identifié pour les gestionnaires de données dans la fonction publique scientifique française complique par ailleurs le recrutement et la pérennisation de ces compétences, pourtant stratégiques.

Enfin, la question de la formation apparaît comme un enjeu transversal. Si quelques initiatives locales ou portées par ODATIS existent, elles restent ponctuelles et reposent sur des individus moteurs. Or, comme le rappellent Mons et al. (2017, 2020), l'intégration des bonnes pratiques de gestion de données dans la formation initiale et continue des chercheurs est un levier essentiel pour transformer durablement les pratiques.

Au regard de ces constats, plusieurs pistes de recommandations se dégagent :

- a. Renforcer les ressources humaines et l'intermédiation
 - i. Créer des postes pérennes de gestionnaires de données, à double compétence scientifique et données.
 - ii. Nommer une personne référente « données » par unité ou par équipe pour servir de relais local avec les centres de données nationaux (SISMER, SEANOE), dans le cas où un poste dédié n'existerait pas au sein du laboratoire.
 - iii. Favoriser l'intermédiation entre centres de données et laboratoires, via des postes de liaison chargés d'assurer la traduction entre exigences techniques, standards internationaux et pratiques scientifiques quotidiennes.
- b. Institutionnaliser la formation et l'acculturation
 - i. Rendre obligatoire une formation initiale à la science ouverte pour tout nouvel entrant, sur le modèle des formations obligatoires à l'hygiène et à la sécurité.
 - ii. Développer les ateliers courts, ciblés et coanimés par chercheurs et gestionnaires pour diffuser les bonnes pratiques.
- c. Intégrer la gestion des données dès la conception des projets
 - i. Exiger un plan de gestion des données dès la soumission et le mettre à jour à chaque étape clé.
 - ii. Allouer systématiquement une part du budget (p. ex. 10–15 %) à la préparation, la documentation et la diffusion des données (moyens technique et/ou humains).
- d. Améliorer l'accompagnement et les outils
 - i. Mettre en place un « guichet unique » national où les chercheurs peuvent poser leurs questions sur les formats, métadonnées, droits et dépôts (par exemple en s'appuyant sur ODATIS).
 - ii. Diffuser des modèles de fichiers préformatés conformes aux vocabulaires contrôlés et standards internationaux.
- e. Valoriser et reconnaître les contributions
 - i. Inclure la production et la mise à disposition de jeux de données dans les critères d'évaluation interne (rapports annuels, promotions, primes).
 - ii. Renforcer la reconnaissance du travail de gestion des données dans l'évaluation des carrières, en intégrant explicitement la production de jeux

- de données FAIR comme critère de valorisation, au même titre qu'une publication scientifique.
- iii. Encourager les projets exploitant les données internes pour en démontrer la valeur ajoutée.

f. Clarifier les incitations et les contraintes

- i. Sensibiliser largement les professionnels de la recherche.
- ii. Fixer des délais clairs de diffusion (p. ex. 2 ans après acquisition sauf dérogation).
- iii. Conditionner certains financements au respect des principes FAIR.

Cette étude repose sur un corpus d'entretiens qualitatifs et sur l'analyse de textes institutionnels, ce qui permet d'identifier des tendances et des logiques d'action, mais ne prétend pas à l'exhaustivité ni à la représentativité statistique. Les propos recueillis reflètent des contextes, disciplines et trajectoires spécifiques au sein de la recherche océanographique française. Les recommandations proposées ne constituent pas un programme prescriptif, mais des pistes issues des observations et analyses menées. Elles visent à nourrir la réflexion et le débat au sein des communautés concernées, en laissant aux acteurs institutionnels et professionnels le soin d'en évaluer la faisabilité, la pertinence et les modalités d'adaptation à leurs environnements.

En définitive, l'appropriation des injonctions de science ouverte dans les laboratoires océanographiques français se situe à un carrefour : les discours généraux percolent jusqu'aux personnes concernées, les outils techniques existent, mais les freins humains, organisationnels et culturels restent puissants. Comme l'ont montré d'autres secteurs scientifiques, la bascule vers une ouverture effective et généralisée ne se fera pas par la seule bonne volonté des acteurs, mais par un ensemble articulé d'incitations, de contraintes, de moyens et de reconnaissance. C'est à cette condition que la gestion des données pourra pleinement contribuer à l'avancée collective des connaissances sur les océans et à une meilleure capacité d'action face aux enjeux environnementaux majeurs.

CORPUS

CNRS. 2019. « Feuille de route du CNRS pour la science ouverte ».

CNRS. 2020. « Plan Données de la Recherche du CNRS ».

Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. 2021.
« 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. 2022.
« Partager les données liées aux publications scientifiques - Guide pour les chercheurs ».

Ouvrir la science. « Passeport pour la science ouverte - Sciences ouverte - Données de la recherche ».

BIBLIOGRAPHIE³⁴

- Akrich, Madeleine. 1998. « Les utilisateurs, acteurs de l'innovation ». *Sociologie de la traduction. Textes fondateurs* 253-65.
- Anichini, Giulia. 2017. « Le côté « obscur » de la science ». *Revue d'anthropologie des connaissances* 11(1). doi:10.3917/rac.034.0065.
- Austin, Claire C., Theodora Bloom, Sünje Dallmeier-Tiessen, Varsha K. Khodiyar, Fiona Murphy, Amy Nurnberger, Lisa Raymond, Martina Stockhause, Jonathan Tedds, Mary Vardigan, et Angus Whyte. 2017. « Key Components of Data Publishing: Using Current Best Practices to Develop a Reference Model for Data Publishing ». *International Journal on Digital Libraries* 18(2):77-92. doi:10.1007/s00799-016-0178-2.
- Blanquart, Marie, Thomas Descous, et Ewen Huet. 2017. « Production et partage des données : vers un droit universel ? » P. 15-30 in *La fabrique numérique des connaissances : Production et valorisation des résultats scientifiques*. ISTE Group.
- Borgman, Christine L. 2017. *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press.
- Caruso, Francesco, Pietro Tedesco, Gerardo Della Sala, Fortunato Palma Esposito, Marco Signore, Simonepietro Canese, Teresa Romeo, Marco Borra, Claudia Gili, et Donatella de Pascale. 2022. « Science and Dissemination for the UN Ocean Decade Outcomes: Current Trends and Future Perspectives ». *Frontiers in Marine Science* 9. doi:10.3389/fmars.2022.863647.
- Claustre, Hervé, Kenneth S. Johnson, et Yuichiro Takeshita. 2020. « Observing the Global Ocean with Biogeochemical-Argo ». *Annual Review of Marine Science* 12(1):23-48. doi:10.1146/annurev-marine-010419-010956.
- CNRS. 2017. « Le travail de la science et le numérique : Données, plateformes, publications ».
- CNRS - Direction de l'Information Scientifique et Technique. 2016. *Livre blanc — Une Science ouverte dans une République numérique*. Laboratoire d'idées. Marseille: OpenEdition Press.
- Collin, Lucile. 2017. « De l'auteur au valorisateur ». P. 57-64 in *La fabrique numérique des connaissances : Production et valorisation des résultats scientifiques*. ISTE Group.
- De Certeau, Michel. 1980. « L'invention du quotidien. 1 ». *Arts de faire, Ed du seuil, Paris*.

³⁴ Bibliographie présentée selon le style *American Sociological Association (ASA) 6th/7th edition*.

- Denis, Jérôme, et Samuel Goëta. 2013. « La fabrique des données brutes. Le travail en coulisses de l'open data ».
- Denis, Jérôme, et Samuel Goëta. 2016. « “Brutification” et instauration des données. La fabrique attentionnée de l'open data ». in *i3 Working Papers Series*.
- Goëta, Samuel. 2016. « Instauration des données, instauration des publics : une enquête sociologique dans les coulisses de l'open data ». phdthesis, Télécom ParisTech.
- Heaton, Lorna, et Florence Millerand. 2013. « La mise en base de données de matériaux de recherche en botanique et en écologie ». *Revue d'anthropologie des connaissances* 7(4). doi:10.3917/rac.021.0885.
- Inist-CNRS. 2022. « DoRANum-Plan de gestion de données : Définition(s) des données de recherche ».
- Jaton, Florian, et Dominique Vinck. 2016. « Processus frictionnels de mises en bases de données ». *Revue d'anthropologie des connaissances* 104(4):489-504. doi:10.3917/rac.033.0489.
- Jouët, Josiane. 2000. « Retour critique sur la sociologie des usages ». *Réseaux. Communication - Technologie - Société* 18(100):487-521. doi:10.3406/reso.2000.2235.
- Larroche, Valérie. 2018. *Le Dispositif*. London: ISTE Editions Ltd.
- Leonelli, Sabina. 2013. « Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology ». *Bulletin of Science, Technology & Society* 33(1-2):6-11. doi:10.1177/0270467613496768.
- Leonelli, Sabina. 2015. « What Counts as Scientific Data? A Relational Framework ». *Philosophy of Science* 82(5):810-21. doi:10.1086/684083.
- Martín Míguez, Belén, Antonio Novellino, Matteo Vinci, Simon Claus, Jan-Bart Calewaert, Henry Vallius, Thierry Schmitt, Alessandro Pititto, Alessandra Giorgetti, Natalie Askew, Sissy Iona, Dick Schaap, Nadia Pinardi, Quillon Harpham, Belinda J. Kater, Jacques Populus, Jun She, Atanas Vasilev Palazov, Oonagh McMeel, Paula Oset, Dan Lear, Giuseppe M. R. Manzella, Patrick Gorringer, Simona Simoncelli, Kate Larkin, Neil Holdsworth, Christos Dimitrios Arvanitidis, Maria Eugenia Molina Jack, Maria del Mar Chaves Montero, Peter M. J. Herman, et Francisco Hernandez. 2019. « The European Marine Observation and Data Network (EMODnet): Visions and Roles of the Gateway to Marine Data in Europe ». *Frontiers in Marine Science* 6. doi:10.3389/fmars.2019.00313.
- Merceur, Frédéric, Loic Petit De La Villeon, et Sybille Van Iseghem. 2023. « SEANOE, un entrepôt thématique ». P. 87-106 in *Partage et valorisation des données de la recherche*. ISTE Group.
- Millerand, Florence. 2011. « Le partage des données scientifiques à l'ère de l'e-science : l'instrumentation des pratiques au sein d'un collectif multidisciplinaire ». *Terrains & travaux* 18(1):215-37. doi:10.3917/tt.018.0215.

- Millerand, Florence. 2012. « La science en réseau : Les gestionnaires d'information « invisibles » dans la production d'une base de données scientifiques ». *Revue d'anthropologie des connaissances* 61(1):163-90. doi:10.3917/rac.015.0201.
- Millerand, Florence, et Karen S. Baker. 2011. « Ocean Informatics Monograph. Ocean Informatics Initiative: an Ethnographic Study (2002-2006). Part 1: Report. » <https://escholarship.org/uc/item/67x0f1g7>.
- Millerand, Florence, et Karen S. Baker. 2020. « Data Infrastructures in Ecology: An Infrastructure Studies Perspective ». in *Oxford Research Encyclopedia of Environmental Science*. Oxford University Press.
- Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. 2021. « Deuxième Plan national pour la science ouverte. Généraliser la science ouverte en France 2021-2024 ».
- Mons, Barend, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos, et Mark D. Wilkinson. 2017. « Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud ». *Information Services & Use* 37(1):49-56. doi:10.3233/ISU-170824.
- Mons, Barend, Erik Schultes, Fenghong Liu, et Annika Jacobsen. 2020. « The FAIR Principles: First Generation Implementation Choices and Challenges ». *Data Intelligence* 2(1-2):1-9. doi:10.1162/dint_e_00023.
- Pendleton, Linwood H., Stella J. Alexandroff, Alison Clausen, Jörn O. Schmidt, et Howard I. Browman. 2023. « Co-designing marine science for the ocean we want ». *ICES Journal of Marine Science* 80(2):342-46. doi:10.1093/icesjms/fsad018.
- Pendleton, Linwood H., Hawthorne Beyer, Estradivari, Susan O. Grose, Ove Hoegh-Guldberg, Denis B. Karcher, Emma Kennedy, Lyndon Llewellyn, Cecile Nys, Aurélie Shapiro, Rahul Jain, Katarzyna Kuc, Terry Leatherland, Kira O'Hainnin, Guillermo Olmedo, Lynette Seow, et Mick Tarsel. 2019. « Disrupting data sharing for a healthier ocean ». *ICES Journal of Marine Science* 76(6):1415-23. doi:10.1093/icesjms/fsz068.
- Proulx, Serge. 2005. « Penser les usages des TIC aujourd'hui: enjeux, modèles, tendances ». *Enjeux et usages des TIC: aspects sociaux et culturels* 1:7-20.
- Rebouillat, Violaine, et Joachim Schöpfel. 2023. « Le dispositif d'entrepôt de données de recherche ». P. 7-37 in *Partage et valorisation des données de la recherche*. ISTE Group.
- Richou, Louki-Géronimo, et Joachim Schöpfel. 2023. « L'option d'un entrepôt national ». P. 127-45 in *Partage et valorisation des données de la recherche*. ISTE Group.
- Robin, Agnès. 2022. *Droit des données de la recherche: science ouverte, innovation, données publiques*. Création information communication. Bruxelles: Larcier.
- Roemmich, Dean, Matthew H. Alford, Hervé Claustre, Kenneth Johnson, Brian King, James Moum, Peter Oke, W. Brechner Owens, Sylvie Pouliquen, Sarah Purkey,

- Megan Scanderbeg, Toshio Suga, Susan Wijffels, Nathalie Zilberman, Dorothee Bakker, Molly Baringer, Mathieu Belbeoch, Henry C. Bittig, Emmanuel Boss, Paulo Calil, Fiona Carse, Thierry Carval, Fei Chai, Diarmuid Ó. Conchubhair, Fabrizio d’Ortenzio, Giorgio Dall’Olmo, Damien Desbruyeres, Katja Fennel, Ilker Fer, Raffaele Ferrari, Gael Forget, Howard Freeland, Tetsuichi Fujiki, Marion Gehlen, Blair Greenan, Robert Hallberg, Toshiyuki Hibiya, Shigeki Hosoda, Steven Jayne, Markus Jochum, Gregory C. Johnson, KiRyong Kang, Nicolas Kolodziejczyk, Arne Körtzinger, Pierre-Yves Le Traon, Yueng-Djern Lenn, Guillaume Maze, Kjell Arne Mork, Tamaryn Morris, Takeyoshi Nagai, Jonathan Nash, Alberto Naveira Garabato, Are Olsen, Rama Rao Patabhi, Satya Prakash, Stephen Riser, Catherine Schmechtig, Claudia Schmid, Emily Shroyer, Andreas Sterl, Philip Sutton, Lynne Talley, Toste Tanhua, Virginie Thierry, Sandy Thomalla, John Toole, Ariel Troisi, Thomas W. Trull, Jon Turton, Pedro Joaquin Velez-Belchi, Waldemar Walczowski, Haili Wang, Rik Wanninkhof, Amy F. Waterhouse, Stephanie Waterman, Andrew Watson, Cara Wilson, Annie P. S. Wong, Jianping Xu, et Ichiro Yasuda. 2019. « On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array ». *Frontiers in Marine Science* 6:439. doi:10.3389/fmars.2019.00439.
- Roemmich, Dean, W. Stanley Wilson, W. John Gould, W. Brechner Owens, Pierre-Yves Le Traon, Howard J. Freeland, Brian A. King, Susan Wijffels, Philip J. H. Sutton, et Nathalie Zilberman. 2022. « The Argo Program ». P. 53-69 in *Partnerships in Marine Research*. Elsevier.
- Rücknagel, Jessika, Paul Vierkant, Robert Ulrich, Gabriele Kloska, Edeltraud Schnepf, David Fichtmüller, Evelyn Reuter, Angelika Semrau, Maxi Kindling, H. Pampel, Michael Witt, Florian Fritze, Stephanie Van De Sandt, Jens Klump, Hans-Jürgen Goebelbecker, Michael Skarupianski, Roland Bertelmann, Peter Schirmbacher, Frank Scholze, Claudia Kramer, Claudio Fuchs, Shaked Spier, et Agnes Kirchhoff. 2015. *Metadata Schema for the Description of Research Data Repositories*. Deutsches GeoForschungsZentrum GFZ. doi:10.2312/RE3.008.
- Schaap, Dick M. A., et Roy K. Lowry. 2010. « SeaDataNet – Pan-European infrastructure for marine and ocean data management: unified access to distributed data sets ». *International Journal of Digital Earth* 3(sup1):50-69. doi:10.1080/17538941003660974.
- Schöpfel, Joachim. 2023. « Enjeux et perspectives des entrepôts de données de recherche ». P. 231-50 in *Partage et valorisation des données de la recherche*. ISTE Group.
- Schöpfel, Joachim, Eric Kergosien, et Hélène Prost. 2017. « “ Pour commencer, pourriez-vous définir 'données de la recherche' ? ” Une tentative de réponse ». in *Atelier VADOR : Valorisation et Analyse des Données de la Recherche; INFORSID 2017*. Toulouse, France.
- Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, et Mike Frame. 2011. « Data Sharing by Scientists: Practices and Perceptions » édité par C. Neylon. *PLoS ONE* 6(6):e21101. doi:10.1371/journal.pone.0021101.

- Tenopir, Carol, Natalie M. Rice, Suzie Allard, Lynn Baird, Josh Borycz, Lisa Christian, Bruce Grant, Robert Olendorf, et Robert J. Sandusky. 2020. « Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide » édité par S. Lozano. *PLOS ONE* 15(3):e0229003. doi:10.1371/journal.pone.0229003.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino Da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 'T Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene Van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan Van Der Lei, Erik Van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, et Barend Mons. 2016. « The FAIR Guiding Principles for scientific data management and stewardship ». *Scientific Data* 3(1):160018. doi:10.1038/sdata.2016.18.
- Zhu, Mingting, Weibin Zhang, et Chongli Xu. 2024. « Ethical governance and implementation paths for global marine science data sharing ». *Frontiers in Marine Science* 11. doi:10.3389/fmars.2024.1421252.

ANNEXES

Table des annexes

ANNEXE 1 : GRILLE D'ENTRETIEN	69
ANNEXE 2 : GRILLE D'ANALYSE DU CORPUS TEXTUEL	71

ANNEXE 1 : GRILLE D'ENTRETIEN

Introduction et parcours
Quel est votre rôle/fonction et vos missions dans votre laboratoire ?
Depuis combien de temps travaillez-vous dans ce domaine ?
Avez-vous une formation spécifique à la gestion et au partage de données ?
Données en océanographie : définitions et pratiques
Quelles sont les principales données que vous produisez, utilisez ou gérez ?
Quels types de formats ou standards utilisez-vous et pour quoi faire : produire les données / encoder les données ? partager ?
Pouvez-vous décrire les différentes étapes de la gestion des données dans vos projets ?
Infrastructures et outils de stockage / partage
Où et comment stockez-vous ou conservez-vous vos données ?
Utilisez-vous pour les faire des outils (portails ou infrastructures, plateformes, etc.) spécifiques ?
Ces outils permettent-ils de rendre vos données disponibles pour les autres ?
Ces infrastructures et lieu de stockage sont-ils intégrés dans vos habitudes de travail ?
Science ouverte et principes FAIR
Êtes-vous familiers avec les principes dits FAIR ?
Intégrez-vous ces principes dans votre travail ?
Connaissez-vous des documents de votre laboratoire / institutionnels sur ce sujet ? Vous ont-ils servi ?
Avez-vous suivi des formations ou reçu un accompagnement sur ces questions ?

Cadres juridiques et institutionnels
Avez-vous connaissance des droits et obligations associés aux données de recherche ?
Existe-il une politique institutionnelle à ce sujet dans votre laboratoire ? Connaissez-vous une personne ressource s'occupant de ces sujets ?
Freins, tensions et arbitrages
Quels sont, selon vous, les principaux freins au partage des données ? Dans le cadre de votre travail, mais aussi de votre laboratoire, et dans votre domaine ?
Comment percevez-vous les obligations institutionnelles en matière de science ouverte ?
Dimensions sociales et culturelles
Avez-vous observé un impact sur la gestion des données dans le cadre de projets collaboratifs ? (inter-équipes, inter-labos, etc.) ?
Observez-vous des différences de pratiques entre disciplines ?
Perspectives et évolutions
Comment percevez-vous l'évolution des pratiques de gestion de données ?
Quelles améliorations seraient nécessaires pour favoriser de telles pratiques ?
Quel type d'accompagnement vous serait utile pour développer ou renforcer ces pratiques ?
Clôture
Souhaitez-vous être tenu au courant des résultats de cette enquête ?
Y a-t-il des éléments que vous souhaiteriez ajouter / rétracter ?

ANNEXE 2 : GRILLE D'ANALYSE DU CORPUS TEXTUEL

Document - date	Thème	Citation	Type d'injonction	Formulation	Commentaires
CNRS - Feuille de route - 2019	Données de la recherche	« Les données doivent être aussi ouvertes que possible, et fermées autant que nécessaire. »	Forte	Obligation implicite	Principe fondamental de la politique CNRS alignée sur les standards FAIR.
	Plans de gestion des données (DMP)	« Les DMP sont demandés lors de la rédaction des projets de recherche par la commission européenne et [...] l'ANR. »	Forte	Obligation administrative	Exigence de plus en plus institutionnalisée au niveau national et européen.
	Culture du partage des données	« Il est essentiel de développer et de partager des bonnes pratiques (FAIR) entre acteurs de la recherche. »	Forte	Recommandation forte	Enjeu majeur de sensibilisation, variable selon disciplines et maturité institutionnelle.
	Infrastructures de données	« Le CNRS soutiendra le développement de nouveaux réservoirs et services de données thématiques. »	Forte	Engagement institutionnel	Politique de soutien conditionnée à des critères d'évaluation (impact, qualité, certification).
	Diffusion & entrepôts	« Plus généralement le CNRS encourage la diffusion de toutes données structurées par leur dépôt dans des réservoirs de données thématiques ou généralistes, éventuellement accompagné de la publication de data papers. »	Moyenne	Recommandation	Appel à utiliser les réservoirs et les data papers plutôt que les entrepôts des éditeurs.
	Dépôt conjoint publication/données	« Les données sur lesquelles s'appuient les publications doivent être rendues accessibles et réutilisables dès que possible. »	Forte	Obligation encouragée	Recommandation à forte incitation, appuyée sur standards internationaux.

	Référentiels et outils FAIR	« Un réseau de personnes ressources, responsables de données dans les structures de recherche et les projets, formées à l'utilisation de services particuliers. »	Faible	Incitation	Reconnaissance émergente du rôle de data stewards, enjeu RH.
	Répertoire des services/dépôts	« Créer et afficher un répertoire des dépôts et des services de données dont le CNRS est responsable. »	Implicite	Recommandation	Objectif de visibilité institutionnelle dans le cadre des contributions à la science ouverte.
CNRS - Plan Données – 2020	Exception à l'ouverture	« Ouvert autant que possible, fermé autant que nécessaire »	Forte (par défaut)	Obligation conditionnelle	L'ouverture est la norme, mais sous réserve de confidentialité, concurrence, souveraineté, etc.
	Cycle de vie des données	« création ou collecte, traitement, analyse, préservation, partage, réutilisation »	Faible	Recommandation	Cadre général visant à une gestion complète des données, sans obligation directe.
	FAIRisation	« promouvoir une véritable “culture de la donnée” »	Forte	Incitation / orientation forte	Nécessité de transformation culturelle. Ambiguïté : rôle concret du CNRS vs. attentes des chercheurs.
	Stockage & archivage	« La question du stockage [...] est différente de celle de l'archivage »	Faible	Recommandation technique	Besoin de clarifier pratiques et outils. Tension : diversité d'interprétations selon communautés scientifiques.
	Plans de gestion (DMP)	« L'obligation dans les appels à projets européens et dorénavant de l'ANR, de produire un plan de gestion des données [...] constitue une bonne opportunité pour inciter les chercheurs et les chercheuses à prévoir les conditions d'entreposage, de curation et de	Forte (externe)	Obligation via financeurs	Obligation imposée par tiers (ANR, Europe), pas directement par le CNRS. Le CNRS accompagne cette dynamique.

		diffusion de leurs données tout au long de leur cycle de vie »			
	Référencement & entrepôts	« La recommandation de déposer les données dans un entrepôt en accès libre. Ceci est particulièrement important avec la multiplication des incitations à déposer les données dans des entrepôts gérés par les éditeurs des journaux, etc. »	Moyenne	Recommandation explicite	Tension avec pratiques éditoriales commerciales (entrepôts d'éditeurs). Soutien à l'autonomie scientifique.
	Certification des entrepôts	« Certification des dispositifs de prise en charge des données »	Moyenne à forte	Recommandation + standardisation	Appel à un alignement sur des critères reconnus (CoreTrustSeal). Tension potentielle : capacité des structures à répondre aux exigences techniques.
	Formation	« La culture de la donnée doit être développée par la formation permanente »	Forte	Obligation implicite	Vise tous les profils (chercheurs, IT, doc, etc.). Demande croissante de formations.
	Valorisation dans les carrières	« Il faudra agir pour que les investissements en support à la FAIRisation [...] soient considérés dans l'évaluation »	Moyenne	Incitation / évolution RH	Refonte potentielle des critères d'évaluation. Enjeu stratégique fort.
	Gouvernance	« Une nouvelle Direction fonctionnelle des données ouvertes de la recherche [...] aura pour mission de proposer et d'accompagner la mise en application d'une politique et d'une stratégie pour l'ouverture des données »	Forte	Obligation organisationnelle	Centralisation à la DDOR. Rôle structurant du comité de pilotage incluant les instituts. Mise en cohérence inter-instituts.
Guide pour les chercheurs - MESRI - 2022	Partage des données liées aux publications	"Il est recommandé de ne pas confier les données à partager aux éditeurs des revues"	Faible	Recommandation	Conseil négatif avec justification technique et juridique

	Choix d'entrepôts de données	"Il est donc plutôt recommandé d'utiliser pour le partage des données des entrepôts de données institutionnels, généralistes ou disciplinaires"	Faible	Recommandation	Orientation vers des solutions institutionnelles
	Documentation des données	"Décrire les données afin de les rendre intelligibles à toute personne n'ayant pas participé à leur production constitue une étape préalable à leur diffusion."	Implicite	Nécessité présentée comme évidente	Approche pédagogique sans contrainte formelle
	Standards techniques	"Quel que soit l'entrepôt choisi pour partager les données, celui-ci se doit en particulier d'offrir les fonctionnalités suivantes"	Forte	Obligation technique	Critères impératifs pour le choix d'entrepôt
	Citation des données	"il est recommandé d'introduire une section spécifique « Disponibilité des données » avant les références bibliographiques"	Faible	Recommandation	Conseil pratique pour la rédaction
2e Plan national MESRI - 2021	Transformation durable	"faire de la science ouverte la pratique commune et partagée"	Forte	Vision stratégique	Objectif de transformation systémique
	Structuration des données	"les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR"	Forte	Obligation progressive	Mise en conformité obligatoire
	Diffusion obligatoire	"Mettre en œuvre l'obligation de diffusion des données de recherche financées sur fonds publics"	Forte	Obligation légale	Application de la loi République numérique
	Infrastructure nationale	"Créer Recherche Data Gouv, la plateforme nationale fédérée des données de la recherche"	Forte	Décision de création	Infrastructure souveraine
	Plans de gestion	"L'ANR demande l'élaboration de plans de gestion des données (DMP) pour les projets de recherche qu'elle finance"	Forte	Exigence réglementaire	Acquis du premier plan
Passeport pour la Science Ouverte - Données de la	Principe général d'ouverture	"les données de la recherche doivent être aussi ouvertes que possible et pas plus fermées que nécessaire"	Forte	Obligation légale	Principe inscrit dans le Code de la recherche
	Obligations légales	"L'article L. 533-4 du Code de la recherche [...] prévoit [...] après publication, un principe de	Forte	Obligation légale	S'applique aux recherches financées à

recherche – MESRI – 2024		libre réutilisation des données de recherche, lorsque celles-ci : sont issues d'une recherche financée à plus de 50 % par des fonds publics, ne sont pas protégées par un droit spécifique ou par une réglementation, ont été rendues publiques par l'établissement ou l'organisme de recherche."			plus de 50% par fonds publics
	Plan de gestion des données	"le plan de gestion de données ou PGD est un outil qui permet de décrire leur gestion"	Faible	Recommandation/outil	Souvent imposé par les financeurs
	Principes FAIR	"Une bonne gestion de ses données vise à les rendre Faciles à trouver, Accessibles, compréhensibles par les humains et les machines, c'est-à-dire Interopérables, et Réutilisables"	Forte	Recommandation normative	Standard international adopté par les politiques publiques
	Licences de diffusion	"il est fortement recommandé d'y apposer une licence afin de définir comment celles-ci peuvent être réutilisées"	Faible	Recommandation	Liste de licences autorisées pour les administrations
	Entrepôts de confiance	"Le choix de l'entrepôt est déterminant car les entrepôts sont plus ou moins compatibles avec les principes FAIR"	Faible	Recommandation	Critères de sélection définis par Science Europe
	Exceptions à l'ouverture	"La décision de maintenir des données fermées doit s'appuyer sur des motifs découlant d'autres dispositifs juridiques"	Forte	Obligation de justification	Données personnelles, secret défense, propriété intellectuelle
	Impact environnemental	"il est donc primordial [...] de permettre la réutilisation (principes FAIR) des données existantes avant de chercher à en produire de nouvelles"	Faible	Recommandation éthique	Enjeux de sobriété numérique émergents
	Accompagnement	"de nombreux experts peuvent vous accompagner à chaque étape d'un projet de recherche"	Implicite	Incitation	Réseau d'acteurs institutionnels mobilisés

TABLE DES ILLUSTRATIONS

Figure 1 : Cycle de vie d'un flotteur profileur Argo. ©Thomas Haessig.	14
Figure 2 : Cycle de vie des données de recherche (d'après Cat-OPIDoR).	18
Figure 3 : Les principes FAIR.	20

TABLE DES MATIÈRES

SIGLES ET ABREVIATIONS	7
INTRODUCTION.....	9
I. COMPRENDRE LE PAYSAGE : CONCEPTS, OUTILS ET REFERENTIELS	11
1. Définir et caractériser les données océanographiques	11
2. Stocker et diffuser : infrastructures, portail et diversité des pratiques	16
3. Les principes FAIR comme cadre du partage des données	19
II. CADRES DE REFERENCE ET OBSTACLES AU PARTAGE DES	
DONNEES DE RECHERCHE	21
1. Environnement institutionnel et juridique de la science ouverte	21
1. Orientation des acteurs institutionnels.....	21
2. Règles, obligations et zones grises du cadre juridique	22
2. Les pratiques et tensions au cœur du travail scientifique.....	23
1. Entre ouverture et rétention : une lecture socio-anthropologique	23
2. Science ouverte et carrière académique : concilier injonctions et	
intérêts individuels	25
III. METHODOLOGIE ET DEMARCHE D'ENQUETE.....	27
1. Une approche qualitative fondée sur un corpus mixte	27
1. Le corpus textuel : sources et critères de sélection	27
Feuille de route du CNRS pour la science ouverte (CNRS, 2019)	28
Plan Données de la Recherche du CNRS (CNRS, 2020)	28
Deuxième Plan national pour la science ouverte (MESRI, 2021)	29
Guide pour les chercheurs : Partager les données liées aux publications	
scientifiques (MESRI, 2022)	30
Passeport pour la science ouverte (Ouvrir la science, 2024).....	30
2. Le corpus d'entretiens : choix des enquêtés et grille d'entretien	31
2. Méthodes d'analyse : de la grille d'entretien à la confrontation avec	
la littérature	33
IV. ENTRE PRESCRIPTIONS ET REALITES : APPROPRIATION	
CONTRASTEE DES INJONCTIONS.....	34
1. Ce que disent les textes : un cadrage institutionnel de la gestion des	
données	34

2. Ce que font les acteurs : des conceptions situées et contextualisées des données	36
3. Les infrastructures au milieu des pratiques de gestion	39
4. Science ouverte en pratique : entre incitation, obligation et stratégies d'évitement	42
5. Ouvrir... mais sous conditions : arbitrages, tensions et rapports de force	45
6. Ressources humaines, organisation et leviers d'action pour renforcer la gestion des données.....	51
CONCLUSION.....	57
CORPUS.....	61
BIBLIOGRAPHIE	63
ANNEXES	68
Annexe 1 : Grille d'entretien	69
Annexe 2 : Grille d'analyse du corpus textuel	71
TABLE DES ILLUSTRATIONS	77
TABLE DES MATIÈRES	79