



HAL
open science

ÉTAT DES LIEUX DU RÉSEAU DE PLATEFORMES TECHNOLOGIQUES DATA

Coralie Morand

► **To cite this version:**

Coralie Morand. ÉTAT DES LIEUX DU RÉSEAU DE PLATEFORMES TECHNOLOGIQUES DATA. Réseau national des Maisons des Sciences de l'Homme. 2025. hal-04955196v1

HAL Id: hal-04955196

<https://hal.science/hal-04955196v1>

Submitted on 18 Feb 2025 (v1), last revised 4 Mar 2025 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉTAT DES LIEUX DU RÉSEAU DE PLATEFORMES TECHNOLOGIQUES DATA

Coralie Morand

Chargée de mission réseau des plateformes Data, RnMSH

Janvier 2025



TABLE DES MATIÈRES

INTRODUCTION	5
PRÉSENTATION DU RÉSEAU DE PLATEFORMES TECHNOLOGIQUES DATA	5
OBJECTIFS DE L'ENQUÊTE	8
MÉTHODOLOGIE D'ENQUÊTE	8
I - PRÉSENTATION GÉNÉRALE DES PLATEFORMES DATA	11
PANORAMA DES MEMBRES DU RÉSEAU DATA	11
GOUVERNANCE ET GESTION OPÉRATIONNELLE	12
DES PLATEFORMES ANCRÉES DANS LEUR TERRITOIRE	14
BUDGET	15
II - MEMBRES DU PERSONNEL QUI FONT VIVRE LE RÉSEAU DATA	17
INGÉNIEURS	17
RÉFÉRENTS SCIENTIFIQUES	20
PERSONNELS DE SOUTIEN	21
III - ACTIVITÉS ET SERVICES RENDUS À LA COMMUNAUTÉ SCIENTIFIQUE	23
UTILISATEURS	23
ACCOMPAGNER LES CHERCHEURS VERS L'UTILISATION DES DONNÉES QUANTITATIVES	25
COMMUNIQUER AUTOUR DES DONNÉES QUANTITATIVES EN SHS	27
CONSTRUIRE DES RÉSEAUX MÉTIERS ET THÉMATIQUES	29
SPÉCIALISATION THÉMATIQUE ET COLORATION LOCALE	31
OUTILS	34

IV - PARTENARIATS	39
UN PARTENAIRE PRIVILÉGIÉ : L'IR* PROGEDO	39
PROJETS IMPLIQUANT PLUSIEURS MSH	41
INSTITUTIONS ET PARTENAIRES LOCAUX	41
PARTENARIATS INTERNATIONAUX	43
V - ENJEUX ET PERSPECTIVES	44
FINANCEMENT ET PÉRENNISATION DES PLATEFORMES	44
VALORISATION DES ACTIONS MENÉES PAR LES PLATEFORMES DATA	45
CONCLUSION	47
ANNEXES	49
LISTES DES SIGLES ET ACRONYMES	50
PRÉSENTATION DES PLATEFORMES DATA	54



Figure 1 : Le réseau des plateformes Data, juin 2024.
Réalisation : C. Morand, sur canva.com

INTRODUCTION

PRÉSENTATION DU RÉSEAU DE PLATEFORMES TECHNOLOGIQUES DATA

Le RnMSH¹ définit une plateforme technologique comme « [un] ensemble de moyens techniques organisés de façon systématique qui, allié à des compétences d'ingénierie de haut niveau, constitue une instrumentation spécifique et mutualisée au service de la recherche en SHS »². Lors de la mise en réseau de ces plateformes, le RnMSH a déterminé cinq ensembles thématiques réunissant des compétences et ressources propres : Audio-Visio et Cogito³, Data, Scripto⁴ et Spatio⁵. Le réseau Data est défini comme un « corpus de données quantitatives ou données traitées de manière quantitative »⁶. Le caractère particulièrement contextuel des données en fait un concept complexe à définir ; elles s'entendent comme des « informations collectées par les scientifiques dans la perspective d'être utilisées comme preuves d'une théorie scientifique »⁷. La notion de donnée quantitative a, elle aussi, une acception variable selon les interlocuteurs. Si, pour certains, une donnée quantitative se définit comme une donnée numérique incluse dans une masse, pour d'autres il s'agit de toute donnée qui peut être traitée selon une méthode quantitative, c'est-à-dire statistique au sens très large du terme. Les plateformes Data traitent essentiellement, mais pas exclusivement, des données issues de la statistique publique, qu'il s'agisse de sondages ou de sources administratives. Ces données ont été collectées de manière organisée soit dans un but administratif, soit dans un but de recherche scientifique.

Le réseau se compose de 17 plateformes technologiques labellisées Data par le RnMSH, situées dans 16 MSH : un observatoire démographique (DemoMed) et 16 plateformes d'accompagnement à la recherche et de recherche labellisées PUD (Plateformes Universitaires de Données) par l'IR* PROGEDO. Les PUD « font partie du dispositif national de mise à disposition de la recherche des grands fichiers d'enquête produits par des organismes publics ou parapublics »⁸, tels que l'INSEE, l'INED, le CNRS, les ministères, etc. La mission du réseau Data, centrée autour des données quantitatives, est au cœur de sa particularité puisque les PUD sont des relais locaux de l'IR* PROGEDO dont l'objectif est de « développer la culture

- 1 Le RnMSH (UAR 3603) prend en charge l'animation et la coordination des 22 MSH qui en sont membres.
- 2 Vendrix P. et Wolikow S., 2016, Le Réseau national des Maisons des sciences de l'Homme : une infrastructure de recherche au service des SHS, Paris : RnMSH.
- 3 Boisseleau L. et Perret C., 2022, « État des lieux de deux réseaux des plateformes technologiques Audio-Visio et Cogito », CNRS, RnMSH, <https://hal.science/hal-04598352>.
- 4 Dupuis L., 2024, « État des lieux du réseau des plateformes technologiques Scripto », CNRS, RnMSH, <https://rnmsh.hal.science/hal-04598669>.
- 5 Grelier A., 2024, « État des lieux du réseau des plateformes technologiques Spatio », CNRS, RnMSH, <https://rnmsh.hal.science/hal-04716023v1>.
- 6 RnMSH, « Data », <https://www.msh-reseau.fr/plateformes#data>.
- 7 Rebouillat V., 2019, « Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs », thèse de doctorat en sciences de l'information et de la communication, Paris : CNAM.
- 8 Marie M. et Niaré A., 2022, « Plateforme Universitaire de Données de Caen. Rapport d'activité 2021-2022 ».

des données d'enquêtes pour la recherche en sciences sociales»⁹. La culture des données est ici entendue comme un ensemble de bonnes pratiques en matière d'acquisition, utilisation, partage et réutilisation des données pour la recherche scientifique¹⁰. Cette implantation massive des PUD au sein des MSH permet à l'IR* PROGEDO de bénéficier d'un maillage territorial conséquent et de favoriser l'usage FAIR¹¹ des données quantitatives dans un large éventail de laboratoires SHS. Les objectifs des plateformes Data sont, d'une part, de soutenir les utilisateurs de données quantitatives ; d'autre part, pour les PUD, de stimuler l'utilisation des données disponibles dans le catalogue Quetelet-Progedo-Diffusion¹².

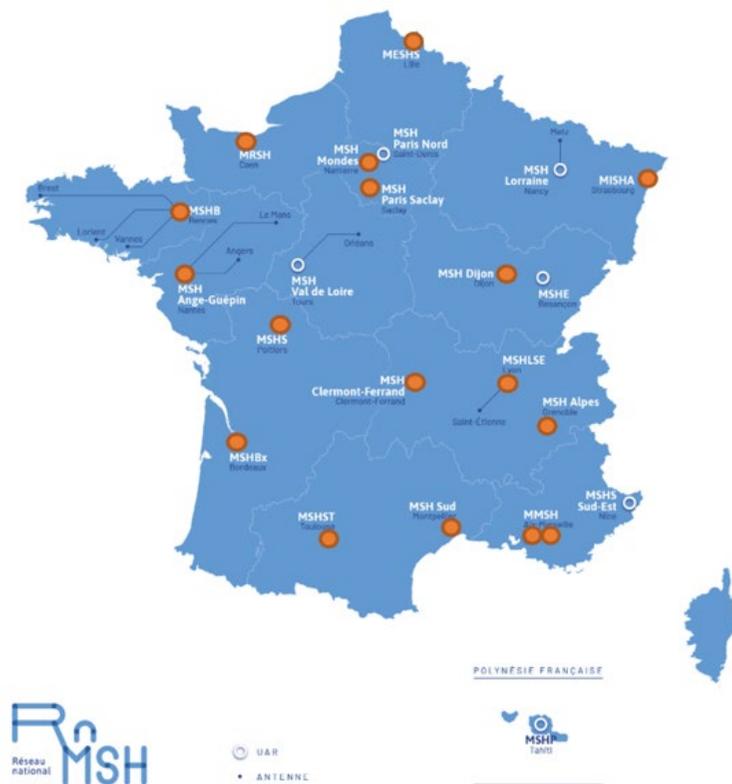


Figure 2 : Répartition géographique des plateformes Data, janvier 2024.

⁹ PROGEDO, « Organisation et gouvernance », <https://www.progedo.fr/progedo/organisation-et-gouvernance/>.
¹⁰ MSH Paris-Saclay, « Plateforme Universitaire de Données Paris-Saclay », <https://msh-paris-saclay.fr/pud-ps/>.
¹¹ Ensemble de principes directeurs pour gérer les données de la recherche et visant à les rendre Faciles à trouver, Accessibles, Interopérables et Réutilisables (INIST, Service Formation DoRANum, « Principes FAIR et cycle de vie des données », 15.02.024, https://www.inist.fr/wp-content/uploads/2024/02/Webinaire_Principes-FAIR-Cycle-de-vie_2024-03-19.pdf).
¹² Le catalogue Quetelet-Progedo-Diffusion est l'entrepôt et catalogue de données de l'IR* PROGEDO, financé par le MESR et le CNRS, en collaboration avec l'INED. Il propose plus de 1 600 jeux de données.

Le réseau Data est au cœur de la définition première de ce que sont les Maisons des Sciences de l'Homme et de la « doctrine des trois I »¹³, soit des lieux dans lesquels les SHS développent l'interdisciplinarité, la coopération institutionnelle et l'ouverture internationale¹⁴. Ces plateformes sont en effet des espaces de soutien à la recherche en SHS toutes disciplines confondues ; elles sont au cœur de partenariats avec des institutions diverses. Certaines s'inscrivent dans une stratégie d'ouverture internationale par leur inscription dans des réseaux de recherche internationaux. Par leurs activités d'appui aux chercheurs et de recherche, elles représentent l'idéal de travail collaboratif qui sous-tendait, dès les années 1960, la création de la première Maison des sciences de l'Homme¹⁵. Enfin, du fait de leur vocation à accompagner la recherche scientifique au niveau local, elles répondent également à deux axes non théorisés, mais structurants ces dix dernières années au sein des MSH : une identité scientifique propre et une inscription territoriale affirmée.

Début 2024, au lancement effectif de l'enquête, 17 plateformes technologiques hébergées par des MSH s'inscrivaient dans la définition du réseau Data, proposant des services mutualisés d'accès aux données quantitatives et d'accompagnement à leur traitement. Suite à la campagne de labellisation des plateformes technologiques du RnMSH fin 2023, 14 plateformes ont été labellisées ou relabellisées Data, trois plateformes sont en projet de labellisation. Les plateformes Data (hors DemoMed) sont intégrées au réseau des PUD de l'IR* PROGEDO, ce qui en fait un réseau particulièrement dynamique et solidaire du fait, notamment, des Journées Inter-PUD, semestrielles, qui permettent aux membres de ce réseau de se connaître. Les PUD disposent également d'une liste de diffusion sur laquelle sont mutualisées des compétences et connaissances.

Plateformes technologiques labellisées Data		
DemoMed	Observatoire démographique de la Méditerranée	Maison Méditerranéenne des Sciences de l'Homme (MMSH)
PANELS	Plateforme technologique pour les sciences sociales	Maison des Sciences de l'Homme Lyon Saint-Étienne (MSHLSE)
Progedo-Loire		Maison des Sciences de l'Homme Ange-Guépin
PUD-AMU	Plateforme Universitaire de Données d'Aix-Marseille Université	MMSH
PUD-Bx	Plateforme Universitaire de Données de Bordeaux	Maison des Sciences de l'Homme de Bordeaux (MSHBx)
PUDC	Plateforme Universitaire de Données de Caen	Maison de la recherche en Sciences Humaines (MRSH)

¹³ Guibentif P., 2020 [2019], « Les principes fondateurs du Réseau national des Maisons des Sciences de l'Homme », in Commaille J. (dir.), *Avenir de la recherche et Maisons des Sciences de l'Homme*, Paris : Alliance Athéna.
¹⁴ Bentaboulet M., 2020 [2019], « Histoire d'une structuration de la recherche : de la Maison au Réseau », in Commaille J. (dir.), *Ibidem*.
¹⁵ *Ibidem*.

PUD-CA	Plateforme Universitaire de Données de Clermont Auvergne	MSH Clermont-Ferrand
PUD-D	Plateforme Universitaire de Données de Dijon	MSH Dijon
PUD-GA	Plateforme Universitaire de Données de Grenoble-Alpes	MSH Alpes
PUD-L	Plateforme Universitaire de Données de Lille	Maison Européenne des Sciences de l'Homme et de la Société (MESHS)
PUD-N	Plateforme Universitaire de Données de Nanterre	MSH Mondes
PUD-S	Plateforme Universitaire de Données de Strasbourg	Maison Interuniversitaire des Sciences de l'Homme - Alsace (MISHA)
PUD-T	Plateforme Universitaire de Données de Toulouse	Maison des Sciences de l'Homme et de la Société de Toulouse (MSHS-T)
PUD-UP	Plateforme Universitaire de Données de l'Université de Poitiers	Maison des Sciences de l'Homme et de la Société (MSHS)
Plateformes technologiques en projet de labellisation Data		
Montpellier		MSH SUD
PUD-B	Plateforme Universitaire de Données de Bretagne	Maison des Sciences de l'Homme en Bretagne (MSHB)
PUD-PS	Plateforme Universitaire de Données de Paris-Saclay	MSH Paris Saclay

Figure 3 : Liste des membres du réseau Data, janvier 2024.

OBJECTIFS DE L'ENQUÊTE

Cette enquête, financée par le MESR, s'inscrit dans un projet de valorisation des réseaux de plateformes du RnMSH. Du fait du rôle stratégique que jouent les données quantitatives au niveau national, elle vise également à repenser la collaboration entre le RnMSH, l'IR* PROGEDO et l'IR* Huma-Num afin d'accompagner la montée en compétences des PUD qui émaillent le territoire. Enfin, elle permet de mettre en lumière la grande diversité des plateformes Data, reflet de leur bonne implantation dans leurs environnements scientifiques respectifs.

MÉTHODOLOGIE D'ENQUÊTE

La méthodologie appliquée est une réadaptation des quatre enquêtes précédentes, portant sur les autres réseaux de plateformes du RnMSH. Entre janvier et mars 2024, un formulaire, inspiré des grilles d'entretien réalisées pour ces enquêtes, a été transmis aux personnes en charge des plateformes technologiques Data, permettant une première collecte de données comparables entre elles. La PUD de Montpellier ayant été créée en janvier 2024

elle n'a pas été intégrée à cette première phase d'enquête. Une recherche bibliographique a été menée en parallèle avec pour corpus les dossiers de demande de labellisation 2023 du RnMSH, les rapports d'activité des plateformes Data et de l'IR* PROGEDO et diverses communications scientifiques en lien avec les données quantitatives. Une restitution des premières données recueillies et analysées lors de cette étape a été réalisée en avril 2024 lors des Journées Inter-PUD organisées par la MSH Dijon. Cette première phase de l'enquête a été complétée par des visites sur sites pour rencontrer les ingénieurs dans leur espace de travail (de mars à juillet 2024). À l'instar des précédentes enquêtes, ces temps comportaient une visite de la MSH et de la PUD, la présentation aux membres du personnel de la MSH, ainsi que des entretiens avec les ingénieurs et, lorsque cela a été possible, les référents scientifiques, des membres de la direction de la MSH, des utilisateurs des services de la PUD et toute autre personne impliquée ou ayant été impliquée dans les activités de la PUD. Lorsqu'une rencontre sur site était impossible, un entretien par Zoom était organisé ; 8 entretiens ont été effectués selon cette modalité. Au total, 63 temps d'échanges formels ont été organisés, auprès de 57 personnes¹⁶, réparties sur 14 sites.

DemoMed	PANELS	PRODEGO-Loire	PUD-AMU	PUD-B	PUD-Bx	PUDC	PUD-CA
1	2	4	6	5	2	6	4

PUD-D	PUD-GA	PUD-L	PUD-N	PUD-PS	PUD-S	PUD-T	PUD-UP
2	6	3	1	1	8	7	7

Figure 4 : Nombre de personnes rencontrées pour chaque plateforme.

Membres de la direction des MSH	Référents scientifiques	Ingénieurs	Utilisateurs des services Data	Autres fonctions
16	14	18	9	8

Figure 5 : Nombre de personnes rencontrées par fonction liée aux plateformes.

L'accueil a toujours été chaleureux et bienveillant, ce qui a facilité la mise en place d'une méthode inspirée de l'anthropologie sociale : observation des relations interpersonnelles, entretiens semi-dirigés ou informels. Ainsi, contrairement à ce qui a été mis au point pour les précédentes enquêtes, il n'existait pas de grille d'entretien prédéfinie pour cette seconde phase. Des questions étaient préparées en amont de la visite, adaptées à chacune des plateformes et issues de l'analyse de l'ensemble de la documentation disponible sur cette plateforme.

En plus de ces rencontres, des entretiens ont été menés auprès de cinq personnes gravitant autour des PUD, qu'il s'agisse de responsables de services humanités numériques et données, de membres d'infrastructures de recherche en SHS ou de participants aux Ateliers de la donnée par exemple. À des fins comparatives, une visite de la PUD Grands-Moulins (Université Paris Cité) et un entretien avec son ingénieure ont été effectués en mai 2024. À ces entretiens formels doivent s'ajouter des rencontres informelles, formations et séminaires ayant nourri la réflexion sur les enjeux actuels autour des données de la recherche.

¹⁶ Certaines personnes occupent des fonctions différentes auprès des PUD et ont donc été interrogées sur ces divers statuts.

L'ensemble des matériaux ainsi collectés ont été traités et analysés au fil de l'eau sur Excel et sur la base de données relationnelles Heurist créée pour les besoins des précédentes enquêtes, permettant une rédaction du présent rapport entre juillet et octobre 2024. Enfin, les principaux interlocuteurs pour cette enquête ont été les ingénieurs des plateformes, puis les référents scientifiques. Les autres entretiens étaient plus succincts et portaient davantage sur des questions de stratégie locale voire nationale que sur les activités des plateformes.

I- PRÉSENTATION GÉNÉRALE DES PLATEFORMES DATA

La première plateforme a été créée à Lille en 2002 par le laboratoire Clersé (Centre lillois d'études et de recherches sociologiques et économiques), au sein de l'IFRESI (Institut Fédératif de Recherche sur les Économies et les Sociétés Industrielles), devenu la MESHS. La deuxième ouvre à Lyon en 2008, puis la troisième à Caen en 2011. S'ensuivra alors une réplification du modèle lillois sur la quasi-totalité du territoire universitaire français avec une implantation de PUD dans les MSH hexagonales, complétées par la PUD Grands Moulins, rattachée au SCD (Service Commun de Documentation) d'Université Paris Cité, et par le CDSP (Centre de Données Socio-Politiques) de Sciences Po Paris. Elles voient le jour à l'issue de concertations entre les tutelles des MSH, les MSH elles-mêmes et l'IR* PROGEDO. Les activités d'une PUD démarrent dès lors qu'un référent scientifique est nommé et qu'il y a un ingénieur en poste. En 2015, une campagne de labellisation du RnMSH permet de donner une visibilité à ces plateformes technologiques accueillies au sein des MSH et de constituer un nouveau réseau officiel de plateformes technologiques, inscrivant dès lors les membres du réseau Data comme parties prenantes des stratégies politiques et scientifiques des MSH. À partir de 2019, le Plan national pour la science ouverte (PNSO 2018-2021)¹⁷ et le Deuxième Plan national pour la science ouverte (PNSO 2021-2024)¹⁸ permettent une accélération de la création de nouvelles PUD pour participer à la structuration des données de la recherche et à la diffusion de la culture des données sur l'ensemble du territoire. En plus de ce réseau de PUD, le réseau Data compte sur la plateforme DemoMed, seule plateforme non PUD, observatoire démographique fondé sur un projet collectif de chercheurs internationaux. En 2023, le RnMSH lance une nouvelle campagne de labellisation auprès des directions des MSH ; l'ensemble des plateformes Data ayant déposé un dossier obtiennent leur labellisation pour les cinq prochaines années, preuve de leur pertinence scientifique et de leur bonne implantation au sein des MSH.

PANORAMA DES MEMBRES DU RÉSEAU DATA

Les plateformes Data ont en moyenne 7,6 ans, avec un âge médian de 6 ans. Il y a toutefois une forte disparité liée, notamment, à la vague de création de nouvelles plateformes à la fin des années 2010. La plateforme la plus ancienne, la PUD Lille, a été officiellement créée en 2002 alors que la plus récente est celle de Montpellier, créée en début d'année 2024. Les plateformes de ce réseau se caractérisent par un paradigme à première vue paradoxal : elles sont d'une grande diversité, tout en partageant un socle commun fort, le fait d'être des relais locaux de l'IR* PROGEDO (à l'exception de DemoMed). Cette caractéristique leur donne une identité forte avec une mission claire centrée autour du développement des compétences quantitatives de la communauté de recherche en SHS. Cette ambition se décline autour de trois axes définis comme suit sur le site internet de l'IR* :

¹⁷ MESRI, 2021, « Bilan du Plan national pour la science ouverte 2018-2021 », mai 2021, <https://www.ouvrirlascience.fr/wp-content/uploads/2021/07/Bilan-PNSO-2018-2021.pdf>.

¹⁸ MESRI, 2021, « Deuxième Plan national pour la science ouverte. Généraliser la science ouverte en France 2021-2024 », juillet 2021, <https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2021-09/2e-plan-national-pour-la-science-ouverte-12968.pdf>.

- « l'aide à la recherche, à l'acquisition et à la mise en œuvre des fichiers de données » ;
- « la formation à la compréhension, à l'utilisation et au traitement de ces données » ;
- « la formation à l'utilisation d'outils informatiques et statistiques »¹⁹.

Au sein de ces grandes branches d'activités se déclinent des identités diverses, liées aux identités locales de chaque MSH, ainsi que des caractéristiques variées en termes de missions, de ressources humaines, d'outils et matériel disponibles, de gouvernance et de fonctionnement opérationnel.

GOVERNANCE ET GESTION OPÉRATIONNELLE

La plateforme est souvent intégrée à d'autres pôles au sein de la MSH, comme c'est le cas de PANELS par exemple où la plateforme Data intègre les pôles « Statistiques » et « Science ouverte, éditions, données », permettant de mutualiser les compétences techniques et opérationnelles de chacun, tout en facilitant la lisibilité des services proposés à la communauté scientifique. Les MSH font en effet souvent le choix de regrouper les pôles données et humanités numériques au sein d'un seul et même service.

recherche accompagnés selon les compétences de chacun, sans que cela n'entraîne nécessairement de travail concret entre les deux pôles. Ce constat n'empêche cependant pas les plateformes Data d'être pleinement intégrées aux MSH et d'être identifiées comme un service des MSH auprès des communautés scientifiques locales.

La gestion opérationnelle des activités des plateformes est généralement confiée à l'ingénieur en poste, avec un soutien plus ou moins important du référent scientifique. La PUD-Université de Poitiers est une exception à ce système avec un consortium de quatre personnes (gestionnaire de la PUD-UP, responsable scientifique de la PUD-UP, responsable de la plateforme technologique de la MSHS et responsable du pôle recherche et innovation de la MSHS) qui se réunissent une fois par mois afin d'aborder l'ensemble des aspects liés à la gestion opérationnelle de la plateforme, à ses perspectives et à son ancrage scientifique local. Ce mode de gestion permet d'ancrer pleinement la plateforme Data dans un réseau de plateformes interne à la MSH, tout en mutualisant des compétences et connaissances diverses. Lorsque des binômes sont présents au sein de la plateforme, il est de coutume que la répartition des tâches se fasse de façon informelle entre les ingénieurs. La PUD-Bretagne fait exception à cette règle en proposant deux fiches de poste structurantes au sein de la plateforme : un poste de coordinatrice et un poste d'ingénieur en traitement et analyse de données. La coordinatrice prend en charge les activités structurantes de la plateforme : organisation d'activités diverses, rédaction de rapports, pilotage de réseaux thématiques et métiers, communication, etc. L'ingénieur est quant à lui responsable des accompagnements aux chercheurs, car il risque de créer une répartition inégale de la charge de travail. De plus, le fait de tenir la coordinatrice éloignée des activités de traitement et analyse de données fait courir le risque d'une perte de compétences opérationnelles, alors même que ce type de poste nécessite un niveau de qualifications élevé sur ces sujets.

La gouvernance des plateformes Data est elle aussi assez variée, et répond à des impératifs locaux divers. Si la majeure partie des décisions sont prises en collaboration entre l'ingénieur et le référent scientifique, d'autres organes, généralement consultatifs, ont parfois été créés tant pour bénéficier de compétences complémentaires que pour valoriser les actions mises en place par la plateforme auprès de publics extérieurs aux MSH. La PUD Lille par exemple, s'est dotée de deux organes consultatifs : un comité de pilotage qui rassemble des instances concernées par les activités de la PUDL et un comité d'utilisateurs qui réfléchit aux orientations futures de la plateforme, constitué des unités dont les membres participent aux activités de la PUDL. Le comité de pilotage réunit, quelques semaines après le comité des utilisateurs, la Vice-Présidente Recherche en SHS de l'Université de Lille, la direction de la MSH, la direction de l'IR* PROGEDO et le responsable du service Plateformes et grands équipements de l'Université de Lille.

Cette grande diversité est le reflet de l'histoire des plateformes et du contexte local au sein duquel elles sont implantées. Elle permet une grande flexibilité opérationnelle, mais peut également être un frein au développement d'une stratégie nationale. Au niveau supra, l'absence de formalisation des fonctions de l'IR* PROGEDO, du RnMSH et des MSH auprès des plateformes peut créer un flou auquel la charte des PUD²⁰, dans sa version actuelle, ne répond pas. Si les relations entre les différents acteurs ont historiquement été sereines, une clarification formelle des rôles de chacun des acteurs permettrait d'éviter d'éventuels conflits à

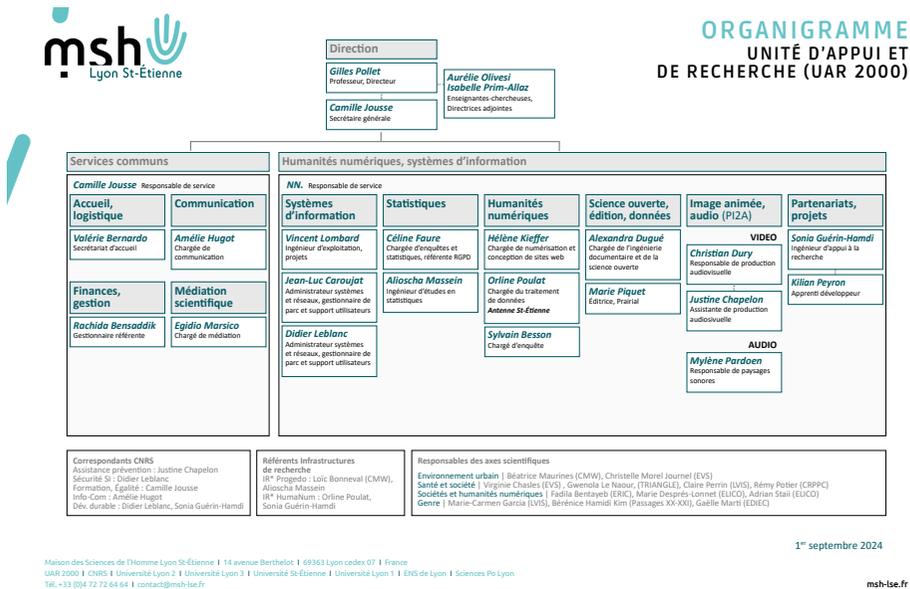


Figure 6 : Organigramme de la MSH LSE, septembre 2024.

Toutefois, de l'avis général, ces liens sont essentiellement une question d'organigramme et n'entraînent que peu de travail commun entre ces deux pôles, à l'exception de quelques projets de formations logicielles communes. D'autres ingénieurs envisagent leur collaboration avec les ingénieurs en humanités numériques comme des transferts de projets de

¹⁹ PROGEDO, « Plateformes Universitaires de Données », <https://www.progedo.fr/services/plates-formes-universitaires-de-donnees/>.

²⁰ PROGEDO, « Charte des Plateformes Universitaires de Données », https://www.progedo.fr/app/uploads/2020/07/Charte_des_PUD_2020-06-15.pdf.

venir et simplifierait le travail des ingénieurs sur le terrain en leur donnant des directives claires quant à leurs fonctions et leurs interlocuteurs aux injonctions parfois contradictoires.

DES PLATEFORMES ANCRÉES DANS LEUR TERRITOIRE

Les MSH disposent d'implantations territoriales diverses, ayant un impact réel sur l'intégration des plateformes Data au sein de la communauté de recherche locale²¹. Ainsi, lorsque la MSH se trouve au cœur du campus universitaire SHS, développer des activités au service des chercheurs est facilité comme c'est le cas pour les PUD-Université de Poitiers, PUD Grenoble-Alpes, PUD-Toulouse, PUD Paris-Saclay, PUD Clermont Auvergne ou PUD Nanterre. Il est en effet plus simple pour les chercheurs de « s'arrêter en passant », tout comme les actions de communication et de formation mises en place par la plateforme sont facilitées et atteignent plus aisément leur public. Ce type d'implantation permet de développer un réel service de proximité comme à la PUD Caen où tant les personnels de la MSH que les chercheurs vantent les mérites de cette proximité géographique avec l'ingénieur. Parfois, la MSH peut être au cœur d'un campus, mais les sites universitaires SHS peuvent être répartis à plusieurs endroits d'une ville, voire d'une région, comme cela se constate à la PUD Bretagne, à PANELS ou à PROGEDO-Loire où seuls 17 des 27 laboratoires rattachés à la MSH se trouvent à Nantes. Une implantation centrale permet également de faire de la MSH un acteur pleinement identifié de la recherche locale, particulièrement lorsque les locaux ont été pensés en ce sens. Par exemple, la MSH Dijon bénéficie d'un bâtiment récent, pensé par et pour la recherche puisqu'architectes et membres du personnel ont construit les plans ensemble. Ce type de bâtiment facilite la dynamique des projets accompagnés par les équipes ainsi que l'interconnaissance entre les membres des différents pôles.

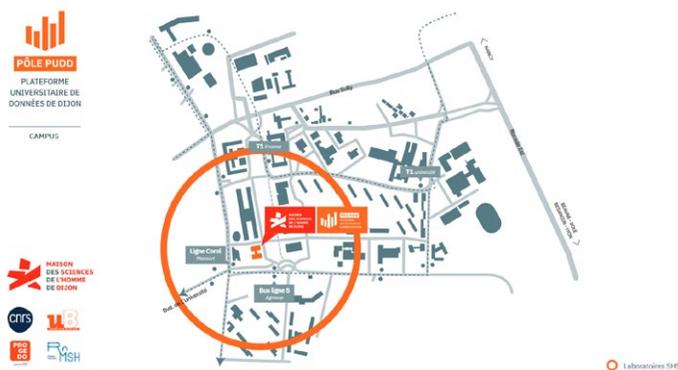


Figure 7 : Plan du campus dijonnais.

La MISHA est un autre exemple de bâtiment particulièrement adapté et dynamique du fait de sa conception autour d'un vaste hall d'entrée, conçu comme un espace de dialogue entre sciences et arts. De nombreux espaces de détente ont été aménagés à différents étages, invitant les étudiants et chercheurs à pleinement s'approprier le lieu, faisant de celui-ci un centre de vie des SHS particulièrement bien identifié sur le campus strasbourgeois.

Maison Interuniversitaire des Sciences de l'Homme-Alsace

1er étage
Situation janv. 2019

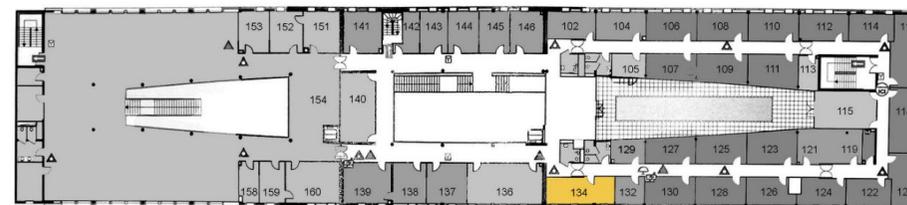


Figure 8 : Plan du premier étage de la MISHA, avec localisation du bureau de la PUD-S.

La MSH Bordeaux, bien qu'implantée sur le campus Montesquieu, se trouve à l'écart des pôles d'enseignement, avec des locaux vieillissants et l'absence notable de restauration administrative sur le site. D'autres, enfin, se situent à l'écart des sites universitaires, rendant complexe le travail des ingénieurs tant en matière de diffusion de l'information que d'organisation d'activités de formation ou d'événements scientifiques, c'est notamment le cas de la PUD-Aix-Marseille Université et de DemoMed, de PROGEDO-Loire ou de la PUD Lille où le récent déménagement (mars 2024) dans les locaux de l'INSPÉ, sur le campus Flers, pose de réelles questions d'accessibilité.

BUDGET

Sous l'impulsion, notamment, de Sébastien Oliveau, alors directeur de la TGIR PROGEDO (2018-2023), les PUD ont bénéficié de fonds du premier PNSO afin d'accompagner leur création et leur pérennisation. Cette enveloppe ministérielle a permis la création de nombreux postes d'ingénieurs dont la pérennité devait être ensuite assurée par les universités d'accueil. Seule exception depuis le PNSO : la PUD-Paris Saclay qui s'est ouverte sans fonds d'amorçage. La fin de ce financement (mieux connu sous le nom d'« enveloppe Vidal »), a créé un « trou d'air financier » selon les termes d'un référent scientifique, qui a mis en danger certaines plateformes. C'est notamment le cas de la PUD-Grenoble Alpes où cette instabilité financière a débouché sur le départ de l'ingénieur alors en poste, entraînant une mise en sommeil de la plateforme de juin 2022 à la fin de la même année. De plus, face à la nouvelle conjoncture économique, certaines universités n'ont pas pu intégrer ces postes à leur budget de fonctionnement de façon pérenne, poussant les MSH à trouver de nouvelles voies de financement pour tenter de maintenir les activités de leurs plateformes Data. C'est là une des spécificités du réseau de plateformes Data par rapport aux autres réseaux de plateformes des MSH. En effet, la très grande majorité des salaires des ingénieurs sont financés par les universités, cette homogénéité, ajoutée à la prévalence du statut de contractuel (86 %), fragilise les plateformes Data lorsque la présence de personnels CNRS (fonctionnaires), dans les autres plateformes permet d'assurer la viabilité du service. Certaines universités ont pu bénéficier de CPER (Contrats de plan État-Région) qui ont réellement permis le déploiement des plateformes comme un outil entre infrastructures et laboratoires de recherche, comme cela a été le cas pour la MSH en Bretagne. D'autres MSH, enfin, à l'image de la MSH Mondes, de la MSHS-Toulouse ou de la MSH Dijon, financent des postes d'ingénieurs sur ressources propres, ce qui permet d'avoir un second ingénieur sur

²¹ Marin B. et Siron V., 2017, *Les Sciences humaines dans leurs maisons*, Marseille : Parenthèses.

la plateforme, bien que sur des postes particulièrement précaires. Ceci est d'autant plus vrai pour le modèle de la PUD de Dijon puisque celle-ci répond à des AAP (appels à projets) de collectivités territoriales pour lesquelles elle mène des études entièrement mises en œuvre par l'ingénieure, de la phase de montage à la restitution.

De plus, les budgets disponibles pour les activités des PUD sont assez restreints, voire quasi inexistantes. La majorité des budgets de fonctionnement des plateformes Data est assurée par les MSH d'accueil. Lorsque ce n'est pas le cas, celui-ci est issu de projets accompagnés par les ingénieurs des plateformes du fait d'une politique budgétaire quasi généralisée imposant d'ajouter des lignes budgétaires dédiées aux projets accompagnés et faisant l'objet de financements externes (ANR par exemple). Il est toutefois impossible d'établir une moyenne réelle des fonds disponibles pour les plateformes, celles-ci ayant des modes de gestion extrêmement divers. Certaines bénéficient d'une enveloppe annuelle fixe, lorsque d'autres se voient alloués des fonds par leur MSH en fonction de leurs besoins et sur demande de l'ingénieur. D'autres, enfin, prennent en charge le volet RH de la plateforme en faisant appel à des partenaires extérieurs.

Malgré cette disparité, les retours de l'enquête montrent qu'un budget de fonctionnement annuel de 5 à 7 000 € permet de faire fonctionner une plateforme correctement. Cette somme se répartit alors sur les actions suivantes :

- Organisation d'évènements scientifiques,
- Organisation de réunions visant à impulser des collectifs autour des données quantitatives,
- Organisation de formations avec intervenants extérieurs défrayés ou rémunérés,
- Formation des ingénieurs en poste,
- Participation à des évènements nationaux divers (Journées inter-PUD, Journée annuelle Maté-SHS, etc.).

II - MEMBRES DU PERSONNEL QUI FONT VIVRE LE RÉSEAU DATA

Les plateformes ne vivent que grâce à l'engagement et au dynamisme de leurs ingénieurs. Il est commun de dire « il n'y a pas de PUD sans ingénieur ». Toutefois, ces agents ne travaillent pas seuls et bénéficient du soutien d'un nombre conséquent de personnes²².

INGÉNIEURS

Profil des ingénieurs

Le réseau Data fonctionne grâce à ses 21 ingénieurs aux expériences et statuts extrêmement divers ; 52 % sont des ingénieures. Trois personnes sont fonctionnaires (PANELS, PUD Lille et PUD Nanterre) et 18 sont contractuels, dont deux en CDI (PUD Caen et PUD-Toulouse) et trois devraient signer un CDI en fin d'année 2024. Les contrats des ingénieurs vont de huit mois à une durée indéterminée. Les plateformes Data disposent, en moyenne, de 1,40 personnels dédiés. Cette moyenne cache une importante disparité de ressources avec une valeur médiane d'une personne salariée par plateforme. DemoMed n'a en effet pas d'ingénieur, alors que PANELS dispose de trois ingénieurs et une assistante ingénieure, avec des temps de travail dédiés aux activités de la PUD variables. 71 % des ingénieurs sont à temps plein sur les activités de la PUD, permettant ainsi aux plateformes de bénéficier de 0,98 ETP, hors temps des référents scientifiques²³. Les personnes travaillant à temps partiel le sont par choix personnel. La plateforme PANELS est une exception avec des personnels à temps plein mais dont les fiches de poste incluent une participation à temps partiel aux activités de la PUD, le reste du temps se répartissant sur d'autres activités internes à la MSH-LSE.

Les parcours des ingénieurs sont également variés, permettant ainsi une complémentarité des compétences sur l'ensemble du réseau. Ainsi, les disciplines représentées au sein du réseau Data sont la démographie, la sociologie, l'économie, la géographie, l'histoire, les mathématiques appliquées aux sciences sociales et, dans une moindre mesure, les sciences politiques et l'anthropologie. 38 % des ingénieurs sont titulaires d'un doctorat et 33 % bénéficient du statut d'ingénieur de recherche (IR). Plusieurs d'entre eux ont un parcours post-doctoral dans des programmes de recherche et un ingénieur a été enseignant-chercheur à l'international. Les avis quant au niveau de diplôme nécessaire divergent. Pour certaines directions de MSH, la vocation d'accompagnement technique des ingénieurs PUD rend le doctorat superflu. Pour d'autres, le fait d'être titulaire d'un doctorat, et du parcours post-doctoral lié, est un réel atout car cela facilite la compréhension des impondérables liés au métier de chercheur, permet d'offrir des services ayant une haute valeur scientifique, et facilite le dialogue avec les chercheurs et les responsables de laboratoires tant dans l'organisation d'évènements scientifiques que pour diffuser l'offre de services auprès des chercheurs. Le statut d'ingénieur de recherche permet également aux ingénieurs d'intégrer des projets de recherche au nom de la plateforme, l'ancrant ainsi dans le paysage local comme un acteur technique et scientifique incontournable. C'est

²² Les données présentées ci-dessous ont été arrêtées à la fin du mois de mai 2024.

²³ Équivalent Temps Plein. Calcul réalisé sur une base de temps plein = 35h/semaine.

le cas, par exemple, de la PUD-Toulouse avec une participation aux projets ANR ObARDI (Infrastructure de Données sur l'Ancien Régime fondé sur des ontologies) et PANELVICO (Observer les conséquences sociales de la crise sanitaire Covid-19 sur le long terme). Ce cas de figure se retrouve également à la PUD-Université de Poitiers où une des utilisatrices des services de la PUD a fortement valorisé les apports scientifiques de l'ingénieure en prévoyant de l'inscrire comme co-auteur de publications à venir ; ainsi qu'à PROGEDO-Loire avec l'implication de l'ingénieure dans les projets « Les ouvriers des ports à l'époque moderne » et SUBWORK « Les espaces suburbains de production : quels emplois populaires ? » pour lequel elle a constitué des jeux de données dont une cinquantaine ont été déposés sur Nakala fin 2023.

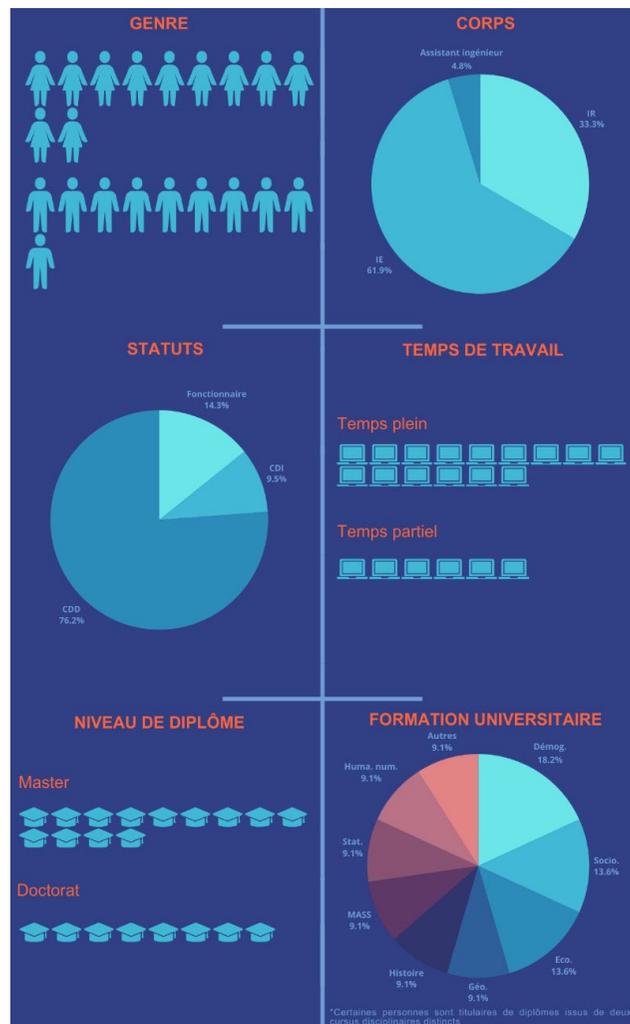


Figure 9 : Profil des ingénieurs, juin 2024.
Réalisation : C. Morand, sur canva.com

Recrutement et pérennisation des ingénieurs

L'engagement personnel des ingénieurs sur leur poste est important. Toutes et tous mettent en avant leur amour du métier, de sa polyvalence, l'acquisition de nouvelles connaissances et compétences qu'il permet. Toutes et tous font preuve d'un réel dynamisme dans l'accomplissement de leurs missions, qualité jugée indispensable par les personnes en charge du recrutement des ingénieurs. Le choix de s'investir professionnellement au sein de l'ESR est souvent lié à une vocation personnelle et à une valorisation du sens de « la mission du service public » pour reprendre les termes utilisés par un ingénieur interrogé. Cependant, cet investissement contraste fortement avec les problématiques RH dont font état des référents scientifiques et directions de MSH. En premier lieu, toutes les MSH ont fait part de difficultés au moment du recrutement du fait du haut niveau d'expertise technique qui est demandé, peu en adéquation avec les salaires proposés, surtout si ce dernier est mis en comparaison avec les salaires proposés dans le secteur privé pour des *data analysts*. À cela s'ajoute la précarité du contrat puisque seuls cinq ingénieurs (24 %) ont un poste pérenne. Parmi les personnes recrutées en CDD, la situation est là aussi extrêmement diverse et fonction tant du financeur du poste que des politiques RH appliquées par l'université tutelle. Ainsi, un des ingénieurs est en CDD depuis six ans et est encore en attente d'une pérennisation du poste. D'autres sont financés par l'*EquipEx+ : LifeObs*²⁴ – projet co-piloté, entre autres, par l'IR* PROGEDO et l'INED – (PUD-Bordeaux et PUD-Strasbourg) et ont donc des CDD de trois ans renouvelables une fois. Lorsque les PUD disposent de plusieurs ingénieurs, la norme générale est que le second poste soit financé par des projets de recherche qui apportent quelques mois de financement de postes d'ingénieurs (PUD-Toulouse et PUD Nanterre par exemple). De plus, les calendriers des universités ne permettent parfois d'annoncer un renouvellement de CDD que quelques semaines avant son échéance, ajoutant encore à la précarité des postes proposés. Cette situation fait courir le risque de voir partir les ingénieurs en cours de route et entraîne un *turn-over* dommageable pour la mission d'accompagnement à des projets de recherche qui s'inscrivent dans le temps long. Ces difficultés se sont notamment présentées à la PUD Caen où trois ingénieurs se sont succédé entre 2011 et 2016, ou à la PUD-Bretagne où six ingénieurs ont été en poste depuis 2018. Ces mouvements de personnels interrompent la continuité des actions de la plateforme et ne permettent pas d'établir des stratégies de développement à long terme, instabilité qui a également été soulignée par de nombreux ingénieurs en CDD.

²⁴ INED, 2021, « *EquipEx+ : LifeObs*, un nouvel équipement scientifique national pour mieux comprendre les tendances démographiques », communiqué de presse, 07.12.2021, https://www.ined.fr/fichier/s_rubrique/32271/ined_communique_lifeobs.vdef.fr.cor.fr.pdf.

RÉFÉRENTS SCIENTIFIQUES

Profil des référents scientifiques

Chaque plateforme dispose d'un référent scientifique (RS), voire deux pour la PUD-Bretagne. Les RS forment un maillage de personnes aux statuts, compétences et rôles divers en fonction des besoins locaux, de la façon dont la relation s'est construite avec l'ingénieur, de la physionomie des locaux et de leur propre personnalité, entre autres.

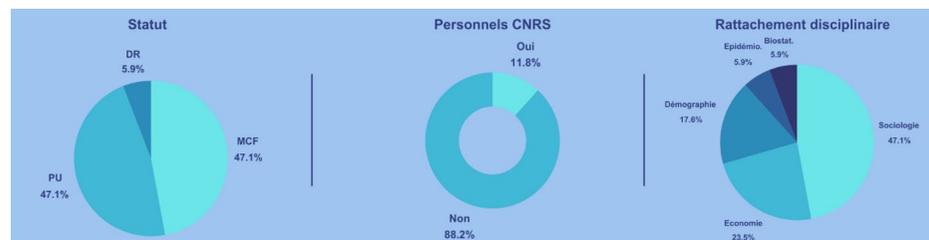


Figure 10 : Profil des référents scientifiques, juin 2024.
Réalisation : C. Morand, sur canva.com

Sur les 17 RS que compte le réseau Data, seuls deux sont des personnels CNRS, les autres étant des personnels universitaires, soit maîtres de conférences, soit professeurs des universités. Bien que la principale discipline représentée soit la sociologie, les RS sont issus de disciplines aussi diverses que l'économie, la démographie, l'épidémiologie ou la biostatistique. Le nombre d'heures qu'ils consacrent à la plateforme est extrêmement variable. Si certains se contentent d'intervenir lors de la Semaine Data-SHS, d'autres assurent une présence informelle quasi quotidienne auprès de l'ingénieur de la plateforme. Tous ne bénéficiant pas de décharges officielles pour cette responsabilité, il est malheureusement extrêmement compliqué de donner une estimation fiable du temps dédié, d'autant plus que celui-ci varie tout au long de l'année selon la charge de travail incombant à la plateforme. Toutefois, d'après les déclarations des RS et ingénieurs, on estime que les RS consacrent en moyenne entre 35 et 40 heures par an à la plateforme. Plusieurs facteurs déterminent les modalités de cette implication :

- ▮ La présence physique facilitée par une proximité géographique des bureaux ;
- ▮ La possibilité de bénéficier d'une décharge d'heures pour se consacrer à la plateforme ;
- ▮ Le niveau de compétences et l'ancienneté de l'ingénieur ;
- ▮ L'intérêt réel porté aux activités de la plateforme.

Les modalités concrètes d'interaction avec les ingénieurs sont diverses et ont souvent été construites sur le temps long, en concertation avec l'ingénieur. Si certains préfèrent organiser des rencontres régulières et formalisées, d'autres optent pour des contacts *ad hoc*, plus informels, rendant d'autant plus complexe une estimation réaliste du temps dédié à la plateforme.

Rôle des référents scientifiques

Le rôle des RS n'étant pas formalisé, il diffère selon les sites et les personnes concernées ; une constante cependant : la présence du RS auprès de l'ingénieur diminue à mesure que

celui-ci acquiert de l'expérience sur son poste. De façon générale, le RS est impliqué dans les recrutements au sein de la plateforme et participe de façon active à l'organisation de la Semaine Data-SHS soit en co-construisant le programme avec l'ingénieur, soit en servant de relais entre la plateforme et les chercheurs invités, soit en intervenant au cours de la semaine. Selon les sites, les RS ont ensuite endossé des responsabilités diverses telles que la participation aux négociations pour la pérennité des postes, la diffusion d'informations sur la plateforme, la relecture de différents rapports, ou la participation à des décisions plus politiques ou techniques liées aux activités de la plateforme. Ceci est particulièrement vrai pour la PUD Dijon et la PUD-Strasbourg où les RS occupent des postes de direction au sein de la MSH (respectivement directeur et directeur adjoint). D'autres configurations, comme celle de la PUD Université de Poitiers, permettent de partager la responsabilité avec d'autres personnes. En effet, à Poitiers, les orientations et missions à venir de la plateforme sont décidées par un consortium de quatre personnes (gestionnaire et référent scientifique de la plateforme, responsable de la plateforme technologique de la MSH et responsable du pôle recherche et innovation) qui se réunissent au moins une fois par mois.

Le choix du RS est déterminé de diverses manières et peut se faire soit par le bouche-à-oreille entre collègues, soit par un recrutement formalisé avec publication d'une fiche de poste comme cela a été le cas à la PUD Lille ou à la PUD Aix-Marseille Université. Cette modalité de nomination a un impact sur le rôle politique occupé par le RS auprès de la plateforme. En effet, un RS rattaché à un laboratoire ne faisant que peu appel à la plateforme Data permettra de resserrer les liens entre ces deux entités et ainsi attirer un public avec une diversité disciplinaire plus importante. L'exemple de la PUD Bretagne est parlant à ce titre puisque les deux RS sont PU à l'EHESP et exercent dans des disciplines à la frontière entre les SHS et les sciences expérimentales. Le choix de faire appel à ces deux personnes a notamment permis de resserrer les liens entre l'EHESP et la MSH, la dernière étant installée dans les locaux de la première. L'exemple de PROGEDO-Loire est également intéressant de ce point de vue puisque le RS a souhaité être présent auprès de la plateforme jusqu'à l'obtention d'un poste pérenne pour celle-ci et transmettra ensuite ses fonctions à un membre d'un autre laboratoire afin d'éviter que la plateforme ne soit perçue comme le domaine réservé des sociologues nantais. Enfin, la PUD-Toulouse a fait le choix d'un RS (seul RS non titulaire) rattaché à la Toulouse School of Economics (TSE), site historiquement moins proche de la plateforme tant du fait de son emplacement géographique que des compétences de ses chercheurs en matière de données quantitatives.

PERSONNELS DE SOUTIEN

Il faut ajouter à ces personnes les personnels des MSH en charge de la gestion, la communication, la science ouverte ou la direction qui sont également partie prenante des activités des plateformes Data, bien que de façon plus ponctuelle. Comme cela a été souligné par l'ensemble des personnels Data, il s'agit là d'une des grandes forces de l'intégration des plateformes au sein des MSH. Cela permet de bénéficier d'un réseau de professionnels qui aident à la communication autour des événements scientifiques et de l'offre de formation, de gestionnaires capables de prendre en charge la gestion budgétaire des plateformes, d'informaticiens ou, tout simplement, des « collègues » avec lesquels échanger au quotidien. Les liens les plus évidents se font avec les ingénieurs responsables des humanités numériques, liens qui sont d'autant plus naturels lorsque les bureaux sont partagés comme

à la PUD Lille par exemple. La répartition des activités est souvent pensée et résumée comme une répartition entre données quantitatives pour Data et données qualitatives pour les humanités numériques. Parfois, les accompagnements aux chercheurs nécessitent de faire appel aux deux pôles de compétences et impliquent donc l'ensemble des personnels, comme cela s'est déroulé pour certains accompagnements réalisés à la MSH de Dijon. Toutefois, les interactions entre ces pôles ne vont pas toujours de soi et ne se mettent parfois en place que sur les organigrammes, sans application concrète sur le terrain. D'autres plateformes, comme la PUD-Aix-Marseille Université avec le projet HyFlex²⁵ à l'été 2024, comptent ponctuellement sur des étudiants stagiaires qui participent aux activités quotidiennes de la plateforme tout en bénéficiant de l'expertise technique de la plateforme pour leurs propres travaux. DemoMed ne disposant pas d'ingénieurs profite de la présence de doctorants pour assurer une partie de l'actualisation de sa base de données. Enfin, la PUD-Strasbourg bénéficie d'un poste de maître de conférences contractuel dédiant une quotité de son temps de travail à la plateforme, notamment pour des activités de formation aux données quantitatives et de promotion de la PUD-S.

III - ACTIVITÉS ET SERVICES RENDUS À LA COMMUNAUTÉ SCIENTIFIQUE

UTILISATEURS

En théorie, les services des plateformes Data sont accessibles à l'ensemble des personnes du milieu de la recherche en SHS rattachées aux UAR et UMR elles-mêmes rattachées aux MSH, soit les étudiants de Master, les doctorants, les enseignants-chercheurs et chercheurs ainsi que les personnels d'appui à la recherche. Ces derniers ne représentent toutefois qu'une portion congrue des personnes accompagnées et bénéficient, le plus souvent, d'une aide de proximité « entre collègues ». Ces accompagnements sont plus perçus par les ingénieurs comme des actions de mutualisation des compétences que comme de réels accompagnements à la recherche. En pratique, il est complexe d'établir une cartographie des utilisateurs des services de la plateforme car les ingénieurs répondent parfois à des demandes ponctuelles et rapides (moins d'une heure d'intervention) sans en tenir le décompte, ou sans demander à l'utilisateur les informations permettant de le rattacher à une unité de recherche. Cependant des analyses peuvent être réalisées à partir des informations consignées par les ingénieurs sur l'année 2023.

En moyenne²⁶, en 2023, les plateformes Data (hors DemoMed) ont accompagné 30 chercheurs (médiane de 28) sur le cycle de vie des données qu'il s'agisse d'une aide à la collecte des données, à leur traitement et analyse, à leur dépôt ou à la recherche de jeux de données, chiffre probablement sous-estimé. Sur la période 2020-2023, DemoMed a formé 10 étudiants de Licence et Master ; des doctorants ont également été accompagnés par la plateforme, pour un total de 1540 heures en 2022. Les accompagnements collectifs (formations logicielles, présentation de la plateforme et des jeux de données disponibles, séminaires de recherche autour des données quantitatives²⁷) ont, quant à eux, concerné en moyenne, 145,44 personnes en 2024. Toutefois, ce chiffre est à pondérer car la médiane des accompagnements collectifs est de 41 personnes. Cet écart important se justifie par le fait que certaines plateformes privilégient les accompagnements individuels, proposant donc une offre collective plus limitée, quand d'autres font le choix inverse. Ces choix sont toujours dictés par les besoins locaux. Il faut également tenir compte de l'ancienneté des plateformes ; les plus anciennes étant bien implantées localement, elles ont donc plus de facilité à attirer du public. La localisation géographique des plateformes est également un enjeu important concernant l'assistance car l'éloignement des campus SHS réduit l'attractivité des formations collectives proposées. La PUDL, par exemple, compte une moyenne de 600 participants par an aux actions collectives, notamment du fait de son cycle de séminaires dédié aux méthodes mixtes de recherche.

Les données disponibles ne permettent pas d'élaborer de profil-type des utilisateurs, tant en

²⁵ Fournier P., 2022, « Dessiner la transition vers l'hydrogène dans les territoires : quelle flexibilité des projets entre risques environnementaux et risques socio-politiques? », AAP « Transition en action » juin 2022, https://www.univ-amu.fr/system/files/2023-01/ITEM_HyFlex.pdf.

²⁶ Calculs effectués sur la base des données fournies par 11 plateformes. Les autres plateformes ne disposaient pas de chiffres actualisés pour l'année 2023 ou ne tiennent pas un calcul précis du nombre d'utilisateurs du fait de la variabilité des demandes, préférant calculer le temps dédié par l'ingénieur à chaque activité. La PUD-Paris Saclay ayant été inaugurée en décembre 2023, elle ne disposait pas de données sur le sujet.

²⁷ Les participants à la Semaine Data-SHS n'ont pas été pris en compte dans cette section.

matière de rattachement disciplinaire que de statut. Deux raisons principales : les statuts des principaux utilisateurs sont très variables selon les sites et dépendent d'un nombre important de facteurs (liens avec les laboratoires, écoles doctorales et responsables de Master ; profil disciplinaire du site, entre autres). Il semble toutefois que les utilisateurs les plus fréquents, notamment pour l'accompagnement collectif, soient les chercheurs de demain (masterants et doctorants). Une portion congrue de formations sont proposées dès la L3, en réponse à des demandes des enseignants responsables de parcours. Les plateformes se saisissent de ce public pour former la nouvelle génération de chercheurs et ainsi assainir les pratiques autour des données dès le début des processus de recherche. Cela répond à un autre constat dressé par les ingénieurs : les chercheurs ont généralement développé leurs propres méthodes et outils et n'ont que peu de temps disponible pour se former à de nouveaux usages. Concernant l'accompagnement individuel, la tendance semble s'inverser avec une participation plus importante de chercheurs et enseignants-chercheurs (42 % des utilisateurs²⁸), notamment du fait de la politique de généralisation des PGD et de FAIRisation des données, directement suivis par les doctorants (35 %). Un profil disciplinaire est cependant difficile à définir du fait de la complexité du paysage scientifique français, doté de nombreux laboratoires pluridisciplinaires. Toutefois, des tendances émergent des données les plus récentes (2022-2023) fournies par 7 plateformes.

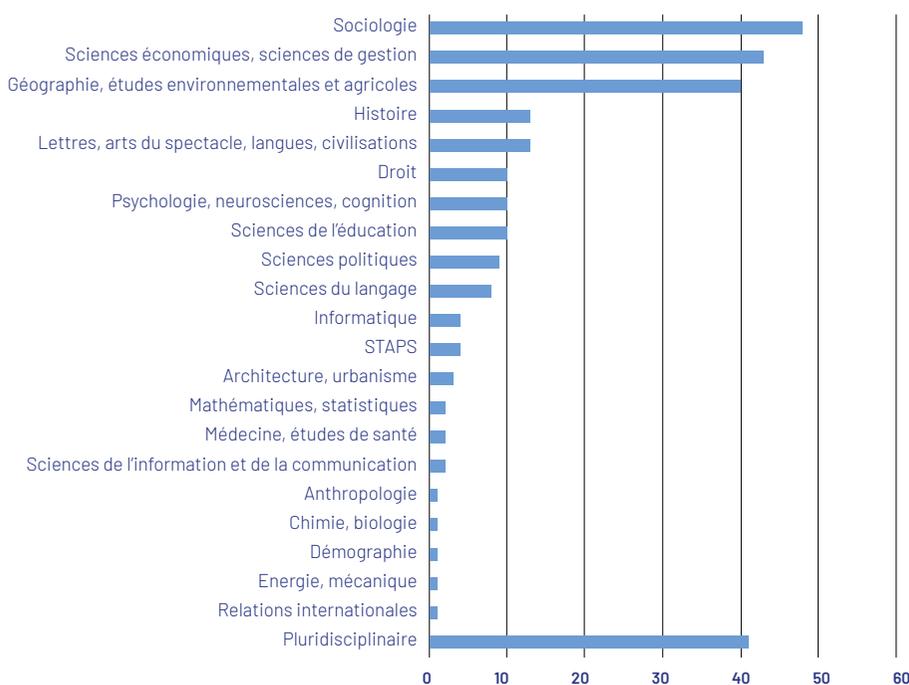


Figure 11 : Profil disciplinaire des utilisateurs des services des plateformes Data.

²⁸ Calcul réalisé sur la base des informations les plus récentes (2022-2023) fournies par 6 plateformes.

Champ : personnes accompagnées par sept plateformes Data en une année, comprise entre 2022 et 2023

Les chercheurs les plus fortement utilisateurs des accompagnements individuels des plateformes seraient rattachés à des laboratoires de sociologie, de sciences économiques et de gestion, pluridisciplinaires et de géographie et études environnementales et agricoles. Cependant, le spectre disciplinaire des utilisateurs est bien plus vaste, démontrant un intérêt croissant pour la culture des données, y compris dans des domaines disciplinaires supposés plus imperméables à ces sujets. De façon contre-intuitive, les démographes ne font que peu appel aux plateformes Data. Cela s'explique par le haut niveau de qualification de ces chercheurs en matière statistique, ainsi que par leur bonne connaissance des jeux de données ouvertes accessibles à la recherche.

ACCOMPAGNER LES CHERCHEURS VERS L'UTILISATION DES DONNÉES QUANTITATIVES

La mission principale des plateformes Data est l'accompagnement aux chercheurs dans la gestion des données quantitatives par la transmission de bonnes pratiques. Cet accompagnement est proposé selon deux modalités complémentaires : un accompagnement collectif et un accompagnement individuel. Ces activités s'ancrent dans une stratégie globale de développement des activités des plateformes. Accompagnements collectifs et individuels sont souvent conçus comme complémentaires et créent un vivier de participants et intervenants pour les différents événements scientifiques organisés par les plateformes. Ils permettent également de constituer des réseaux locaux de chercheurs travaillant avec des données quantitatives.

Accompagnement collectif

L'accompagnement collectif représente un enjeu important pour les plateformes car « [il] permet de mutualiser les services, de créer des temps de rencontre et d'échange (et ainsi favoriser l'interdisciplinarité), d'asseoir la culture de la donnée en formant notamment les jeunes chercheurs, d'assurer la pérennité des projets et même leur suite en rendant plus autonomes les chercheuses et les chercheurs »²⁹. Certains accompagnements sont inscrits dans les catalogues de formations des ED, facilitant ainsi le travail de communication et d'organisation des ingénieurs. Les accompagnements collectifs sont généralement limités à un public d'une dizaine de personnes et sont à destination des chercheurs et EC, des doctorants et des masterants. Certaines plateformes proposent des cycles de formations récurrents, comme la PUD-Aix-Marseille Université qui offre un accompagnement logiciel sur R (du niveau débutant à avancé) et Microsoft Excel (niveau débutant) avec comme objet la découverte de ces environnements de travail et l'introduction à certaines notions de collecte, de traitement, d'analyse et de visualisation de données.

En moyenne, les plateformes forment, chacune, 140 personnes par an³⁰, en majorité sur l'utilisation de logiciels de traitement, analyse et visualisation de données quantitatives. D'autres sessions sont organisées sur le RGPD, sur les jeux de données disponibles via

²⁹ PROGEDO-Loire, 2024, « Rapport d'activité 2023 ».

³⁰ Calcul effectué sur la base des données les plus récentes (2021-2023) fournies par 12 plateformes.

PROGEDO, sur les bonnes pratiques de collecte de données ou sur les méthodes d'analyses statistiques en fonction des demandes qui émergent de la communauté scientifique locale. Dans une optique de développement de la science ouverte, des plateformes proposent des tutoriels d'auto-formation comme la PUD-Clermont Auvergne sur les tableaux croisés dynamiques ou la PUD-Aix-Marseille Université sur R et Excel, par exemple. Ces outils sont ensuite réutilisés par d'autres plateformes dans un objectif de mutualisation des ressources disponibles.

Certaines plateformes conçoivent ces ateliers comme des formules hybrides mêlant formation et accompagnement individuel. PROGEDO-Loire propose des formations « ouvertes à tous », conçues comme complémentaires à l'accompagnement individuel. Chaque participant travaille sur ses propres jeux de données pendant les ateliers et bénéficie, en complément, d'un accompagnement individuel par l'ingénieure. Cette formule permet un accompagnement complet des chercheurs et fait donc face à une hausse des demandes depuis 2023. Elle permet également à l'ingénieure de mutualiser ses services et ainsi d'accompagner un nombre plus important de chercheurs (12 participants par atelier en moyenne en 2023). L'atelier le plus demandé est une initiation à R pour les SHS, tendance qui semble se vérifier au niveau national.

Accompagnement individuel

Les accompagnements individuels proposés aux chercheurs en SHS par les plateformes recouvrent l'ensemble du cycle de vie de la donnée en fonction des besoins exprimés par les utilisateurs. Ce service est véritablement conçu comme une plus-value de la MSH par les utilisateurs qui en ont bénéficié. Il peut se faire à différentes étapes du projet selon le moment de l'enquête au cours duquel la plateforme est sollicitée. Cela peut-être dès la phase de montage du projet, auquel cas la plateforme est directement intégrée au projet ; ou bien la plateforme peut être sollicitée en cours de projet. Il s'agit là d'un accompagnement qui est généralement prévu comme ponctuel, mais qui va bien souvent s'inscrire sur le temps long, les chercheurs interrogés saluant l'importante plus-value technique et scientifique de ce type d'accompagnement. Ainsi, les accompagnements individuels sont un réel levier d'amélioration des pratiques sur le temps long du fait de la systématisation de l'accompagnement qu'ils entraînent. PROGEDO-Loire dresse le même constat en notant un doublement des demandes d'accompagnement à l'ouverture des données en 2023 du fait des politiques gouvernementales certes, mais aussi des formations et accompagnements mis en œuvre par la plateforme en 2022. Ces accompagnements permettent également de positionner les plateformes comme des acteurs centraux de la donnée par les mises en relation qu'elles proposent, particulièrement lorsqu'il n'y a pas d'Atelier de la donnée opérationnel sur le site.

Les stratégies de gestion des flux de demandes varient d'une plateforme à l'autre, et sont souvent le fait d'une politique mise en œuvre par la MSH. La majeure partie des plateformes ont, jusqu'à présent, opté pour des accompagnements au fil de l'eau, c'est-à-dire que le calendrier des accompagnements est totalement dépendant des demandes. D'autres MSH ont mis en place des systèmes de tri qui permettent de hiérarchiser les demandes et ainsi lisser les activités sur l'année. La MSHB, par exemple, propose un AMI pour bénéficier d'un accompagnement prioritaire de la PUD Bretagne. Autre exemple, la direction de la MSH Ange-Guépin priorise les projets accompagnés dès leur démarrage par la MSH. Une constante cependant, toute demande légitime et pertinente par rapport au domaine de compétences des plateformes doit être traitée.

En moyenne, les plateformes accompagnent 30 projets par an³¹. Les demandes qui sont adressées aux plateformes peuvent être classées selon cinq grandes catégories :

- La création, la collecte et la préparation de données ;
- L'identification de sources de données pertinentes ;
- Le traitement et l'analyse statistique des données, la visualisation de données ;
- L'aide à la constitution de projets et de PGD ;
- Autres types de demandes (science ouverte, expertise de dossier ANR ou HCERES, etc.).

Toutes les plateformes ne sont pas confrontées à chaque catégorie de demandes. Il est fréquent, par exemple, que les SCD prennent en charge l'accompagnement à la rédaction de PGD dans le cadre des Ateliers de la donnée. Afin de limiter les accompagnements spécifiquement dédiés au RGPD, PROGEDO-Loire, en collaboration avec Huma-Num Loire et la Cellule d'édition électronique, a créé, en décembre 2022, un blog sur lequel sont diffusées les questions posées à ce sujet et les réponses qui y ont été apportées. En janvier 2024, il a enregistré 271 visites, preuve de la nécessité de développer ce type d'outils.

COMMUNIQUER AUTOUR DES DONNÉES QUANTITATIVES EN SHS

Les actions de communication mises en place par les plateformes permettent d'améliorer leur implantation locale tout en diffusant les bonnes pratiques autour des données quantitatives au niveau national. L'organisation d'événements de communication est à la charge des ingénieurs, souvent soutenus par les RS. La stratégie de communication des plateformes est à la fois harmonisée au niveau national (page dédiée sur les sites des MSH, sous-collection HAL-PROGEDO et carnet Hypothèses PROGEDO à l'exception de DemoMed), tout en laissant une grande marge de manœuvre aux ingénieurs. La PUD Lille et DemoMed sont les seules plateformes à bénéficier de sites internet dédiés. Toutefois, cette autonomie risque de mener à un éclatement et à une déperdition de l'information. Ainsi, les plateformes communiquent par le biais de newsletters (propre à la plateforme ou à la MSH), de listes de diffusion (internes, thématiques ou métiers), de réseaux sociaux (X, Bluesky), de Canal U (en propre ou non), des sites internet des universités, d'OPIDoR, de sites internet d'acteurs locaux (consortiums de plateformes ou d'acteurs de la donnée, acteurs institutionnels, partenaires divers), de sciencesconf.fr.

Faire connaître les jeux de données du catalogue Quetelet-Progedo-Diffusion

Des actions de sensibilisation sont proposées au moment de la rentrée universitaire et visent des publics divers : étudiants de Master, doctorants (que cela passe, ou non, par les écoles doctorales) et chercheurs par le biais des laboratoires. Les PUD interviennent régulièrement dans les laboratoires et séminaires de rentrée afin d'ancrer l'habitude d'intégrer la plateforme dans les projets de recherche, ce qui, pour l'instant, n'est pas systématique. D'autres initiatives visent le même objectif, tel que le projet « Chercher ses données »

³¹ Calcul effectué sur la base des données les plus récentes (2021-2023) fournies par 14 plateformes.

auquel participent la PUD Lille et la PUD Université de Poitiers. La PUD Nanterre organise des séminaires thématiques visant à communiquer sur les grandes enquêtes et bases de données administratives. En plus de ces temps forts, toute sollicitation est une opportunité de présenter le réseau et les activités des PUD, comme lors des Wiki cafés locaux ou nationaux ou des Cafés du SPD du CNRS³² par exemple. Il s'agit généralement d'une présentation des activités mises en place par la plateforme et des services proposés aux chercheurs. Selon certains ingénieurs, ces actions sont particulièrement chronophages pour un faible retour sur investissement, ce qui les rend peu valorisantes. Ces temps sont également des moments privilégiés de présentation des données accessibles via le catalogue Quetelet-Progedo-Diffusion, dont les PUD sont une des premières portes d'entrée. Ces présentations visent à renouveler le vivier de participants aux différentes activités des plateformes, tout en étant une première étape de sensibilisation à l'usage des données quantitatives, toutes disciplines confondues. Elles sont d'ailleurs généralement suivies d'une hausse des demandes d'accompagnement, qui ne compense pas toujours l'investissement mis dans ces activités de communication. Enfin, elles permettent de créer un premier lien, ou de le renouveler, amorçant ainsi le travail de préparation de la Semaine Data-SHS.

mêle des ateliers de formation à des outils de traitement et analyse des données, des présentations des jeux de données accessibles par le biais de la plateforme, des séminaires de présentation du RGPD, ainsi que des journées d'études au cours desquelles des chercheurs présentent leurs travaux et méthodes de travail autour des données quantitatives. Plusieurs plateformes font appel à des personnes accompagnées afin de leur permettre d'échanger avec leurs pairs sur des travaux en cours.

Les plateformes comptabilisent une moyenne de 86 participants à la Semaine Data-SHS³⁴. Toutefois, cette moyenne n'est qu'approximative car certaines plateformes comptabilisent le nombre de participants par session quand d'autres calculent le nombre de participants uniques. De plus, cette moyenne ne reflète pas la grande disproportion qui existe entre les différents sites puisque certaines plateformes reçoivent une trentaine de personnes quand d'autres approchent les 200. Ces disparités sont liées à des facteurs divers tels que la taille du campus, l'importance du quantitatif sur le site, les moyens de communication disponibles et l'accessibilité des lieux. En général, l'organisation de la Semaine Data-SHS requiert un travail d'une dizaine d'heures, réparties sur les trois derniers mois de l'année, et mobilise un nombre conséquent de personnes tant pour l'organisation logistique, la recherche d'intervenants que la communication autour de l'évènement. La participation aux différentes sessions étant extrêmement variable, elle entraîne parfois un sentiment de démotivation important lié au gouffre entre l'investissement personnel et le faible nombre de participants (sur certaines sessions, seules 4 à 5 personnes participent). De plus, les campus sur lesquels peu de chercheurs sont impliqués dans les données quantitatives n'offrent qu'un vivier limité d'intervenants potentiels. Cela rend les programmes quelque peu redondants et donc potentiellement moins intéressants pour le public. Afin de pallier ces difficultés, certaines plateformes mutualisent le travail en proposant des sessions communes, comme l'ont fait les PUD-Strasbourg, Nanterre et Toulouse en 2022, ou, comme le font les PUD franciliennes, en proposant un programme multisite, n'engageant ainsi chaque plateforme que sur une journée. Pour les MSH, la Semaine Data-SHS est une véritable vitrine des actions et services mis en place par les plateformes. Elle permet également de fédérer les chercheurs travaillant avec des données quantitatives, participant ainsi de la mission d'impulsion de réseaux qui incombe aux plateformes Data.

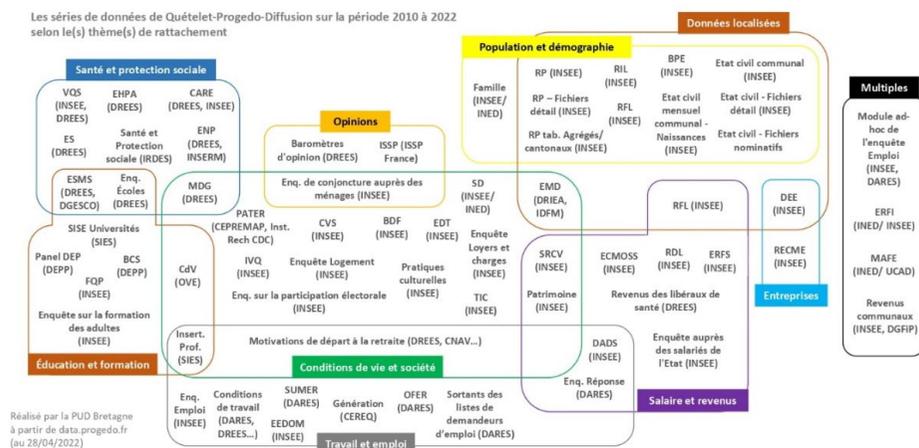


Figure 12 : Données accessibles sur le catalogue Quetelet-Progedo-Diffusion par thèmes, avril 2022. Source : PUB Bretagne à partir de data.progedo.fr

Semaine Data SHS

Créée en 2017 par la PUD Lille, la Semaine Data-SHS est, depuis 2019, un temps fort de l'année pour les PUD et l'IR* PROGEDO. Organisée chaque année, elle se tient la deuxième semaine de décembre dans chaque site disposant d'une PUD. « Cette semaine vise à mettre en avant l'utilisation et la réutilisation des données quantitatives en SHS, en particulier les grandes enquêtes et les données issues de la statistique publique. Elle vise aussi à promouvoir les bonnes pratiques autour de leur production (questionnaire, échantillonnage, collecte, documentation, diffusion) et leur usage (méthodes, outils, valorisation) »³³. Cette semaine

CONSTRUIRE DES RÉSEAUX MÉTIERS ET THÉMATIQUES

Une des missions des plateformes Data, plus ou moins formalisée selon les sites, est de contribuer à construire des réseaux autour des données quantitatives en SHS. Les actions d'accompagnement et de communication qu'elles déploient sont un des meilleurs outils pour constituer ces réseaux.

³² Cafés du SPD, PUD-T, « Les plateformes universitaires de données », 24.09.2024.

³³ PROGEDO, « Semaine Data-SHS », <https://www.progedo.fr/services/plates-formes-universitaires-de-donnees/semaine-data-shs/>.

³⁴ Moyenne calculée sur la base des informations fournies par 14 plateformes concernant les informations les plus récentes disponibles, soit entre les années 2021 et 2023.

Des réseaux autour des données quantitatives

Le fait de coordonner, ou simplement d'accompagner, la construction de réseaux est une véritable action de soutien aux communautés de recherche locales. Les plateformes impulsent des synergies et offrent des espaces qui permettent aux chercheurs de se rencontrer, de constituer des groupes de travail et d'identifier les interlocuteurs pertinents sur leur site de rattachement. La quasi-totalité des plateformes ne propose de formations qu'en présentiel entre autres pour permettre aux participants de se rencontrer et toutes privilégient les intervenants locaux pour ces mêmes raisons. Plusieurs plateformes mettent en avant ce travail de mise en réseau, qui peut passer par des biais différents. Ainsi, PROGEDO-Loire profite de la Semaine Data-SHS pour proposer des espaces de travail collaboratifs aux chercheurs bénéficiant de son accompagnement. Lorsqu'elle est impliquée dès la phase de construction de questionnaires, la PUD Nanterre devient un intermédiaire qui met en relation l'ensemble des acteurs d'un projet : chercheurs, service informatique, DPO, etc. Ce travail de mise en réseau peut engager jusqu'à six mois de préparation lorsque l'ensemble de ces acteurs sont impliqués dans un projet. D'autres plateformes profitent des événements scientifiques qu'elles organisent ou coorganisent pour œuvrer à cette mise en réseau. Ainsi, l'école d'été coorganisée par la PUD-Bretagne sert officiellement cet objectif : « L'objectif de l'école d'été est d'ouvrir un espace inédit de formation, de recherche et de travail collectif autour des méthodes mixtes [...]. Les différents ateliers constituent l'occasion pour les participant-e-s qui le souhaitent de préparer des manuscrits d'articles qui pourront être soumis dans le cadre de cette section »³⁵. Une des ingénieures du réseau exprime le fait que les formations ont certes pour objectif de familiariser des publics variés aux usages des données quantitatives dans la recherche en SHS, mais elle les voit également comme un prétexte à la création d'une communauté d'utilisateurs, pilier fondamental de la mission d'une plateforme d'appui à la recherche.

Participation aux réseaux métiers

Les ingénieurs des plateformes Data sont également fortement impliqués dans des réseaux métiers qui leur permettent de construire du collectif et de participer à des actions de formation et d'auto-formation. Un des principaux réseaux dans lequel la communauté s'investit est le réseau Mate-SHS du CNRS SHS qui regroupe des ingénieurs en production, traitement, analyse et représentation des données. Mate-SHS comporte plus de 1000 membres, coordonnés par un comité de pilotage, au sein duquel a été élue l'ingénieure en charge de la PUD Lille. Ce réseau regroupe également des groupes locaux et thématiques dans lesquels se retrouvent plusieurs ingénieurs Data soit en tant que membres, soit en tant qu'acteurs de la création et du pilotage de ces groupes, comme l'ingénieur de la PUD-Aix-Marseille Université au sein du réseau CoDaPro (Collectif Data Provence).

Mate-SHS propose régulièrement des séminaires et tutoriels³⁶ autour du traitement des données au sens large, ainsi que diverses actions de formation (ANF, écoles thématiques). L'ANF 2024 est consacrée aux méthodes et traitements appliqués à l'analyse longitudinale en SHS, plusieurs ingénieurs de plateformes Data ont été sélectionnés afin de maintenir et développer leurs compétences métiers. Une rencontre est organisée chaque année pour faire le bilan de l'année écoulée, proposer de nouveaux projets et échanger sur des thèmes

³⁵ MSHB, « École d'été – Méthodes mixtes 2024 », <https://www.mshb.fr/event/ecole-dete-methodes-mixtes-2024>.

³⁶ Mate-SHS, « Les Tuto@Mate », <https://mate-shs.cnrs.fr/actions/tutomate/>.

en lien avec le métier d'ingénieur en SHS. En 2024, cette rencontre a été coorganisée par l'ingénieure de la PUD Lille et a bénéficié d'une large présence des ingénieurs Data. Elle portait sur l'intelligence artificielle et ses applications et implications pour le métier d'ingénieur en SHS, ce qui a permis à l'ingénieur de la PUD-Grenoble Alpes de présenter l'outil de transcription automatique d'entretiens Taddam³⁷. Ainsi, la participation active du réseau des plateformes Data au réseau Mate-SHS permet de construire une communauté nationale autour du métier d'ingénieur en SHS, tout en contribuant à la nécessaire formation permanente que le métier d'ingénieur implique.

SPÉCIALISATION THÉMATIQUE ET COLORATION LOCALE

Réseau des PUD : quelles spécificités locales ?

Le RnMSH répond à la volonté de territorialiser la recherche en construisant une identité scientifique propre à chaque MSH et une complémentarité entre les différents acteurs qui constituent ce réseau³⁸. Les plateformes Data s'intègrent dans cette stratégie en développant des compétences spécifiques à chacune d'entre elles et répondant à des problématiques locales, compétences qui sont ensuite valorisées au niveau national en faisant d'elles des expertes reconnues dans leur domaine. Ainsi, la plateforme DemoMed et la PUD-Aix-Marseille Université participent à la volonté de la MMSH de construire des projets de recherche internationaux sur le pourtour méditerranéen. La PUD Lille propose des séminaires sur les méthodes mixtes³⁹ et se positionne comme centre de référence pour la constitution d'un réseau national autour de ces méthodes de recherche encore peu développées et fédérées en France. Depuis 2023, la MSH-Bretagne a rejoint ce mouvement en participant à une école d'été sur les méthodes mixtes de recherche. La PUD-Bordeaux et la PUD-Strasbourg sont des acteurs incontournables sur les questions démographiques et les parcours de vie du fait des profils historiques des ingénieurs et référents scientifiques et de leur implication dans l'*EquipEx+ : LifeObs*. La PUD-Toulouse, PROGEDO-Loire et la PUD-Université de Poitiers proposent une approche technique et scientifique des projets accompagnés du fait des parcours scientifiques antérieurs des ingénieurs. La PUD-Grenoble Alpes propose une coloration « intelligence des données », se tournant vers des solutions innovantes, notamment par la création de l'outil Taddam. À ces tendances s'ajoutent d'autres formes de spécialisation, peut-être moins marquées, liées aux volets d'activités proposés. En effet, certaines plateformes privilégient l'accompagnement individuel, d'autres vont accentuer les formations collectives quand d'autres encore misent sur l'organisation d'événements visant à construire des collectifs de chercheurs autour des données quantitatives. Cependant, ces tendances vers la spécialisation ne sont pas issues d'une stratégie politique supérieure, mais sont une réponse à la demande locale, complétée par le recrutement d'ingénieurs correspondant aux besoins spécifiques de chaque site.

³⁷ Le projet TADDAM (Transformations, Analyses et Développements de Données et d'Archives Multimédias) est porté par la PUD-GA et le laboratoire PACTE (UMR 5194), avec le soutien du GRICAD et de l'UGA.

³⁸ Commaille J., Rouillard P. et Wolikow S., 2020 [2019], « La politique du Réseau national des Maisons des Sciences de l'Homme », in Commaille J. (dir.), *op. cit.*

³⁹ Chmielewsky A. et Guergoat-Larivière M., 2023, « Plateforme Universitaire de Données de Lille. Bilan d'activité 2022 ».

Pour autant, les chercheurs n'entrant pas dans le cadre de ces thématiques privilégiées peuvent également bénéficier d'un accompagnement par la plateforme. Certains craignent qu'afficher une coloration ne freine certains utilisateurs qui se trouveraient en-dehors du champ, mais d'autres soulignent la nécessité de construire des pôles d'expertise reconnus afin d'avoir une assise européenne claire. Toutefois, la précarité des postes d'ingénieurs et l'important *turn-over* que certaines connaissent est un frein réel à la définition d'une stratégie d'expertise thématique de long terme.

Cas particulier de la plateforme technologique DemoMed

La plateforme DemoMed est le seul membre du réseau à ne pas être relais de l'IR* PROGEDO et donc à ne pas être une PUD. Cette plateforme est une émanation d'un projet de recherche, lancé en 2009 et pensé par des démographes d'Aix-Marseille Université, accompagnés par des géographes et démographes de divers pays du pourtour méditerranéen. Son objectif est de rendre interopérables des données démographiques issues de différents pays méditerranéens et produites à des niveaux administratifs et époques diverses. Elle se construit autour de six axes de recherche pluridisciplinaires :

- ▮ « Sources et méthodes » ;
- ▮ « Dynamiques démographiques de population » ;
- ▮ « Structure de la population » ;
- ▮ « Mobilités et migrations » ;
- ▮ « Démographie de la famille » ;
- ▮ « Prospective démographique ».

DemoMed est dirigée par Isabelle Blöss-Widmer, également RS de la PUD-AMU, et codirigée par Yoann Doignon, ancien ingénieur de la PUD-AMU, et Sébastien Oliveau, ancien directeur de l'IR* PROGEDO et actuel directeur de la MSH-PS. DemoMed entretient des liens étroits avec la PUD-AMU et ce d'autant plus qu'elle est pilotée par des enseignants-chercheurs rattachés au laboratoire MESOPOLHIS (Centre méditerranéen de sociologie, de science politique et d'histoire), accueilli à la MMSH et utilisateur fréquent des services de la PUD.

Depuis 2017, ce réseau de chercheurs internationaux s'est doté d'une plateforme numérique technologique d'accès aux données démographiques structurées autour de ses axes de recherche. Son site internet permet une visualisation graphique, cartographique et tabulaire interactive d'indicateurs démographiques sur les populations méditerranéennes. Cet outil innovant consiste en une base de données de type SQL, accessible gratuitement, et est capable d'intégrer des données, stockées sur des serveurs de l'IR* Huma-Num, ayant des structures originelles particulièrement diverses. L'outil est ainsi en mesure de gérer les transformations des unités territoriales et donc de varier selon le temps et l'échelle considérés. Cette conception FAIR permet à des néophytes de s'emparer facilement de l'outil. « Finalement, la base de données actuelles (*sic*) intègre donc tous les niveaux géographiques existant en Méditerranée dans 23 pays [...], à toutes les époques, en conservant la possibilité d'en intégrer de nouveaux sans restructurer la base existante. On peut donc la décrire comme structurellement résiliente au changement territorial et alors envisager

son utilisation dans une perspective plus historique le jour venu »⁴⁰. Elle s'adresse donc aussi bien aux chercheurs qu'à la société civile, qu'il s'agisse de citoyens, journalistes ou décideurs politiques.



Figure 13 : Page d'accueil du site de DemoMed, juillet 2024.
Source : <https://demomed.org/index.php/fr/>.

Seule plateforme sans personnel dédié, DemoMed fonctionne grâce à l'investissement de sa directrice. Elle remplit également une mission de formation à la recherche par la recherche en accueillant des étudiants dès la licence auxquels sont proposés des tutorats de stages, des suivis de dossiers et mémoires. Aux niveaux master et doctorat, les jeunes chercheurs sont intégrés aux axes de recherche par la participation à des publications et communications scientifiques. La plateforme bénéficie d'un financement annuel de la MMSH à hauteur de 2 500 €, somme qui permet de couvrir les frais de fonctionnement et de maintenance informatique. Toutefois, ce montant permet uniquement de maintenir la base de données et non de la développer. Ainsi, la directrice de DemoMed fait part de son besoin d'accueillir un ingénieur qui pourrait penser les développements technologiques de l'outil, tout en participant à la valorisation des données et des travaux qui en sont issus.

En conclusion, DemoMed est tant un collectif de chercheurs qu'une plateforme technologique. Ce projet allie transdisciplinarité, réflexions autour des évolutions des pratiques vers une meilleure compréhension et prise en compte des outils et méthodes appliqués dans les champs de recherche autour des populations, et innovation technologique. L'observatoire est également un projet d'humanités numériques et de science ouverte par sa grande accessibilité. Enfin, DemoMed est un dialogue entre « deux rigueurs » : celle de l'outillage informatique et celle de l'épistémologie propre à la recherche, à ses terrains et problèmes étudiés⁴¹.

⁴⁰ Oliveau S., Doignon Y. et Blöss-Widmer I., 2021, *op. cit.*

⁴¹ Boulaire C. et Carabelli R., 2017, « Du *digital naive* au bricoleur numérique : les images et le logiciel Omeka », in Cavalié É., et al., *Expérimenter les humanités numériques*, Montréal : PUM.

OUTILS

Les PUD⁴² utilisent une grande variété d'outils informatiques pour la gestion de leurs activités quotidiennes et pour les accompagnements individuels et collectifs qu'elles proposent. Le terme « outils » est ici entendu au sens que lui donne Jacomy⁴³, c'est-à-dire au sens technologique du terme, comme l'outillage ou l'appareillage scientifique sur lequel repose une partie des théories scientifiques.

Applications, progiciels⁴⁴, logiciels et outils informatiques

Dans un souci de correspondre à la politique de développement de la science ouverte prônée par les MSH, les PUD manient 53 % d'outils libres de droits. Plusieurs plateformes font part d'une baisse des demandes d'accès à SPSS, SAS et STATA⁴⁵. Toutefois, d'autres notent un maintien des usages de logiciels sous licence, comme la PUD Dijon qui signale même une hausse des demandes d'accès à NVivo et SAS. Bien que les principaux outils utilisés permettent de réaliser des analyses et traitements de données statistiques, leurs activités ne se limitent pas à cela.

- | | |
|--|--|
| • Affichage PDF | • Création de questionnaires en ligne |
| • Cartographie | • Réseau privé virtuel |
| • Contrôle de version | • Réunions en ligne et visioconférences |
| • Création et gestion de base de données | • Stockage |
| • Dessin vectoriel | • Traitement et analyse de données qualitatives, méthodes mixtes |
| • E-learning | • Traitement et analyse de données statistiques |
| • Gestion de codes | • Traitement et analyse de données textuelles, méthodes mixtes |
| • Présentations et diaporamas | • Traitement et analyse de données textuelles, méthodes mixtes |
| • Prise de notes | • Traitement de texte |
| • Programmation (et distribution) | • Transcription automatique |
| • Projection | |

Figure 14 : Liste des fonctions des outils et logiciels informatiques, par ordre alphabétique

⁴² La plateforme DemoMed proposant un système de base de données spécifique, elle n'a pas été incluse dans cette partie. Les PUD de Montpellier et de Paris-Saclay ne disposant pas d'ingénieur en poste au moment de la collecte de ces données, elles ont également été exclues. Les données ici présentées correspondent donc aux 14 autres plateformes du réseau Data.

⁴³ Jacomy M., 2024, « Compagnonnage, infrastructure et communs pour les ingés de recherche : un regard d'anthropologue et de créateur d'outils », Séminaire MetSem, 16.05.2024, <https://archive.org/details/metssem44>.

⁴⁴ Un progiciel est un « ensemble complet et documenté de programmes conçu pour être fourni à plusieurs utilisateurs en vue d'une même application ou d'une même fonction » (Journal Officiel du 17 janvier 1982, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000842020>).

⁴⁵ SPSS est un logiciel d'analyse statistique, SAS une plateforme d'analyse de données et STATA un logiciel de statistiques et d'économétrie. Les trois sont des solutions propriétaires.

Les outils les plus utilisés⁴⁶ par les PUD sont, par ordre de plateformes concernées⁴⁷ :

- Microsoft Excel et R/Rstudio pour l'analyse et le traitement de données statistiques (14 plateformes) ;
- Les autres applications du pack Microsoft Office, soit PowerPoint et Word (14 plateformes) ;
- LimeSurvey pour l'établissement de questionnaires en ligne (12 plateformes) ;
- Zotero pour la création et la gestion de bases de données bibliographiques (11 plateformes) ;
- Zoom pour la tenue de réunions en ligne et de visioconférences (10 plateformes) ;
- IRaMuTeQ pour l'analyse et le traitement statistique de données textuelles (9 plateformes) ;
- QGIS pour des usages cartographiques (8 plateformes) ;
- GitHub/GitLab pour des contrôles de versions (8 plateformes).

Hormis les applications du Pack Office, l'ensemble de ces outils sont libres de droits. Si les outils les plus utilisés, Excel et R/Rstudio, sont directement liés à l'ADN originel des PUD, soit les données quantitatives, nombre d'entre elles ont élargi leur offre de services en proposant également de travailler à partir d'outils d'analyse de données textuelles comme IRaMuTeQ. Elles ont ainsi développé une palette de compétences correspondant à une acception élargie de la notion de données quantitatives qui inclue le traitement statistique de (corpus de) textes. La PUD Dijon a également pris ce tournant en étant la seule plateforme à utiliser NVivo, progiciel informatique d'analyse de données qualitatives aidant à la recherche par méthodes qualitatives et mixtes. Le même choix a été effectué par la plateforme PANELS avec le logiciel Atlas.ti. La plateforme PUD-Grenoble Alpes a pleinement investi ce choix avec la création d'un outil de transcription automatique libre de droits pour les chercheurs rattachés à la MSH-Alpes (Taddam).

Enfin, le réseau Data se caractérise par sa grande variété de compétences logicielles spécifiques à son champ de métier. En excluant les outils de bureautique classiques, ce ne sont pas moins de 23 applications, logiciels et progiciels qui sont utilisés au sein des plateformes, chacun proposant des modalités de collecte, traitement, analyse et visualisation des données distinctes. Parmi ceux-là, 10 sont spécifiquement dédiés au traitement et à l'analyse de données statistiques. Cette diversité permet de proposer aux chercheurs des accompagnements personnalisés de pointe dans le développement de leurs travaux. Parmi ces 23 outils, 15 sont libres de droits, inscrivant pleinement les plateformes dans la politique de développement de la science ouverte impulsée par les MSH.

⁴⁶ Soit les outils et logiciels utilisés par plus de la moitié des plateformes répondantes.

⁴⁷ Pour plus d'informations : RnMSH, « Les plateformes technologiques du RnMSH », https://heurist.huma-num.fr/heurist/?db=RnMSH_Plateformes&website&id=761&lang=FR&pageid=528.

Outils	Fonction	Plateformes
Eval&GO	Questionnaire en ligne	1
GitHub/GitLab	Contrôle de version	8
Heurist	Création et gestion de base de données relationnelle	2
InkScape	Dessin vectoriel	2
IRaMuTeQ	Traitement et analyse de données textuelles pour méthodes mixtes	9
Jamovi	Traitement et analyse de données statistiques	6
Jupyter Notebook	Gestion de codes	4
LibreOffice Calc	Traitement et analyse de données statistiques	4
LimeSurvey	Questionnaire en ligne	12
Microsoft Excel	Traitement et analyse de données statistiques	14
Microsoft Power BI	Traitement et analyse de données statistiques	1
NVivo	Traitement et analyse de données qualitatives pour méthodes mixtes	1
Philcarto	Cartographie	3
PSPP	Traitement et analyse de données statistiques	4
Python	Programmation et distribution	7
QGIS	Cartographie	8
R/Rstudio	Traitement et analyse de données statistiques	14
Sphinx	Traitement et analyse de données statistiques	2
SPSS	Traitement et analyse de données statistiques	3
STATA	Traitement et analyse de données statistiques	3
TADDDAM	Transcription automatique	1
XLSTAT	Traitement et analyse de données statistiques	2
Zotero	Création et gestion de base de données bibliographiques	11

Figure 15 : Liste des outils métiers utilisés par les plateformes, par ordre alphabétique.

SD-Box : un environnement numérique sécurisé

La SD-Box est un matériel d'accès à l'infrastructure centrale du CASD (Centre d'Accès Sécurisé aux Données)⁴⁸. Disposant d'un système d'authentification poussé, il permet d'accéder, de façon sécurisée, à 512 sources de données sensibles par un système de verrouillage physique (biométrie) empêchant la récupération de fichiers.

⁴⁸ Le CASD est un groupement d'intérêt public (GIP) créé en 2018 et rassemblant l'INSEE (en tant que représentant de l'État), le groupe des écoles nationales d'économie et statistique (GENES), le CNRS, l'École polytechnique, HEC Paris et la Banque de France. Pour plus d'informations sur sa mission et son mode de gouvernance : <https://www.casd.eu/le-centre-daccs-securise-aux-donnees-casd/gouvernance-et-missions/>.

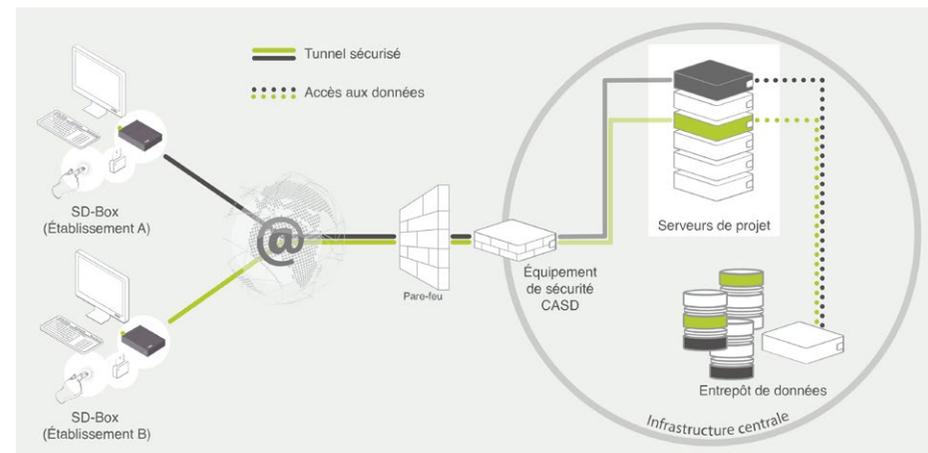


Figure 16 : Schéma de fonctionnement du système de SD-Box
Source : www.casd.eu⁴⁹

Il permet donc à des membres d'équipes de recherche de travailler sur des données sensibles inaccessibles par ailleurs. Les données accessibles par le biais des SD-Box ne peuvent être extraites du serveur et nécessitent des aménagements spécifiques visant à les sécuriser.

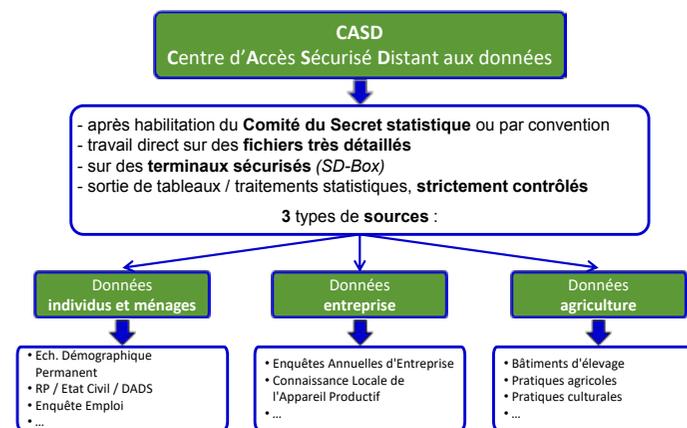


Figure 17 : Schéma d'accès aux données du CASD
Source : De Belsunce 2021⁵⁰

Cinq plateformes mettent une à trois SD-Box à disposition de leurs utilisateurs : les PUD-Bretagne, PUD-Bordeaux, PUD Dijon, PUD Lille et PUD-Strasbourg. Elles font face à des volumes d'utilisation très variés. Lorsqu'elle en avait une, PROGEDO-Loire recensait

⁴⁹ CASD, « Infrastructure », <https://www.casd.eu/technologie/infrastructure/>.

⁵⁰ De Belsunce C., 2021, « Comment trouver et accéder à mes données ? Recensement et présentation des sources et producteurs de données », Semaine Data-SHS, déc. 2021, <https://amu.hal.science/hal-03524177>.

moins d'une utilisation par an, alors que la PUD Lille en dénombre entre trois et quatre selon les années. La PUD-Strasbourg souhaiterait obtenir une troisième SD-Box afin de répondre aux demandes des doctorants. La PUD-Aix-Marseille Université ou la PUD Nanterre ont renvoyé les SD-Box du fait de l'absence de demande. En effet, sur ces campus universitaires, les personnes ayant besoin d'une SD-Box en disposent directement dans leurs laboratoires de rattachement. Ces volumes d'utilisation variés ont comme corollaire des stratégies financières distinctes, adaptées aux problématiques locales.

L'accès aux données du CASD nécessite deux abonnements : un abonnement à la SD-Box (pris en charge par la plateforme et d'un montant de 37 €/mois⁵¹) et un abonnement permettant aux chercheurs d'accéder aux bases de données (variable⁵²). Face aux difficultés financières des plateformes, l'arrêt de l'abonnement est parfois présenté comme une solution simple pour réaliser des économies, comme cela a été le cas pour la PUD-Aix-Marseille Université. À l'inverse, dans d'autres plateformes disposant de ressources plus importantes, une aide financière est apportée aux chercheurs ayant besoin d'un accès aux données du CASD. La PUD Lille, par exemple, propose une prise en charge des frais d'accès aux abonnements pour les doctorants, en contrepartie d'un soutien apporté à la PUDL par leur laboratoire de rattachement. Sur un principe similaire, la PUD-Strasbourg propose aux doctorants une prise en charge de leurs frais d'abonnement. Cette question de l'abonnement représente une contrainte réelle pour les PUD, certaines souhaitant en bénéficier de façon pérenne au motif que « l'offre crée la demande ». Ainsi, un ingénieur a soulevé la question des conditions partenariales qui lient l'IR* PROGEDO au CASD, mentionnant qu'un tel partenariat devrait permettre aux PUD de disposer de SD-Box sans s'acquitter de l'abonnement mensuel afin que les plateformes deviennent des points d'accès de référence aux données du CASD.

⁵¹ CASD, « Tarifs », <https://www.casd.eu/tarifs-2/>.

⁵² À titre d'exemple, un accès de 12 mois à des données de santé de taille modeste, pour deux chercheurs dont un doctorant, représente un coût d'environ 4 500 €. Pour plus d'informations, consulter le simulateur en ligne du CASD : <https://www.casd.eu/estimation-de-cout/>.

IV - PARTENARIATS

Les plateformes Data s'intègrent pleinement dans la mission de coopération inter-institutionnelle des MSH. Les principales institutions qui leur sont liées leur apportent également une coloration particulière : les MSH, l'IR* PROGEDO et les universités. Pour le résumer, ainsi que cela a été formulé par plusieurs ingénieurs interrogés, PROGEDO fournit le sens de la mission et coordonne les activités, les universités sont le terrain sur lequel celles-ci sont déployées ainsi que les correspondants purement administratifs, et les MSH sont le cadre de travail et le soutien quotidien des membres du personnel des plateformes Data. Cependant, les partenariats de ces plateformes ne peuvent se résumer à ce schéma tripartite car elles participent à d'autres projets, faisant intervenir des acteurs divers, répondant ainsi à leur mission première d'implantation au cœur d'un tissu scientifique local. Au niveau national, les autres principaux partenaires des plateformes Data sont Mate-SHS du CNRS SHS et différents organes liés aux données quantitatives tels que l'INED, l'ADISP ou l'INSEE. L'URFIST est également un important partenaire de plusieurs plateformes qui organisent des formations dans ce cadre. Au niveau local, les plateformes sont intégrées aux différents réseaux pertinents auxquels appartiennent les MSH, aux Ateliers de la donnée et à tous les consortiums traitant des données. Elles travaillent également de façon rapprochée avec les laboratoires rattachés à leur MSH, les écoles doctorales, les SCD et les universités de leurs sites.

UN PARTENAIRE PRIVILÉGIÉ : L'IR* PROGEDO

L'IR* PROGEDO est dirigée par Nicolas Sauger depuis le 1er janvier 2024. Depuis 2022, elle est labellisée centre de référence thématique de Recherche Data Gouv. L'objectif de l'IR* PROGEDO est « de développer la culture des données, d'impulser et de structurer une politique des données d'enquêtes pour la recherche en sciences sociales »⁵³. Créée en 2008, elle vise non seulement à diffuser des bases de données quantitatives, mais également à les rendre « intelligibles »⁵⁴ auprès de la communauté scientifique des sciences humaines et sociales. Elle soutient et coordonne la réalisation de grandes enquêtes européennes et la mise à disposition de données issues d'enquêtes nationales. Elle s'est donné comme mission principale d'assurer la préservation et l'accès sécurisé aux données quantitatives en SHS françaises (documentation rigoureuse des métadonnées, respect du RGPD et des standards internationaux, promotion de l'utilisation de formats réutilisables). Elle a vocation à structurer les réseaux de chercheurs SHS en données quantitatives, tout en promouvant les bonnes pratiques quant à leur traitement, analyse et conservation. Elle propose des accès à trois grandes catégories de catalogues de jeux de données pseudonymisées, à usage scientifique uniquement :

Quetelet-Progedo-Diffusion : entrepôt de l'IR* PROGEDO, financé par le MESR et le CNRS, en collaboration avec l'ADISP, qui diffuse plus de 1 600 jeux de données issus de l'INSEE et de l'ADISP, avec des données allant de 1954 à nos jours pour la France. Il repose sur les serveurs de l'IR* Huma-Num ;

⁵³ PROGEDO, « Organisation et gouvernance », <https://www.progedo.fr/progedo/organisation-et-gouvernance/>.

⁵⁴ Zurbach J., 2023, « PROGEDO au service de l'intelligibilité des données quantitatives », *Communication & organisation*, no 64.

- Le catalogue du CASD : accès à des jeux de données confidentielles ;
- Enquêtes internationales : CDC, DBnomics, ESS, EVS, GGP, ICPSR, ISSP, LIS, SHARE.

PROJETS IMPLIQUANT PLUSIEURS MSH

Du fait de leur réseau très structuré, les plateformes labellisées PUD par l'IR* PROGEDO ont de nombreux modes de collaboration. De ce fait, les projets impliquant des plateformes de plusieurs MSH sont plus perçus comme des activités inter-PUD qu'inter-MSH car ils n'impliquent que peu, voire pas, d'autres réseaux de plateformes. Ces projets sont impulsés et gérés par les ingénieurs eux-mêmes, véritable pierre angulaire de ce réseau. Le fait que deux postes d'ingénieurs soient financés par l'*EquipEx+ : LifeObs* crée d'importants liens entre la PUD-Bordeaux et la PUD-Strasbourg. La Semaine Data-SHS est également un bel exemple du travail inter-plateformes avec un travail collaboratif sur les thématiques à développer chaque année, des partages de retours d'expérience, ainsi qu'une organisation de sessions communes lorsque cela est possible. Autre projet important au sein des PUD : la constitution de kits pédagogiques. Ce projet implique notamment la PUD-Bordeaux, PUD Nanterre, PUD-Strasbourg et PUD-Toulouse, ainsi que le CDSP. Ces kits visent à mettre en valeur des jeux de données issus de grandes enquêtes telles que l'enquête ERFI - GGS (Étude des relations familiales et intergénérationnelles - *Generations and Gender Survey*) de l'INED par exemple. Ce sont des fiches d'auto-formation sur des méthodes statistiques appliquées à des jeux de données anonymisés et diffusables. Ce projet vise à permettre à un public débutant d'acquérir quelques notions de base en traitement statistique ainsi qu'à dédramatiser les logiciels de traitement. Ces kits pédagogiques seront publiés en accès libre et visent à devenir une porte d'entrée vers les données quantitatives. Les projets financés par l'ANR sont une autre opportunité de mutualiser les compétences techniques et disciplinaires des ingénieurs lorsque plusieurs sites sont engagés dans le projet. C'est notamment le cas de l'ANR VICO « La vie en confinement » au sein duquel de nombreuses plateformes sont impliquées (PUD-Aix-Marseille Université, PUD-Grenoble Alpes, PUD-Toulouse...). Enfin, le collectif POF (Paris-Ouest-France), initié par des enseignants-chercheurs de Paris Nanterre et de Paris 8 regroupe désormais une dizaine d'universités, avec l'aide et l'accompagnement de plusieurs plateformes. Ce collectif est né d'une volonté de mutualiser des méthodes d'enseignement de méthodes d'enquête statistique en Licence. Il consiste en la création, passation et analyse d'un questionnaire par des étudiants des universités impliquées. Les questionnaires sont diffusés dans des salles de classe choisies aléatoirement sur chaque site et permettent aux étudiants de travailler sur la quasi-totalité du cycle de vie de la donnée en situation d'enquête réelle. Ce ne sont là que quelques exemples des modalités de collaboration déployées par les plateformes au sein du réseau Data. Cependant, rares sont les exemples de collaborations formalisées entre des plateformes technologiques d'une même MSH ou entre des plateformes Data et des pôles d'autres réseaux de plateformes du RnMSH. Les activités les plus fructueuses semblent se faire avec les référents Huma-Num, mais ne sont souvent que ponctuelles malgré une volonté de plusieurs MSH de les regrouper au sein d'un seul et même service.

INSTITUTIONS ET PARTENAIRES LOCAUX

Du fait de leur fort ancrage local, les MSH et les plateformes ont développé des réseaux régionaux qui leur permettent, par exemple, de financer des seconds postes d'ingénieurs au sein des plateformes, ou d'obtenir un budget de fonctionnement. Bien que les plateformes privilégient les partenaires du monde de l'ESR, il existe quelques exceptions à cette règle. La PUD Dijon est à ce titre un exemple intéressant. Labellisée par le Comité d'Orientation Stratégique (COS) des Plateformes de Bourgogne (Franche-Comté (BFC)), la plateforme peut participer aux AAP de la région. Ce mode de gestion présente l'avantage d'implanter la plateforme comme un acteur incontournable du paysage scientifique régional, tout en mettant au service des collectivités territoriales les compétences présentes dans la MSH. Cependant, ce modèle a également des limites. La vocation originelle des plateformes est de soutenir la recherche en SHS, or mener des projets pour le compte des collectivités est particulièrement chronophage et réduit, de fait, le temps disponible pour l'accompagnement des chercheurs. Bien que conçue comme une solution pragmatique face au manque de ressources disponibles dans le secteur SHS, elle est loin de faire l'unanimité tant auprès des ingénieurs que des RS. La principale crainte exprimée est de voir les plateformes Data devenir des cabinets d'études aux tarifs moins élevés que ceux du secteur privé. Particulièrement attachés à leur mission d'appui et de recherche, plusieurs ingénieurs envisagent de quitter leur poste si ce modèle devait se généraliser.

D'autres plateformes bénéficient de fonds externes visant généralement à financer un second poste d'ingénieur. Ces financements sont souvent issus de projets ANR dans lesquels les plateformes ont été intégrées en amont du dépôt de dossier (PUD-Toulouse, par exemple). Ces projets sont accompagnés par la MSH et peuvent donc intégrer plusieurs plateformes au sein d'une même MSH. Un autre cas de figure existe avec des partenariats dédiés à et créés par la plateforme Data. C'est le cas notamment de la Fondation Pierre Deniker qui finance des projets de recherche sur la santé mentale et avec laquelle l'ingénieur de la PUD Nanterre a entièrement construit un projet de mécénat⁵⁵.

Un acteur complémentaire localement : le réseau des Ateliers de la donnée

Depuis 2021, les Ateliers de la donnée sont le maillon local de l'écosystème Recherche Data Gouv⁵⁶ dont ils sont les relais locaux, déployés dans le cadre du deuxième PNSO en partenariat avec l'INRAE. Ils sont portés par les universités, hormis l'Atelier de la Donnée Centre-Val de Loire qui est porté par la MSH locale. Ils visent à accompagner les équipes de recherche en leur apportant « une première expertise dans la gestion raisonnée des données de recherche »⁵⁷. Il s'agit de soutenir la communauté universitaire dans la mise en place de bonnes pratiques quant à la gestion de ces données, afin de systématiser le travail d'ouverture des données, codes et logiciels. Ce système d'accompagnement est complété par l'entrepôt Recherche Data Gouv. Fin juin 2024, le maillage territorial comportait vingt Ateliers de la donnée, dont seize Ateliers labellisés et quatre « sur la trajectoire de labellisation ». En tant que relais de l'IR* PROGEDO, les PUD sont membres des Ateliers de la donnée sur leurs sites, participant ainsi à la structuration locale des activités liées aux données quantitatives.

⁵⁵ Chauvel B., 04.10.2024, « Appel à questions : "Accompagnement et représentations sociales des maladies neuro-évolutives", PROGEDO, DOI : 10.58079/12f8o.

⁵⁶ <https://recherche.data.gouv.fr/fr>.

⁵⁷ Ouvrir la science!, « Recherche Data Gouv », <https://www.ovvirlascience.fr/recherche-data-gouv/>.

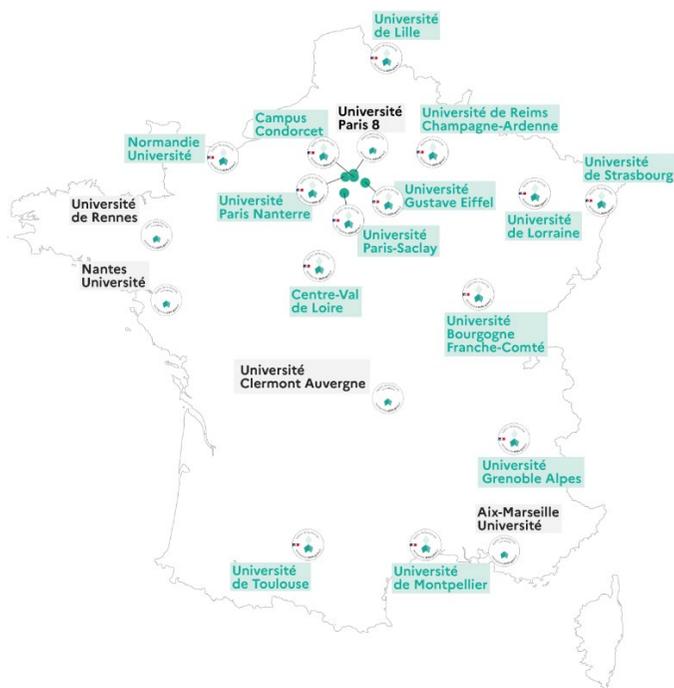


Figure 18 : Répartition géographique des Ateliers de la donnée, juin 2024.
Source : Recherche Data Gouv58

Les Ateliers fonctionnent selon un système de guichet unique, coordonné par les SCD, qui redirige les demandes de la communauté scientifique vers les acteurs locaux compétents. Les SCD prennent en charge les formations pour débutants, toutes disciplines confondues. Les PUD sont au cœur de ce dispositif en apportant l'expertise technique de l'IR* PROGEDO et la réflexivité scientifique des MSH. Ce nouvel acteur des données a pu être vu comme concurrent lors de son déploiement, mais s'est finalement révélé un bon complément en permettant aux MSH de se recentrer sur des accompagnements nécessitant une haute expertise technique et scientifique. Ce principe de départ supposait que la création des Ateliers ne représenterait pas de tâche supplémentaire pour les ingénieurs Data, tout en augmentant leur visibilité. Ce dernier point est difficile à mesurer, mais l'enquête de terrain n'a pas fait ressortir de réel changement du flux après l'arrivée d'un Atelier de la donnée sur un territoire. De plus, l'implication des ingénieurs dans les comités de pilotage représente une charge de travail supplémentaire, de l'ordre de quelques heures par mois, qui participe toutefois de leur mission de constitution de réseaux locaux autour des données quantitatives. Au moment de la rédaction du rapport, peu d'Ateliers étaient réellement actifs et bien implantés, la diversité des acteurs impliqués freinant la mise en place concrète du guichet unique. Enfin, les SCD ont bénéficié d'apports financiers liés à la coordination de

ces Ateliers qui bénéficient d'un nombre d'ETP conséquent (5,81 ETP⁵⁹), bien que réparti entre de très nombreux acteurs. L'écart d'ETP entre les Ateliers et les plateformes Data pourrait, à terme, fragiliser ces dernières en absorbant les fonds disponibles sur ces sujets, au détriment de l'accompagnement spécifique aux chercheurs en SHS. Cependant, le grand nombre d'acteurs impliqués dans les Ateliers entraîne une certaine rigidité de fonctionnement, là où le réseau Data est plus souple.

PARTENARIATS INTERNATIONAUX

En lien avec la politique d'internationalisation du RnMSH, certaines plateformes développent des partenariats hors des frontières françaises. Le séminaire sur les méthodes mixtes de recherche en SHS de la PUDL, par exemple, a une attractivité internationale et s'est construit suite à des formations suivies par l'ingénieur hors de France. En effet, ce type de méthodes de recherche étant encore assez peu développé, et surtout théorisé, en France, ce programme de séminaires se veut une référence, complétée par l'école d'été à laquelle la PUD Bretagne participe. Toutefois, du fait de la vocation locale des PUD, les partenariats internationaux sont assez peu développés par ce réseau. À l'inverse, de par sa nature même, DemoMed est un acteur central de la politique d'internationalisation du réseau Data. Cette plateforme a développé des partenariats privilégiés avec l'INED d'une part, et avec plusieurs universités du pourtour méditerranéen d'autre part : Université Complutense de Madrid (Espagne), Université La Sapienza de Rome (Italie), Université de Thessalie à Volos (Grèce), Université de Zadar (Croatie) et The American University in Cairo (Égypte). Malgré ces efforts de tendre vers une internationalisation des activités et à quelques rares exceptions près (participation à l'Inter-university Consortium for Political and Social Research par exemple), les infrastructures de recherche qui coordonnent les plateformes Data (RnMSH et IR* PROGEDO) ne disposent pas d'une masse salariale suffisante pour accompagner les plateformes vers un niveau d'expertise européen ou international.

58 Recherche.data.gouv, « Ateliers de la donnée : des services généralistes pour tout le territoire », <https://recherche.data.gouv.fr/fr/ateliers-de-la-donnee>.

59 Moyenne calculée à partir des données disponibles pour 12 Ateliers de la donnée en juillet 2024, *ibidem*.

V- ENJEUX ET PERSPECTIVES

FINANCEMENT ET PÉRENNISATION DES PLATEFORMES

Les données et leur FAIRisation sont devenus un enjeu politique central pour les organismes de recherche français et européens, notamment dans le cadre du développement d'une politique science ouverte affichée comme « priorité ministérielle » par le MESR⁶⁰. Le déploiement des Ateliers de la donnée dans le cadre du PNSO 2021-2024 a été un des axes majeurs développés dans ce cadre par le précédent gouvernement. En parallèle à cela, les infrastructures de recherche en SHS ont été mises en avant par le CNRS comme des axes centraux de la structuration de la recherche scientifique française et européenne. Leur financement et la pérennisation de leurs actions a été pointé comme un enjeu crucial de cet objectif : « [Les IR] sont les instruments d'une stratégie scientifique européenne qui s'inscrit dans le temps long et ont besoin de sécurisation et sanctuarisation budgétaire »⁶¹. Paolo Laj, coordinateur du groupe de travail CNRS dédié aux infrastructures de recherche dans le cadre européen ajoute : « Il est donc essentiel de rappeler que ces infrastructures sont des outils de recherche essentiels pour pouvoir répondre aux questions scientifiques de demain et qui doivent rester pérennes »⁶². Les plateformes Data s'inscrivent donc pleinement dans ces objectifs nationaux en étant tout à la fois les représentantes d'infrastructures de recherche au niveau local et en promouvant la FAIRisation des données.

Pourtant, ces plateformes fonctionnent avec un nombre conséquent de postes non pérennes ainsi qu'avec des budgets de fonctionnement extrêmement limités. Ces faibles dotations ne permettent pas aux ingénieurs en poste de se former régulièrement alors même que leur domaine d'activité est en évolution constante et en lien avec des potentiels d'innovation forts. Cette fragilité est accrue par la proportion importante de postes d'ingénieurs financés par les universités, employeur quasi exclusif au sein du réseau Data. Ce manque de diversité réduit la flexibilité potentielle des plateformes et les soumet à des impératifs politiques qui ne permettent pas d'assurer un accompagnement à la recherche sur le temps long. Qui plus est, le statut de contractuel à durée déterminée d'une portion non négligeable d'ingénieurs (76 %) freine la mise en place de stratégies d'action à long terme, problématique qui se généralise de plus en plus dans le milieu de l'enseignement supérieur et de la recherche : « [...] le caractère bénévole et enthousiaste de la recherche est de plus en plus mis en difficulté par les contrats proposés dans l'Enseignement supérieur et la recherche – contrats qui, quant à eux, sont de moins en moins pérennes »⁶³. Cette fragilité entraîne un fort sentiment de découragement tant auprès des ingénieurs que des référents scientifiques et des directions de MSH. Le référent scientifique d'une plateforme

Data résume ainsi ce sentiment général : « c'est un peu dommage de se dire qu'il faut éviter de fermer ». Enfin, l'absence de pérennité des postes ne permet pas d'envisager de stratégie de développement à moyen ou long terme pour la plateforme.

VALORISATION DES ACTIONS MENÉES PAR LES PLATEFORMES DATA

Comme pour les autres réseaux de plateformes du RnMSH, les actions de communication sont gérées au niveau local et semblent montrer leur efficacité. Certains sites se sont saisis de cet enjeu et ont engagé des actions visant à pallier ce manque. Plusieurs MSH, par exemple, ont installé des affiches de présentation de leurs plateformes technologiques dans le hall d'accueil afin de valoriser leur travail auprès des usagers de la MSH. Toutefois, il n'y a que peu, voire pas, d'harmonisation nationale de la stratégie en matière de communication. En premier lieu, le RnMSH ne dispose pas de moyens suffisants pour se saisir pleinement de cet enjeu, d'autant plus que cet axe n'est pour l'instant pas défini comme une priorité stratégique. Pourtant, l'éclatement des moyens de communication pourrait freiner la mise en place d'une stratégie nationale de valorisation des activités des plateformes.

MAISON DES SCIENCES DE L'HOMME DE DIJON

PLATEFORME HUMANITÉS NUMÉRIQUES

LE PÔLE PUDD

Plateforme universitaire de données de Dijon

Objectifs

La plateforme Universitaire de Données de Dijon est une composante de la plateforme Humanités Numériques de la MSH de Dijon. Elle accompagne les chercheurs, doctorants et étudiants en sciences humaines et sociales dans des projets de recherche nécessitant la collecte et l'analyse de données quantitatives.

Au service des 16 laboratoires en Sciences Humaines et Sociales du Grand campus dijonnais, la PUDD intervient en appui aux programmes de recherche nécessitant la production, l'analyse et l'archivage de données quantitatives.

Comme toutes les PUDD en France, la PUDD prolonge au niveau régional l'action de l'IR* PROGEDO, infrastructure nationale qui a pour but de développer la culture des données, d'imposer et structurer une politique des données d'emplais pour la recherche en sciences sociales.

Missions

- collecte, traitement et exploitation statistique de données
- production de bases de données
- accompagnement sur le plan de gestion des données et le respect du RGPD
- formation aux logiciels d'analyse statistique
- réalisation de prestations de services

Compétences

- Expertise en traitement et analyse de données d'enquêtes.
- Structuration et gestion de bases de données : diffusion, rédaction d'un plan de gestion.
- Capacité à mettre en place un protocole d'analyses multivariées des données.
- Connaissance et accès à des données statistiques nationales et internationales.
- Accompagnement dans la production de données d'enquêtes.
- Développement d'outils pour la collecte et l'archivage des données.

Équipe

L'équipe du pôle PUDD est composée de statisticiens et ingénieurs en traitement et analyse de données.

Participation à des réseaux nationaux

- le pôle PUDD est partenaire et labellisé par l'IR* PROGEDO.
- Réseau QUETLET : réseau français des centres de données pour les sciences humaines.

Matériels et logiciels

- SAS, R Studio, R, SPSS, STATA, NVIVO et les logiciels disponibles au CCUB.
- Photos de travail, les CASO.

Formations

- Outils d'analyse statistique.
- Langages et techniques de programmation.
- Méthodes quantitatives.

SITE WEB & CONTACT :
mah-dijon@bourgnies.fr
pudd.mahdijon@u-bourgogne.fr

Figure 19 : Affiche présentant la PUDD, hall d'accueil de la MSH Dijon.

60 MESR, « Priorités ministérielles », <https://www.enseignementsup-recherche.gouv.fr/fr/science-ouverte-50360>.

61 CNRS Hebdo, « Les infrastructures sont un pilier essentiel pour la stratégie de recherche européenne », 07.06.2024, <https://www.cnrs.fr/fr/actualite/les-infrastructures-sont-un-pilier-essentiel-pour-la-strategie-de-recherche-europeenne>.

62 Ibidem.

63 Saint-Raymond L., 2023, « Digital Viau : histoire et aléas d'un projet en humanités numériques », *Les Cahiers de Framespa* [En ligne], 42, DOI : 10.4000/framespa.14121.

DemoMed, du fait de son réseau de chercheurs actifs, comptabilise 12 publications de rang A pour la période 2020-2023. Le site internet de la plateforme est automatiquement actualisé dès qu'une publication mentionnant la plateforme est disponible, ce qui visibilise fortement le travail des chercheurs de l'observatoire. D'autres plateformes ont investi HAL pour mettre en avant tant les communications scientifiques des ingénieurs au nom de la plateforme que les présentations réalisées par des intervenants extérieurs. C'est le cas, par exemple, de la PUDN qui s'est dotée, en lien avec le pôle Humanités Numériques de la MSH Mondes, d'une collection HAL faisant apparaître son logo et celui de ses tutelles et partenaires (Université Paris Nanterre, MSH-Mondes et IR* PROGEDO). Vingt-deux notices et sept documents y ont été déposés. Des onglets permettent de naviguer dans la collection : accès aux documents, historique des programmes des Semaines Data-SHS et lien vers les dépôts de vidéos sur Canal U. Cette collection HAL permet de mettre en avant les communications de ses ingénieurs, ainsi que les différents séminaires thématiques organisés par la plateforme (données sur les migrations, données localisées ou données historiques par exemple). Certaines MSH réfléchissent actuellement à une meilleure valorisation du travail des ingénieurs dans les projets de recherche, notamment lorsque ceux-ci ont un apport scientifique et technique important dans le projet. Le co-autorat pourrait ainsi être un mode de valorisation de cet engagement, mais il n'est pour l'instant lié qu'à la bonne volonté des chercheurs. Or, faire apparaître le nom des ingénieurs dans les publications scientifiques est non seulement un atout pour leur carrière individuelle, mais aussi une façon de valoriser le haut niveau d'expertise technique et scientifique des plateformes.

Ces initiatives locales sont complétées par une réflexion portée par le RnMSH sur la valorisation des services aux chercheurs proposés par les plateformes technologiques. Ce travail a abouti à la constitution et mise en ligne d'une base de données Heurist⁶⁴ regroupant l'ensemble des plateformes technologiques des cinq réseaux (Audio-Visio, Cogito, Data, Scripto et Spatio) et dans laquelle se trouvent, entre autres, la liste des équipements, la liste des logiciels sur lesquels des compétences spécifiques sont disponibles et les différents services proposés par chaque plateforme.

CONCLUSION

Le réseau Data a une identité particulière dans les réseaux de plateformes technologiques du RnMSH. Par son articulation tripartite entre le réseau et les MSH, l'IR* PROGEDO et les universités, il pourrait occuper une place centrale dans la stratégie politique déployée par les MSH et le RnMSH. À cela s'ajoute le fait qu'il prend en charge un domaine en pleine expansion dans l'ESR en se situant au carrefour des enjeux, nombreux, liés aux données et à la science ouverte. Aujourd'hui essentiellement piloté par l'IR* PROGEDO, le réseau Data bénéficierait d'une meilleure implantation au sein du RnMSH, ses domaines de compétences étant transversaux et donc en lien avec les besoins des autres réseaux de plateformes. À ce sujet, nombreux sont les ingénieurs à avoir exprimé leur envie de mieux connaître le RnMSH et ses membres et personnels.

Toutefois, ces besoins de mise en réseau sont freinés par la grande fragilité financière de ce réseau. La majorité des plateformes qui le constituent fonctionnent avec des personnels précaires, voire avec aucun personnel dédié à DemoMed et, trop souvent, des budgets insuffisants. Il s'agit pourtant de plateformes ayant besoin de frais de fonctionnement limités car elles n'ont besoin que d'ordinateurs et de quelques licences propriétaires pour mener à bien leurs activités. À cela s'ajoutent les déplacements nécessaires aux ingénieurs pour des formations diverses, dépense nécessaire au vu de la rapidité des évolutions technologiques dans leur secteur. Certains ingénieurs sont d'ailleurs en demande d'un élargissement des formations disponibles, notamment sur la question de l'impact écologique des données. Le plus gros poste de dépense est, par conséquent, lié à la masse salariale. Si les fonds du PNSO ont effectivement permis d'amorcer les activités de très nombreuses plateformes, les universités tutélaires n'ont pas toujours pérennisé les postes après épuisement de l'enveloppe Vidal, malgré leur investissement financier non négligeable sur ces profils BAP D habituellement peu, voire pas, présents dans les universités. Cette fragilité financière pousse les ingénieurs à s'investir dans des activités de recherches de financements qui représentent autant de temps en moins pour assurer leur mission de soutien aux chercheurs. La précarité imposée aux ingénieurs est en opposition totale avec leur investissement personnel et professionnel sur des postes nécessitant une grande expertise technique, voire scientifique.

Enfin, malgré son apparente homogénéité et un socle d'activités communes bien identifiées, le réseau Data est, dans les faits, extrêmement divers, ce qui en fait probablement sa richesse. Toutefois, il pâtit de relations encore en construction entre ses deux organes structurants, l'IR* PROGEDO et le RnMSH. Ces difficultés ont été accentuées par le dynamisme avec lequel des plateformes se sont créées ces dernières années, laissant assez peu de temps et d'espace à une réflexion plus globale sur l'animation de ce réseau. Ces travaux sont désormais en cours et devraient également permettre de construire une réelle politique de communication et de valorisation autour des activités des plateformes.

⁶⁴ RnMSH, « Les plateformes technologiques du RnMSH », https://heurist.huma-num.fr/heurist/?db=RnMSH_Plateformes&website&id=761&lang=FR.

ANNEXES

LISTES DES SIGLES ET ACRONYMES	50	Plateforme Universitaire de Données Clermont-Auvergne (PUD-CA)	
SIGLES DES MSH.....	50	MSH Clermont.....	56
SIGLES ET ACRONYMES DES PLATEFORMES.....	51	Plateforme Universitaire de Données de Dijon (PUDD)	
SIGLES ET ACRONYMES DES INSTITUTIONS, LABORATOIRES ET SERVICES.....	52	MSH Dijon.....	56
AUTRES SIGLES.....	53	Plateforme Universitaire de Données Grenoble-Alpes (PUD-GA)	
PRÉSENTATION DES PLATEFORMES DATA	54	MSH-Alpes.....	57
DemoMed		Plateforme Universitaire de Données de Lille (PUDL)	
MMSH.....	54	MESHS.....	57
PANELS		Plateforme Universitaire de Données de Nanterre (PUDN)	
MSH-LSE.....	54	MSH Mondes.....	58
PROGEDO-Loire		Plateforme Universitaire de Données de Montpellier	
MSH Ange-Guépin.....	54	MSH-SUD.....	58
Plateforme Universitaire de Données d'Aix-Marseille Université (PUD-AMU)		Plateforme Universitaire de Données de Paris-Saclay (PUD-PS)	
MMSH.....	55	MSH Paris-Saclay.....	58
Plateforme Universitaire de Données de Bretagne (PUD-B)		Plateforme Universitaire de Données de Strasbourg (PUD-S)	
MSHB.....	55	MISHA.....	58
Plateforme Universitaire de Données de Bordeaux (PUD-Bx)		Plateforme Universitaire de Données de Toulouse (PUD-T)	
MSH Bx.....	55	MSHS-T.....	59
Plateforme Universitaire de Données de Caen (PUDC)		Plateforme Universitaire de Données de l'Université de Poitiers (PUD-UP)	
MRSH.....	56	MSHS Poitiers.....	59

LISTES DES SIGLES ET ACRONYMES

SIGLES DES MSH

MESHs	Maison Européenne des Sciences de l'Homme et de la Société (UAR 3185)
MISHA	Maison Interuniversitaire des Sciences de l'Homme - Alsace (UAR 3227)
MMSH	Maison Méditerranéenne des Sciences de l'Homme (UAR 3125)
MRSH	Maison de la Recherche en Sciences Humaines (UAR 3486)
MSH Ange-Guépin	Maison des Sciences de l'Homme Ange Guépin (UAR 3491)
MSH Clermont-Ferrand	Maison des Sciences de l'Homme de Clermont-Ferrand (UAR 3550)
MSH Dijon	Maison des Sciences de l'Homme de Dijon (UAR 3516)
MSH Mondes	Maison des Sciences de l'Homme Mondes (UAR 3225)
MSH Paris-Saclay	Maison des Sciences de l'Homme Paris-Saclay (UAR 3683)
MSH Poitiers	Maison des Sciences de l'Homme et de la Société de Poitiers (UAR 3565)
MSH SUD	Maison des Sciences de l'Homme Sciences et société Unies pour un autre Développement (UAR 2035)
MSH-Alpes	Maison des Sciences de l'Homme des Alpes (UAR 3394)
MSHB	Maison des Sciences de l'Homme en Bretagne (UAR 3549)
MSH-LSE	Maison des Sciences de l'Homme Lyon Saint-Étienne (UAR 2000)
MSHS-T	Maison des Sciences de l'Homme et de la Société de Toulouse (UAR 3414)

SIGLES ET ACRONYMES DES PLATEFORMES

DemoMed	Observatoire Démographique de la Méditerranée
PANELS	Plateforme technologique pour les sciences sociales
PUD-AMU	Plateforme universitaire de données d'Aix-Marseille Université
PUD-B	PUD de Bretagne
PUD-Bx	PUD de Bordeaux
PUDC	PUD de Caen
PUD-CA	PUD - Clermont-Auvergne
PUDD	PUD de Dijon
PUD-GA	PUD de Grenoble-Alpes
PUDL	PUD de Lille
PUDN	PUD de Nanterre
PUD-PS	PUD de Paris-Saclay
PUD-S	PUD de Strasbourg
PUD-T	PUD de Toulouse
PUD-UP	PUD de l'Université de Poitiers

SIGLES ET ACRONYMES DES INSTITUTIONS, LABORATOIRES ET SERVICES

AMI	Appel à manifestation d'intérêt
ANR	Agence nationale de la recherche
CASD	Centre d'Accès Sécurisé aux Données
CDGA	Cellule Data Grenoble Alpes
CDSP	Centre de données socio-politiques
Clersé	Centre lillois d'études et de recherches sociologiques et économiques
CNRS	Centre national de la recherche scientifique
CoDaPro	Collectif Data Provence
ED	École doctorale
EHESP	École des hautes études en santé publique
GATES	Grenoble ATtractiveness and ExcellenceS
GRICAD	Grenoble Alpes Recherche – Infrastructure de Calcul intensif et de Données
IFRESI	Institut Fédératif de Recherche sur les Économies et les Sociétés Industrielles
INED	Institut national d'études démographiques
INIST	Institut de l'information scientifique et technique
INSEE	Institut national de la statistique et des études économiques
INSPÉ	Institut universitaire de formation des maîtres
MESR	Ministère de l'Enseignement Supérieur et de la Recherche (depuis juin 2022)
MESRI	Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation
PROGEDO	PRoduction et GEstion des DOnnées
SCD	Service commun de documentation
SPD	Service protection des données
TSE	Toulouse School of Economics
ObARDI	Ontology-based Ancien Régime Data Infrastructure
UGA	Université Grenoble Alpes
URFIST	Unité régionale de formation à l'information scientifique et technique

AUTRES SIGLES

AAP	Appel à projets
ADISP	Archives de données issues de la statistique publique
ANF	Action Nationale de Formation
CDC	CESSDA Data Catalogue
CESSDA	Consortium européen des archives de données en sciences humaines et sociales
CPER	Contrat de plan État-Région
EC	Enseignant-chercheur
ESS	European Social Survey
EVS	European Values Study
FAIR	Faciles à (re)trouver ; Accessibles ; Interopérables ; Réutilisables
GGP	Generation and Gender Program
ICPSR	Inter-university Consortium for Political and Social research
ISSP	Internal Social Survey Programme
LIS	The Luxembourg Income Study
PGD	Plan de gestion des données
PNSO	Plan National pour la Science Ouverte
RGPD	Règlement général de protection des données
RS	Référent scientifique
SHARE	Survey of Health, Ageing and Retirement in Europe
SQL	Structured Query Language
STAPS	Sciences et techniques des activités physiques et sportives
UAR	Unité d'appui et de recherche
UMR	Unité mixte de recherche

PRÉSENTATION DES PLATEFORMES DATA

DemoMed MMSH

La plateforme technologique DemoMed est le seul membre du réseau Data à ne pas être membre de l'IR* PROGEDO et donc à ne pas être une PUD. Cette plateforme est une émanation d'un projet de recherche, lancé en 2009 pensé par des démographes d'Aix-Marseille Université accompagnés par des géographes et démographes de divers pays du pourtour méditerranéen dont l'objectif était de rendre interopérables des données démographiques issues de différents pays méditerranéens et produites à des niveaux administratifs et époques divers. Elle propose une plateforme numérique technologique innovante d'accès aux données démographiques. Son site internet permet une visualisation graphique, cartographique et tabulaire interactive d'indicateurs démographiques sur les populations méditerranéennes.

PANELS MSH-LSE

La plateforme PANELS, labellisée PUD par l'IR* PROGEDO, propose un accompagnement aux chercheurs des 54 laboratoires SHS des sites universitaires lyonnais et stéphanois sur l'ensemble du cycle de vie de la donnée, avec une conception large des données. Composante du pôle « Humanités numériques, systèmes d'information », elle intègre des compétences en enquêtes, statistiques, traitement et analyse de données, ingénierie documentaire et science ouverte. Elle accompagne les chercheurs à toutes les étapes d'une recherche requérant des collectes de données, des traitements statistiques, des modélisations et interprétations. Cet appui technique et méthodologique permet aux chercheurs en SHS d'acquérir et de développer de nouvelles compétences en enquêtes statistiques, observation de traces numériques, recueil et moissonnage de données publiques, accès aux données de la statistique publique, bases de données ainsi que traitements statistiques et fouille de données. Enfin, elle propose une mutualisation des compétences et connaissances en données quantitatives par des voies de communication diverses.

PROGEDO-Loire MSH Ange-Guépin

PROGEDO-Loire a été créée en 2015 et couvre les besoins en accompagnement en données quantitatives de 27 laboratoires SHS répartis sur l'ensemble de la région ligérienne (universités de Nantes, Angers et Le Mans). En tant que relais local de l'IR* PROGEDO, elle sensibilise à la réutilisation de données présentes dans l'entrepôt Quetelet-Progedo-Diffusion, ainsi qu'au traitement quantitatif des données. Elle propose des cycles de formations réservés aux doctorants ligériens ainsi qu'un séminaire annuel de formation sur le logiciel R, conçu comme une première étape avant des accompagnements individuels plus poussés. Ceux-ci

peuvent porter sur la création et la préparation des données, leur traitement, leur analyse et divers aspects liés aux données et à la science ouverte tels que le RGPD, la FAIRisation des données, la rédaction de PGD ou le dépôt de données dans Quetelet-Progedo-Diffusion. Son ingénieure participe à des projets de recherche impliquant le traitement quantitatif de données. Enfin, PROGEDO-Loire organise des événements scientifiques autour des données et méthodes quantitatives en SHS.

Plateforme Universitaire de Données d'Aix-Marseille Université (PUD-AMU) MMSH

La PUD-AMU a été créée en 2018 pour couvrir les besoins en accompagnement aux données quantitatives de la communauté scientifique des 10 laboratoires SHS d'Aix-Marseille Université. En tant que relais local de l'IR* PROGEDO, elle sensibilise à la réutilisation de données présentes dans le catalogue Quetelet-Progedo-Diffusion, au traitement quantitatif des données et aux bonnes pratiques de dépôt de données en entrepôts. Elle propose notamment des cycles de formations logicielles aux masterants, doctorants et chercheurs de son périmètre. Son ingénieur participe activement au pilotage du réseau Mate-SHS CoDaPro (Collectif Data Provence), ancrant la plateforme comme un acteur central du paysage quantitatif local.

Plateforme Universitaire de Données de Bretagne (PUD-B) MSHB

La PUD-B a été créée en 2019 et couvre les besoins en accompagnement de la communauté scientifique SHS de l'ensemble de la région Bretagne (38 laboratoires). En tant que relais local de l'IR* PROGEDO, elle sensibilise à la réutilisation de données présentes dans le catalogue Quetelet-Progedo-Diffusion, ainsi qu'au traitement quantitatif des données. Elle propose des formations collectives sur des logiciels de traitement et analyse de données quantitatives, sur, entre autres, la gestion de jeux de données et sur l'accès aux données du CASD pour lesquels elle dispose d'une SD-Box. La plateforme fournit également un accompagnement personnalisé aux chercheurs, doctorants et masterants autour de la recherche de jeux de données, de leur bonne utilisation, interprétation et analyse. Enfin, depuis 2023, elle participe, avec le laboratoire Arènes à une école d'été annuelle sur les méthodes mixtes en sciences sociales.

Plateforme Universitaire de Données de Bordeaux (PUD-Bx) MSH Bx

La PUD-Bx a été créée en mars 2022 et propose un accompagnement aux chercheurs en SHS issus des 14 laboratoires SHS affiliés à la MSH-Bx dans la collecte, le traitement, l'analyse et la diffusion de leurs données. En tant que relais de l'IR* PROGEDO, elle sensibilise à

la réutilisation de données présentes dans le catalogue Quetelet-Progedo-Diffusion. Par sa participation à l'*EquipEx+ : LifeObs*, elle contribue de façon active à la production et la diffusion de données françaises sur les parcours de vie et les tendances démographiques, tout en constituant, en partenariat avec d'autres plateformes Data, des kits pédagogiques visant à devenir des outils d'auto-formation à l'analyse des données quantitatives. La PUD-Bx participe aux travaux de l'URFIST autour du développement de formations à la science ouverte. Elle organise, chaque année, la Semaine Data-SHS à Bordeaux. Enfin, la PUD-Bx propose un accès aux données du CASD.

Plateforme Universitaire de Données de Caen (PUDC) MRSH

La PUDC a été créée en 2011 et est une composante du pôle « Plateformes – recherche », directement rattaché à l'autorité de la direction de la MRSH. Elle bénéficie d'une stabilité durable et d'une bonne implantation locale permettant de répondre aux nombreuses demandes d'accompagnement individuel quant au traitement des données ainsi qu'aux chercheurs en quête de jeux de données préexistants. L'essentiel des formations proposées par la plateforme traite du maniement de Microsoft Excel et de R/Rstudio, bien que d'autres demandes puissent ponctuellement apparaître. Face au déficit de formation quantitative sur le campus, et aux compétences en traitement qualitatif disponibles dans d'autres services, la PUDC centre l'ensemble de ses activités sur la production, le traitement et l'analyse des données quantitatives à destination des membres des 18 équipes du domaine des SHS fédérées par la MRSH

Plateforme Universitaire de Données Clermont-Auvergne (PUD-CA) MSH Clermont

La PUD-CA a été créée en 2021 avec comme objectif de faciliter l'accès et l'utilisation de données quantitatives en SHS. Insérée au sein d'un paysage scientifique local faiblement doté en SHS quantitatives, elle apporte une expertise reconnue dans ses accompagnements. Elle propose également des formations logicielles de niveau débutant afin de faciliter l'entrée vers le traitement quantitatif de chercheurs peu habitués à ces types de méthodes. Ses services sont accessibles à l'ensemble des communautés de recherche des 15 laboratoires associés à la MSH Clermont.

Plateforme Universitaire de Données de Dijon (PUDD) MSH Dijon

La PUDD a été créée en 2015 à la MSH de Dijon au sein de laquelle elle est une composante de la Plateforme Humanités Numériques. Elle est labellisée par le Comité d'Orientation Stratégique (COS) des Plateformes de Bourgogne-Franche-Comté (BFC) au niveau régional.

Cette labellisation lui ouvre l'accès aux AAP de la région BFC, facilitant ainsi sa participation aux financements et aux ressources disponibles dans la région. La PUDD propose des formations d'initiation à des logiciels de traitement statistique, mixte et qualitatif. Elle met à disposition des licences propriétaires telles que SAS et Nvivo pour compléter l'offre du Centre de calcul de l'Université de Bourgogne (CcuB), ainsi qu'un accès aux données du CASD par le biais d'une SD-Box. Une partie de ses frais de fonctionnement est assurée par le biais de fonds extérieurs en réponse à des AAP de partenaires institutionnels et pour lesquels la PUDD met en œuvre des enquêtes statistiques.

Plateforme Universitaire de Données Grenoble-Alpes (PUD-GA) MSH-Alpes

La PUD-GA a été créée en 2019 dans un environnement universitaire marqué par l'innovation, les liens avec la société civile et des compétences solides en traitement et analyse des données quantitatives. Elle bénéficie d'un environnement fortement structuré autour des données, toutes disciplines confondues (Vendredi Quanti, CDGA, projet GATES). Cette structuration s'accompagne d'investissements importants sur les thématiques des données et de l'innovation, permettant ainsi de coordonner et fluidifier les liens entre acteurs de la donnée sur le campus grenoblois. La PUD-GA accompagne la communauté scientifique des 26 laboratoires SHS de l'UGA avec une coloration « intelligence des données ». Elle propose un accompagnement aux projets de recherche sur l'ensemble du cycle de vie de la donnée. Son offre de formation, plus réduite, permet aux étudiants et chercheurs d'acquérir les bases du traitement statistique des données ainsi qu'une acculturation aux données par la diffusion d'informations sur les grandes enquêtes nationales et internationales.

Plateforme Universitaire de Données de Lille (PUDL) MESHS

La PUDL est la plus ancienne des plateformes Data, avec une création en 2002-2003. Dès le départ, son objectif était de « favoriser le développement des recherches de sciences humaines et sociales s'appuyant sur l'exploitation et le traitement de grands fichiers statistiques, en accompagnant les travaux des chercheurs expérimentés, en favorisant la formation des jeunes chercheurs et des doctorants, et en mettant à disposition de l'enseignement et de la recherche des outils statistiques et des informations adaptés »⁶⁵. Le modèle mis en place à Lille a donc ensuite été répliqué dans l'ensemble des PUD ouvertes en France. Désormais, la PUDL accompagne les 46 laboratoires affiliés à la MESH, dont une vingtaine directement concernée par les données quantitatives, en soutenant et en favorisant les recherches basées sur des données quantitatives en sciences sociales, tout en sensibilisant à l'usage de ce type de données au niveau régional, national et international. À ces fins, la PUDL dispose de trois SD-Box permettant d'accéder aux données du CASD,

⁶⁵ « Note sur les plates-formes universitaires pour les données en sciences humaines et sociales », adressée par P. Casella à la Direction de la Recherche du Ministère de la Recherche, 21.02.2002.

d'un service de prêt d'ouvrages de référence sur le traitement de données quantitatives et organise un séminaire méthodologique sur les méthodes mixtes en SHS.

Plateforme Universitaire de Données de Nanterre (PUDN) MSH Mondes

La PUDN a été créée en 2019 à la MSH Mondes. Son implantation centrale sur le campus de l'Université Paris Nanterre ainsi que la stabilité de son ingénieur, présent dès les débuts de la plateforme, ont permis de rapidement ancrer la plateforme dans le paysage scientifique nanterrois. Elle bénéficie ponctuellement du soutien d'ingénieurs en CDD affectés à des projets de recherche précis. Déployée au sein du service *Humanités numériques* de la MSH Mondes, elle accompagne les étudiants, doctorants et enseignants-chercheurs des universités Paris Nanterre et Paris 1 Panthéon-Sorbonne dans l'utilisation de données quantitatives et de méthodes d'analyse quantitatives, afin de contextualiser, d'enrichir et de conduire leurs analyses de recherche. En plus des accompagnements individuels, elle propose des formations R, inscrites sur les catalogues des écoles doctorales. Enfin, la plateforme organise régulièrement des séminaires thématiques visant à diffuser les grandes enquêtes et bases de données administratives.

Plateforme Universitaire de Données de Montpellier MSH-SUD

Créée en janvier 2024, la PUD-Montpellier n'était pas encore en activité lors de la rédaction du rapport.

Plateforme Universitaire de Données de Paris-Saclay (PUD-PS) MSH Paris-Saclay

La PUD-PS a été inaugurée en décembre 2023 et est gérée par la MSH-PS et par l'ENS Paris-Saclay. Elle propose une expertise axée sur l'ingénierie de recherche orientée vers les données ouvertes tout en valorisant la production scientifique participative. N'ayant pas d'ingénieur actuellement en poste, la PUD-PS n'a pas encore pu s'implanter dans le paysage scientifique du plateau de Saclay.

Plateforme Universitaire de Données de Strasbourg (PUD-S) MISHA

La PUD-S a été créée en 2016 et rapidement labellisée Data par le RnMSH, avant d'être rattachée à la MISHA en 2018. Elle est désormais intégrée au Pôle « Données et Savoirs »

avec la Plateforme Humanités Numériques. En tant qu'émanation locale de l'IR* PROGEDO, elle vise à soutenir et promouvoir les analyses quantitatives des données issues des grandes enquêtes et des données administratives. Elle propose donc de faciliter l'accès à ces données ainsi qu'à leur traitement et analyse par le biais de formations collectives et accompagnements individuels. Historiquement marquée par la présence de démographes, elle bénéficie du soutien de l'*EquipEx+ : LifeObs* qui finance un second poste d'ingénieur. La plateforme dispose également d'une maîtresse de conférences contractuelle en partie affectée à ses formations. Elle propose aux chercheurs alsaciens un accompagnement à l'accès aux données du CASD par la mise à disposition de deux SD-Box.

Plateforme Universitaire de Données de Toulouse (PUD-T) MSHS-T

La PUD-T est née en 2019 dans un environnement universitaire et scientifique particulièrement propice aux enjeux liés aux données, notamment du fait de la forte implication du SCD de l'université de Toulouse sur ces questions. Elle est désormais une composante du pôle « Données, science ouverte et informatique » aux côtés de la plateforme Scripto ARCHITOU, d'un développeur d'applications et de la correspondante de l'IR* Huma-Num. Elle propose des services de soutien à la recherche et à la formation à la communauté scientifique des 38 unités fédérées par la MSHS-T. Elle offre un accompagnement à la formation en données quantitatives dès la licence. Disposant d'un haut niveau d'expertise scientifique, elle propose également des partenariats avec des projets de recherche tels que l'ANR ObARDI (2021-2025) et le PANELVICO (2023-2027) qui a pris la suite de l'ANR VICO dans lequel plusieurs plateformes Data étaient impliquées. Enfin, elle est porteuse du projet pédagogique « Démocratiser la réutilisation d'enquêtes de qualité par la formation » financé par le GIS Réseau Urfist.

Plateforme Universitaire de Données de l'Université de Poitiers (PUD-UP) MSHS Poitiers

Créée en 2020, la PUD-UP est une composante du pôle « Soutien à la recherche et innovation » et de son service « Plateformes technologiques ». Elle est pilotée par un comité composé de la gestionnaire et de la référente scientifique de la plateforme, du responsable de la plateforme technologique de la MSHS et de la Responsable du pôle Recherche et innovation. Elle a comme objectif d'accompagner les projets de recherche et les chercheurs et doctorants dans l'appropriation des données quantitatives en sciences sociales, en facilitant l'accès et l'utilisation de bases de données spécialisées et en formant aux méthodes et outils de traitement de données par le biais de cycles de séminaires et formations de tous niveaux et dont certains sont inscrits dans les catalogues des écoles doctorales. Elle s'adresse aux membres de 15 laboratoires rattachés à la MSHS Poitiers. L'ingénieure en poste a également participé au projet de recherche « Le coteau ligérien vu par les naturalistes des Pays de la Loire ». La plateforme est également membre de différents réseaux régionaux liés aux données.

