
PGD du projet PRESPOL

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - Modèle de PGD (français)" fourni par Agence nationale de la recherche (ANR).

Renseignements sur le plan

Titre du plan PGD du projet PRESPOL

Livrable

Version Version initiale

Objet/périmètre du plan Le Plan de gestion des données (PGD) a été conçu collectivement au sein de PRESPOL de manière à soutenir non seulement la recherche actuelle mais aussi de futures études sur l'autonomie économique des personnes handicapées. Cette version initiale du PGD correspond ainsi à l'aboutissement d'une démarche collaborative au sein du projet :

- Avril 2024 : Réunion de lancement du projet, en présence du financeur, au cours de laquelle la question du PGD est abordée avec les correspondants des établissements scientifiques partenaires du projet.
- Juillet 2024 : Réunion du comité de pilotage du projet, au cours de laquelle la démarche du PGD est présentée en détail. La question des modalités les plus pertinentes de partage des données au moment du traitement sont notamment évoquées par les responsables de *work packages* et des *tasks*.
- Juillet 2024 : Formation Juliette Pinon (project manager) au PGD dans le cadre du PPR Autonomie.
- Août 2024 : L'ensemble des responsables de tâches, réparties dans les quatre *work packages* du projet ont répondu à un questionnaire élaboré à partir du modèle de PGD de l'Agence nationale de recherche. Ce formulaire avait pour objectif de mettre en place une réflexion autour des données du projet afin de les rendre FAIR (Faciles à trouver, Accessibles, Interopérables, Réutilisables), en prenant en compte les habitudes et les possibilités de chacune et de chacun. Le formulaire du questionnaire a été relu et testé avant partage par Anne Revillard (responsable scientifique et technique de PRESPOL), Célia Bouchet (responsable de la Tâche 4.1 Science ouverte au sein de PRESPOL) et Emmanuel Collier (coordinateur de l'activité de conseil à la rédaction des PGD de l'Humathèque).
- Septembre 2024 : Rédaction d'une première ébauche du PGD par Juliette Pinon soumise à relecture auprès de Emmanuel Collier (coordinateur de l'activité de conseil à la rédaction des PGD de l'Humathèque) et Cyril Heude (responsable-adjoint de bibliothèque et service Science Ouverte et Appui à la recherche).

- Septembre 2024 : Le projet de PGD est présenté en COPIL et validé collectivement avant soumission à l'ANR.

Cette démarche collective visait à offrir une première vue d'ensemble des données pour l'ensemble du projet Prespol tout en sensibilisant les chercheurs et chercheuses au cycle des données avant que les terrains ne débutent. A ce stade, nous pensons produire une actualisation du PGD à mi-chemin du projet, en Octobre 2026 et à la fin du projet en Mars 2029.

Le projet Prespol est structuré en WP. Chaque WP est divisé en tâches pour lesquelles des livrables ont été identifiés. Les WP sont organisés de manière à ce que les deux responsables scientifiques soient d'établissements, de disciplines et de méthodes complémentaires. Pour la présentation des jeux de données, certaines tâches ont été divisées en deux produits de recherche distincts (tâche 1.2 et tâche 3.1) afin d'assurer la suivi et le contact par la personne en charge de cette opération de recherche spécifiquement.

WP1 : Politiques d'autonomisation économique pour les personnes handicapées, entre emploi et politiques sociales (coord. E. Ferragina, LIEPP, et A. Lejeune, CERAPS)

- **Tâche 1.1 : Théorisation des politiques d'autonomisation économique (coord. E. Ferragina et A. Revillard, LIEPP) - Dates prévisionnelles : Octobre 2024-Décembre 2025**

Livrables : Article sur les régimes de protection sociale et le handicap (livrable 1), article sur le handicap et l'investissement social (livrable 2), cartographie des politiques (livrable 3), article théorique sur les politiques d'autonomisation économique (livrable 4)

- **Tâche 1.2 : Autonomisation entre emploi et politiques sociales : une entrée par les trajectoires (coord. A. Lejeune, CERAPS, et A. Boyer, LIEPP) - Dates prévisionnelles : Janvier 2025-Décembre 2025**

Livrables : Étude de trajectoire : autonomisation/travail/politique sociale (livrable 1), article quantitatif (livrable 2)

- **Tâche 1.3 : Administrer la frontière emploi/politique sociale (coord. A. Lejeune, CERAPS) - Dates prévisionnelles : Octobre 2025-Septembre 2028**

Livrables : Analyse documentaire (dossiers administratifs, dossiers judiciaires) sur les usages de la notion d'employabilité (livrable 1), thèse de doctorat et publications (livrable 2)

WP2 : Handicap et inégalités professionnelles : quelle autonomisation économique ? (coord. C. Bouchet, LIEPP, et J-F Trani, LISE)

- **Tâche 2.1 : La marginalisation des personnes handicapées de l'emploi et au sein de l'emploi : un aperçu général (coord. C. Bouchet, LIEPP) - Dates prévisionnelles : Avril 2024-Décembre**

2027

Livrables : Article sur les effets du travail à temps partiel basé sur EU-SILC (livrable 1), article sur les trajectoires professionnelles en France (livrable 2), article sur l'articulation du temps social en Europe (livrable 3)

- **Tâche 2.2 : Handicap et emploi : quelles discriminations ? (coord. C. Bouchet, LIEPP, et M. Mercat-Bruns, LISE) - Dates prévisionnelles : Avril 2024-Septembre 2028**

Livrables : Présentation à la clinique juridique de Sciences Po (livrable 1), publication dans l'International Journal of Comparative Law (livrable 2), article comparatif sur les discriminations dans l'UE basé sur EU-LFS (livrable 3), article à méthodes mixtes sur les conséquences des discriminations en France (livrable 4)

- **Tâche 2.3 : Accès à l'emploi et autonomisation économique des jeunes handicapés (coord. J-F Trani, LISE) - Dates prévisionnelles : Octobre 2024-Septembre 2028**

Livrables : Présentation à la Human Development and Capabilities Association, à la Society for Social Work and Research, et aux conférences de l'American Public Health Association (livrable 1) ; Article sur les obstacles à l'emploi dans les zones urbaines en Afrique du Sud dans une revue de développement (World Development) (livrable 2) ; Article comparatif sur le handicap et les conditions d'emploi en France, au Royaume-Uni et en Afrique du Sud pour les personnes handicapées (revue sur le handicap, livrable 3).

WP3 : Évaluation de deux dispositifs clés des politiques d'emploi (coord. Y. L'horty, ERUDITE, et C. Spieser, LISE)

- **Tâche 3.1 : Mise en œuvre, appropriation et effets des politiques de quotas (coord. Y. L'horty, ERUDITE, et C. Spieser, LISE) Dates prévisionnelles : Octobre 2024-Septembre 2027**

Livrables : Revue systématique de la littérature sur les politiques de quotas (livrable 1), article basé sur l'enquête testing (livrable 2), article basé sur l'Enquête Autonomie (livrable 3), article sur la mise en œuvre du quota (livrable 4), article sur la réception du quota (livrable 5)

- **Tâche 3.2 : Emploi accompagné (coord. F. Jaffrès, EHESP). Dates prévisionnelles : Janvier 2025-Décembre 2027**

Livrables : Cartographie des dispositifs d'emploi accompagné (livrable 1), articles sur la mise en œuvre (livrable 2) et la réception (livrable 3) des dispositifs d'emploi accompagné

WP4 : Science ouverte, médiation scientifique et formation (coord. E. Fillion, EHESP, et A. Revillard, LIEPP)

- **Tâche 4.1 : Science ouverte (coord. C. Bouchet et A. Revillard,**

LIEPP) - Dates prévisionnelles: Avril 2024-Mars 2029

Livrables : Plan de gestion des données (livrable 1), téléversement des ensembles de données quantitatives et des analyses statistiques (livrable 2), transcriptions d'entretiens (livrable 3) et autres matériaux qualitatifs (livrable 4) sur nakala ; téléversement des résumés d'articles sur la base de données "Handicap et Travail" (livrable 5) ; publication de deux livres en accès libre

- **Tâche 4.2 : Médiation scientifique (coord. E. Payen, EHESP, et A. Revillard, LIEPP) - Dates prévisionnelles : Avril 2024-Mars 2029**

Livrables : Blog de recherche (livrable 1), affiches et clips vidéo (livrable 2), 50 portraits de travailleurs handicapés (livrable 3), 8 notes de politique (livrable 4), exposition (livrable 5), documentaire (livrable 6).

- **Tâche 4.3 : Formation (coord. A. Revillard, LIEPP) - Dates prévisionnelles : Avril 2024-Mars 2029**

Livrables : Travail de terrain collectif de niveau master (livrable 1), École thématique d'hiver (livrable 2), actions de formation initiale (livrable 3), actions de formation continue (livrable 4), 3 notes méthodologiques (livrable 5)

Il est à noter que différentes tâches ont des liens entre elles, notamment en termes de collecte et d'analyse des données. Ces éléments sont anticipés en termes de cohérence et de transparence lors de la collecte, de sécurisation du partage des données et de droits de propriété. A titre informatif, les liens que nous établissons au début du projet, et donc sujets à évolution, sont les suivants :

- Tâche 1.1 - Tâche 1.2 : Partage d'informations sur les politiques d'emploi et les politiques sociales au Royaume-Uni
- Tâche 1.1 - Tâche 1.3 : Partage d'informations sur les politiques d'emploi et les politiques sociales au Royaume-Uni
- Tâche 1.1 - Tâche 3.1 : Revue de littérature partagée sur le mapping des dispositifs de politiques publiques (quotas)
- Tâche 1.1 - Tâche 3.2 : Partage d'informations sur les politiques d'emploi et les politiques sociales en Suède
- Tâche 1.2 - Tâche 2.1 : Mutualisation d'entretiens
- Tâche 2.1 - Tâche 3.1 : Revue de littérature partagée sur le mapping des dispositifs de politiques publiques (discriminations)
- Tâche 2.1 - Tâche 3.2 : Revue de littérature partagée sur le mapping des dispositifs de politiques publiques (emploi accompagné)
- Tâche 2.2 - Tâche 3.1 : Mutualisation d'entretiens
- Tâche 2.2 - Tâche 3.2 : Mutualisation d'entretiens
- Tâche 2.3 - Tâche 3.2 : Mutualisation d'entretiens

Domaines de recherche (selon classification de

5.4 Sociology, 5.5 Law, 5.6 Political science, 5.9 Other social sciences, 5.2 Economics and business, 5.7 Social and economic geography, 5.3 Education, 6.5 Other humanities

l'OCDE)

Langue	fra
Date de création	2024-08-30
Date de dernière modification	2024-09-27
Identifiant	
Type d'identifiant	DOI

Documents (publications, rapports, brevets, plan expérimental...), sites web associés

- Site de présentation du projet Prespol - PPR Autonomie : <https://ppr-autonomie.com/les-projets-finances/projet-prespol/>

Renseignements sur le projet

Titre du projet Promouvoir l'autonomie économique des personnes handicapées par l'emploi et les politiques sociales

Acronyme PRESPOL

Résumé Le projet PRESPOL (Promouvoir l'autonomie économique des personnes handicapées par l'emploi et les politiques sociales) vise à améliorer l'autonomie économique des personnes handicapées en France et dans d'autres pays, en se concentrant sur l'emploi et les politiques sociales comme leviers principaux.

Objectifs principaux

Le projet PRESPOL vise à promouvoir l'autonomie économique des personnes handicapées en améliorant les interventions publiques dans les domaines de l'emploi et les politiques sociales. Il s'articule autour de trois principaux objectifs de recherche :

- (1) développer un cadre théorique original pour analyser les politiques d'autonomisation économique des personnes handicapées (WP1)
- (2) approfondir la compréhension des obstacles et des expériences d'autonomisation des personnes handicapées dans l'emploi (WP2)
- (3) évaluer systématiquement deux dispositifs centraux de politiques d'emploi : l'emploi accompagné et les politiques de quotas (WP3)

Cadre théorique

PRESPOL s'appuie sur une conception interactive et intersectionnelle du handicap, pensée à partir de caractéristiques personnelles et de facteurs environnementaux. Le projet propose une théorisation originale des politiques d'autonomisation économique en tenant compte des régimes de l'État-providence et des investissements sociaux, tout en développant une critique de ces cadres existants à travers le prisme des *Disability studies*. La recherche se positionne ainsi à l'interface des politiques de l'emploi et de politiques sociales indépendantes de l'emploi, incluant des éléments tels que l'accessibilité, le logement, la

formation ou les transports.

Méthodologie

Le projet adopte une approche interdisciplinaire, combinant des méthodes qualitatives et quantitatives, et utilise une comparaison internationale pour évaluer l'efficacité des politiques d'autonomisation. Plus précisément, des études de cas, des comparaisons binationales (France, Royaume-Uni, Suède, Afrique du Sud) et des enquêtes quantitatives sur l'emploi et les conditions de travail des personnes handicapées sont prévues. Une forte dimension participative est intégrée, avec l'implication de plus de 20 organisations associatives et institutionnelles au cours de la recherche, afin de maximiser la pertinence et l'impact des recherches sur les politiques publiques.

Résultats attendus

Les résultats escomptés comprennent la formulation d'un cadre théorique comparatif sur les politiques d'autonomisation économique, une évaluation précise de l'efficacité des politiques existantes et un diagnostic des inégalités liées au handicap sur le marché du travail.

A ce titre, le projet vise à influencer les décideurs politiques et à sensibiliser le public aux discriminations liées au handicap.

Le développement de modalités de participation adaptées aux personnes en situation de handicap et aux personnes qui leur sont proches au cours de la recherche auront également comme impact de faciliter la prise en compte de leurs savoirs et de leurs expériences dans des recherches futures.

Gouvernance

Dans la continuité de ses ambitions scientifiques, le projet s'appuie et développe la collaboration de plusieurs entités de recherche françaises :

- Paris : SciencesPo/LIEPP
- Paris : CNRS/CNAM/Lise (UMR 3320)
- Rennes : EHESP /Arènes (UMR 6041)
- Lille : Université de Lille/Ceraps (UMR 8026)
- Marne-la-Vallée : TEPP/Université de Gustave Eiffel /ERUDITE (EA 437)

Des partenaires scientifiques internationaux pourront également être mobilisés dans certaines tâches, le cas échéant.

La gouvernance repose sur :

- Un comité de pilotage opérationnel représentatif des entités du consortium, à but décisionnel sur des sujets transversaux, comme la gestion des données (minimum 3 COPIL/an). Il réunit tous les responsables de work packages, de tasks et les correspondants d'établissements.
- Un comité de parties prenantes réunissant les partenaires associatifs et institutionnels, avec un suivi formel, à but consultatif (1/an, en amont du rendu du rapport scientifique et financier au financeur).
- Un comité scientifique international chargé d'évaluer les orientations et les résultats du projet, avec un suivi formel, à

but consultatif (3/projet)

Sources de financement

- Agence Nationale de la Recherche : ANR-23-PAVH-0001

Date de début 2024-04-01

Date de fin 2029-03-31

Partenaires

- Ecole des Hautes Etudes en Santé Publique
- Université Gustave Eiffel
- Université de Lille
- CNRS Ile-de-France Meudon
- Conservatoire National des Arts et Métiers Paris

Produits de recherche :

1. Tâche 1.1 : Théorisation des politiques d'autonomisation économique (coord. E. Ferragina et A. Revillard, LIEPP) (Jeu de données)
2. Tâche 1.2 - volet quantitatif : Autonomisation entre emploi et politiques sociales : une entrée par les trajectoires (A. Boyer, LIEPP) (Jeu de données)
3. Tâche 1.2 - volet qualitatif : Autonomisation entre emploi et politiques sociales : une entrée par les trajectoires (A. Lejeune, CERAPS) (Jeu de données)
4. Tâche 1.3 : Administrer la frontière emploi/politique sociale (coord. A. Lejeune, CERAPS) (Jeu de données)
5. Tâche 2.1 : La marginalisation des personnes handicapées de l'emploi et au sein de l'emploi : un aperçu général (coord. C. Bouchet, LIEPP) (Jeu de données)
6. Tâche 2.2 : Handicap et emploi : quelles discriminations ? (coord. C. Bouchet, LIEPP, et M. Mercat-Bruno, LISE) (Jeu de données)
7. Tâche 2.3 : Accès à l'emploi et autonomisation économique des jeunes handicapés (coord. J-F Trani, LISE) (Jeu de données)
8. Tâche 3.1 - volet quantitatif : Mise en œuvre, appropriation et effets des politiques de quotas (Y. L'horty, ERUDITE) (Jeu de données)
9. Tâche 3.1 - volet qualitatif : Mise en œuvre, appropriation et effets des politiques de quotas (C. Spieser, LISE) (Jeu de données)
10. Tâche 3.2 : Emploi accompagné (coord. F. Jaffrès, EHESP) (Jeu de données)
11. Tâche 4.2 : Médiation scientifique (coord. E. Payen, EHESP, et A. Revillard, LIEPP) (Jeu de données)
12. Tâche 4.3 : Formation (coord. A. Revillard, LIEPP) (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
Arrigoni Alessandro - https://orcid.org/0000-0001-7387-003X	Fondation Nationale des Sciences Politiques	
Bouchet Célia - https://orcid.org/0000-0002-1992-6554	Fondation Nationale des Sciences Politiques	<ul style="list-style-type: none">• Personne contact pour les données (Marginal Emploi)

Boyer Aurélien	Fondation Nationale des Sciences Politiques	<ul style="list-style-type: none"> • Personne contact pour les données (Trajectoires quanti)
Dessein Sophie - https://orcid.org/0000-0002-9476-0785	Conservatoire national des arts et métiers	
Ferragina Emanuele - https://orcid.org/0000-0002-6308-1413	Fondation Nationale des Sciences Politiques	<ul style="list-style-type: none"> • Personne contact pour les données (Théorie Autonomie)
Fillion Emmanuelle - https://orcid.org/0000-0003-0286-4315	Ecole des Hautes Etudes en Santé Publique	
Jaffrès Fanny - https://www.researchgate.net/profile/Fanny-Jaffres	Ecole des Hautes Etudes en Santé Publique	<ul style="list-style-type: none"> • Personne contact pour les données (Éval Emploi Acc)
Lejeune Aude - https://orcid.org/0000-0003-3675-9812	Université de Lille	<ul style="list-style-type: none"> • Personne contact pour les données (Frontière Emploi, Trajectoires quali)
L'Horty Yannick - https://orcid.org/0000-0001-6258-8565	Université Gustave Eiffel	<ul style="list-style-type: none"> • Personne contact pour les données (Éval quanti quotas)
Mercat-Bruns Marie - https://www.researchgate.net/profile/Marie-Mercat-Bruns	Conservatoire national des arts et métiers	<ul style="list-style-type: none"> • Personne contact pour les données (Discrim Emploi)
Noet Ngoc-Thao - https://orcid.org/0000-0002-0786-1992	Université Gustave Eiffel	
Ortiz Daniela	Conservatoire national des arts et métiers	
Payen Emeline	Ecole des Hautes Etudes en Santé Publique	<ul style="list-style-type: none"> • Personne contact pour les données (Médiation)
Pinon Juliette	Fondation Nationale des Sciences Politiques	<ul style="list-style-type: none"> • Responsable du plan

Revillard Anne - https://annerevillard.wordpress.com/	Fondation Nationale Sciences Politiques	<ul style="list-style-type: none"> • Personne contact pour les données (Formation) • Responsable
Spieser Catherine - https://orcid.org/0000-0003-0667-7162	Conservatoire national des arts et métiers	<ul style="list-style-type: none"> • Personne contact pour les données (Eval quali quotas)
Trani Jean-François - https://orcid.org/0000-0002-9187-0946	Conservatoire national des arts et métiers	<ul style="list-style-type: none"> • Personne contact pour les données (Jeunes & Emploi)

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

PGD du projet PRESPOL

1. Description des données et collecte ou réutilisation de données existantes

Tâche 1.1 : Théorisation des politiques d'autonomisation économique (coord. E. Ferragina et A. Revillard, LIEPP)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 1.1 : Théorisation des politiques d'autonomisation économique (coord. E. Ferragina et A. Revillard, LIEPP), nous réutiliserons principalement des données qualitatives et quantitatives déjà existantes, initialement collectées par des tiers. Ces données comprendront principalement des rapports et des données législatives, qui serviront à construire des comparaisons internationales. A ce stade, les sources exactes doivent être précisées mais on peut déjà citer le recours aux recommandations et documents d'organisations internationales et des institutions comme l'Organisation internationale du travail (OIT), les Nations Unies (ONU), l'Organisation de coopération et de développement économiques (OCDE) et l'Union européenne (UE).

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives réutilisées

- **Types de données** : Rapports, guidelines, documents administratifs, directives, articles académiques, livres, rapports de recherche, données législatives.
- **Format** : Documents sous formats texte (.docx, .pdf). Durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (TXT (Text)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre de documents analysés. On peut s'attendre à la gestion de plusieurs centaines de pages de documents textuels (entre 200 à 500 pages).

2. Les données quantitatives réutilisées

- **Type de données** : Données quantitatives existantes issues de bases de données à définir.
- **Format** : Données structurées dans des formats numériques standard pour les bases de données statistiques, (.CSV (Comma-Separated Values)), (.XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). Durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra de la part donnée au traitement quantitative dans cette tâche

Tâche 1.2 - volet quantitatif : Autonomisation entre emploi et politiques sociales : une entrée par les trajectoires (A. Boyer, LIEPP)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 1.2 : Autonomisation entre emploi et politiques sociales : une entrée par les trajectoires, les travaux d'Aurélien Boyer s'appuieront sur des données préexistantes pour une analyse quantitative.

Ces données incluent la base de données ENIACRAMS (Enquête Nationale Interministérielle sur les Allocataires de certains Régimes et Allocations de la Sécurité Sociale) et le panel "Tous actifs" de l'INSEE (Institut National de la Statistique et des Études Économiques). Ces bases de données seront exploitées pour mener des analyses de séquence et des modèles de durée, permettant de caractériser les trajectoires d'emploi et de recours aux prestations, de les regrouper en fonction de leur degré de similarité, et d'identifier les facteurs déterminants de ces trajectoires.

Cette réutilisation de données quantitatives existantes facilite une analyse statistique robuste et complémentaire aux données qualitatives recueillies par les entretiens biographiques par Aude Lejeune.

Cette tâche est complémentaire de l'approche du WP3 du projet KAPPA (Conditions d'accès aux aides et politiques publiques de l'autonomie. Origines, implications et perspectives d'évolution de la segmentation par âge), également financé par le PPR Autonomie.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données quantitatives réutilisées

- **Type de données :** Données quantitatives existantes issues de deux bases de données principales :

-L' ENIACRAMS (Enquête Nationale Interministérielle sur les Allocataires de certains Régimes et Allocations de la Sécurité Sociale) est une base de données française qui regroupe des informations détaillées sur les allocataires de certaines prestations sociales et de sécurité sociale. Elle a été constituée dans le but de fournir une meilleure compréhension des caractéristiques socio-économiques, des parcours de vie, et des situations d'emploi des bénéficiaires de différentes allocations sociales. La base de données comporte un ensemble de données démographiques (âge, sexe, nationalité, situation familiale, etc.) ; de données socio-économiques (niveau d'éducation, situation professionnelle, type d'emploi, revenu, etc.), d'informations sur les parcours de vie (historique d'emploi, périodes de chômage ou d'inactivité, transitions entre différents états (emploi, chômage, inactivité) et de données relatives aux prestations sociales (types de prestations perçues, montants des prestations, durée de perception, motifs de non-recours ou de refus, etc.).

-Le panel "Tous actifs" de l'INSEE (Institut National de la Statistique et des Études Économiques) est une base de données longitudinale qui suit les parcours professionnels et les situations socio-économiques des individus en âge de travailler en France. Ce panel offre une vue détaillée sur l'ensemble de la population active, permettant d'analyser les dynamiques de l'emploi, les transitions sur le marché du travail, ainsi que l'impact des politiques publiques sur ces parcours. Ce panel contient des informations détaillées sur un large éventail de variables, notamment les caractéristiques démographiques (âge, sexe, situation familiale, niveau d'éducation, etc.), les situations d'activité (statut d'emploi, secteur d'activité, type de contrat (CDI, CDD, intérim), durée de l'emploi, etc.), les transitions sur le marché du travail (passages entre emploi et chômage, périodes de formation, reprises d'activité après une période d'inactivité, etc.), les conditions de travail et de rémunération (salaire, temps de travail, conditions de travail (télétravail, horaires), avantages sociaux, etc.) et la perception de prestations sociales (allocation chômage, RSA (Revenu de Solidarité Active), indemnités de maladie, retraite, etc.)

- **Format :** Ces données sont structurées dans des formats numériques standard pour les bases de données statistiques (.CSV (Comma-Separated Values) ; .XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). Durant leur exploitation, des formats fermés seront donc utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé :**

- L' ENIACRAMS (Enquête Nationale Interministérielle sur les Allocataires de certains Régimes et Allocations de la Sécurité Sociale) contient des milliers à des dizaines de milliers d'enregistrements relatifs aux bénéficiaires de prestations sociales en France, avec des variables couvrant des informations démographiques, socio-économiques, et des parcours de vie. Le volume de données dépendra de la période d'étude et de la couverture choisie, mais il pourrait représenter plusieurs gigaoctets (GO) de données.

-Le panel "Tous actifs" de l'INSEE (Institut National de la Statistique et des Études Économiques) est longitudinal et suit des milliers d'individus en âge de travailler sur une longue période. Le volume de données est potentiellement important, couvrant de nombreuses variables sur des périodes temporelles étendues, représentant potentiellement plusieurs gigaoctets (GO) de données également.

Tâche 1.2 - volet qualitatif : Autonomisation entre emploi et politiques sociales : une entrée par les trajectoires (A. Lejeune, CERAPS)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 1.2 : Autonomisation entre emploi et politiques sociales : une entrée par les trajectoires, Aude Lejeune produira de nouvelles données, qualitatives.

De nouvelles données seront recueillies à travers une approche d'entretiens biographiques. Ces entretiens permettront de recueillir des informations qualitatives détaillées sur les trajectoires de vie de bénéficiaires potentiels de prestations d'invalidité en France et au Royaume-Uni. Des études de cas se concentreront sur les personnes qui contestent le refus

de ces prestations devant les tribunaux, offrant ainsi un accès à des trajectoires marquées par des périodes d'emploi et des recours (ou absence de recours) à des prestations d'invalidité ou d'autres formes d'assistance sociale.

Ces données qualitatives sont complémentaires à l'analyse statistique d'Aurélien Boyer.

Cette tâche est complémentaire de l'approche du WP3 du projet KAPPA (Conditions d'accès aux aides et politiques publiques de l'autonomie. Origines, implications et perspectives d'évolution de la segmentation par âge), également financé par le PPR Autonomie.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données qualitatives produites

- **Type de données** : Données qualitatives recueillies par des entretiens biographiques et études de cas.
- **Format** : Transcriptions d'entretiens sous format texte (.docx, .pdf), notes de terrain sous format texte (.docx, .pdf), enregistrements audio et/ou vidéo (.mp3, .mp4). Durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (TXT (Text) ; MKVV ; MP3) pour le dépôt, le cas échéant..
- **Volume estimé** : Le volume dépendra du nombre d'entretiens réalisés et de leur durée. Par exemple, si 50 à 100 entretiens sont menés, avec une moyenne de 1 à 2 heures par entretien, les transcriptions pourraient représenter plusieurs centaines de pages (estimé entre 500 à 1 000 pages de texte). Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (GO), en fonction de la qualité d'enregistrement et du format de stockage utilisé.

Tâche 1.3 : Administrer la frontière emploi/politique sociale (coord. A. Lejeune, CERAPS)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 1.3 : Administrer la frontière emploi/politique sociale (coord. A. Lejeune, CERAPS), nous produirons de nouvelles données et réutiliserons des données existantes afin d'analyser la manière dont le principe d'employabilité est mis en œuvre en France et au Royaume-Uni.

D'une part, de nouvelles données qualitatives seront recueillies à travers des entretiens et des observations auprès des professionnels impliqués dans l'allocation des droits aux prestations sociales, notamment des médecins, qui évaluent l'état de santé et la capacité de travail des personnes handicapées, des travailleurs sociaux, qui accompagnent les demandeurs dans leurs démarches pour obtenir des prestations, et des juges des tribunaux sociaux, qui tranchent les litiges concernant l'attribution des prestations. Ces entretiens auront pour but de comprendre comment ces professionnels appliquent le principe d'"employabilité" dans leurs décisions d'attribution des prestations, telles que l'allocation aux adultes handicapés (AAH) en France, et de comparer ces pratiques avec celles du Royaume-Uni.

D'autre part, des données quantitatives et qualitatives existantes sur l'attribution des prestations sociales liées à l'incapacité de travail, en France et au Royaume-Uni, seront utilisées, notamment des documents officiels et rapports des administrations de la sécurité sociale ou des agences gouvernementales, décrivant les critères d'évaluation de l'employabilité et les statistiques sur les bénéficiaires des prestations d'invalidité ; des études de cas juridiques portant sur les décisions de justice concernant les contestations des refus de prestations sociales, qui offriront une base pour comprendre comment le principe d'employabilité est interprété par les juges et des publications académiques et rapports de recherche existants qui explorent déjà les critères d'attribution des prestations sociales et les politiques d'incitation à l'emploi pour les personnes en situation de handicap. Les sources exactes de ces données restent à déterminer.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives produites

- **Types de données** : Données qualitatives recueillies par des entretiens semi-directifs avec les professionnels impliqués dans l'attribution des droits sociaux (médecins, travailleurs sociaux, juges des tribunaux sociaux) et des observations directes des pratiques et des processus de décision concernant l'attribution des prestations sociales.
- **Format** : Transcriptions d'entretiens sous format texte (.docx, .pdf), notes de terrain et d'observations sous format texte (.docx, .pdf), enregistrements audio et/ou vidéo (.mp3, .mp4). Durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (TXT (Text) ; MKVV ; MP3) pour le dépôt, le cas échéant..

- **Volume estimé** : Le volume dépendra du nombre d'entretiens réalisés et de leur durée. Par exemple, si 50 à 100 entretiens sont menés, avec une moyenne de 1 à 2 heures par entretien, les transcriptions pourraient représenter plusieurs centaines de pages (estimé entre 500 à 1 000 pages de texte). Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (Go), en fonction de la qualité d'enregistrement et du format de stockage utilisé. Une compression pourra être à prévoir. Par ailleurs, la question du floutage et/ou du droit d'image sera abordée dans les prochains mois. Les notes de terrain pourraient représenter environ 50 à 100 pages de texte, selon la durée et la fréquence des observations.

2. Les données quantitatives réutilisées

- **Types de données** : Données quantitatives sur les bénéficiaires des prestations d'invalidité et critères d'évaluation de l'employabilité.
- **Format** : Ces données sont structurées dans des formats numériques standard pour les bases de données statistiques (.CSV (Comma-Separated Values) ; .XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). Durant leur exploitation, des formats fermés seront donc utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé** : Plusieurs fichiers de données statistiques, représentant potentiellement plusieurs mégaoctets (Mo) à quelques gigaoctets (Go), selon la taille des bases de données accessibles.

3. Les données qualitatives réutilisées

- **Types de données** : Rapports, documents administratifs, et directives des administrations de la sécurité sociale et des agences gouvernementales en France et au Royaume-Uni, articles académiques, livres, et rapports de recherche portant sur les critères d'attribution des prestations sociales et les politiques d'incitation à l'emploi pour les personnes en situation de handicap, décisions de justice sur les litiges relatifs à l'attribution des prestations sociales et sur les recours contre les refus de prestations.
- **Format** : Documents sous formats texte (.docx, .pdf). Durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (TXT (Text)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre de documents analysés. On peut s'attendre à la gestion de plusieurs centaines de pages de documents textuels (entre 200 à 500 pages).

Tâche 2.1 : La marginalisation des personnes handicapées de l'emploi et au sein de l'emploi : un aperçu général (coord. C. Bouchet, LIEPP)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 2.1 : La marginalisation des personnes en situation de handicap sur le marché du travail (coord. C. Bouchet, LIEPP), nous produirons de nouvelles données qualitatives en réalisant des entretiens biographiques avec des personnes en situation de handicap en les complétant avec des données quantitatives préexistantes provenant d'enquêtes statistiques françaises et européennes. Cette approche permettra de mieux comprendre les inégalités d'emploi vécues par les personnes handicapées, au-delà des mesures standard comme le "disability employment gap", en explorant les barrières à l'accès et au maintien dans l'emploi selon le contexte de vie, le type de handicap et d'autres inégalités sociales.

D'une part, de nouvelles données seront recueillies à travers des entretiens biographiques avec des personnes en situation de handicap. Ces entretiens viseront à explorer les expériences individuelles de ces personnes sur le marché du travail, notamment en identifiant les mécanismes d'inégalités spécifiques auxquels elles sont confrontées (comme les asymétries de légitimité sur le lieu de travail, l'accès réduit aux ajustements, etc.) et les significations subjectives qu'elles attachent à ces expériences. Ils permettront également d'examiner comment les inégalités d'emploi varient en fonction de facteurs tels que le type de handicap, le contexte de vie, ou d'autres formes d'inégalités sociales (âge, genre, origine ethnique, etc.).

D'autre part, des données quantitatives préexistantes seront réutilisées pour analyser les inégalités d'emploi des personnes handicapées à différents niveaux. Les sources utilisées incluent des enquêtes statistiques françaises et européennes, telles que l'enquête "Autonomie 2022" en France, ainsi que des enquêtes européennes telles que EU-SILC (European Union Statistics on Income and Living Conditions), EU-LFS (European Union Labour Force Survey), et HETUS (Harmonised European Time Use Surveys). Ces enquêtes fourniront des données quantitatives sur l'état de l'emploi (inactivité, chômage), les conditions de travail (position, salaire, conditions de travail), et d'autres indicateurs socio-économiques pour les personnes en situation de handicap. Cette tâche s'appuiera également sur des publications académiques et des rapports de recherche.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives produites

- **Type de données** : Données qualitatives recueillies par des entretiens biographiques.
- **Format** : Transcriptions d'entretiens sous format texte (.docx, .pdf), notes de terrain sous format texte (.docx, .pdf), enregistrements audio et/ou vidéo (.mp3, .mp4). Durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (TXT (Text) ; MKVV ; MP3) pour le dépôt, le cas échéant..
- **Volume estimé** : Le volume dépendra du nombre d'entretiens réalisés et de leur durée. Par exemple, si 50 à 100 entretiens sont menés, avec une moyenne de 1 à 2 heures par entretien, les transcriptions pourraient représenter plusieurs centaines de pages (estimé entre 500 à 1 000 pages de texte). Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (GO), en fonction de la qualité d'enregistrement et du format de stockage utilisé. Les notes de terrain pourraient représenter environ 50 à 100 pages de texte, selon la durée et la fréquence des observations.

2. Les données quantitatives réutilisées

- **Type de données** : Données quantitatives existantes issues de quatre bases de données principales :
 - L'enquête "Autonomie 2022" est une enquête française centrée sur les conditions de vie et d'autonomie des personnes en situation de handicap ou de dépendance. Elle a été élaborée pour fournir des informations détaillées sur divers aspects de la vie des personnes en perte d'autonomie. Cette enquête regroupe des données démographiques telles que l'âge, le sexe, le statut marital, la composition du ménage et le niveau d'éducation. Elle intègre également des données socio-économiques comme le statut d'emploi, le type de contrat, le revenu et les conditions de travail. En outre, elle collecte des informations sur la santé et la dépendance, incluant l'état de santé, les types et niveaux de handicap, les besoins d'assistance, ainsi que le recours aux services de soins et d'aide à domicile.
 - EU-SILC (European Union Statistics on Income and Living Conditions) est une enquête européenne qui vise à recueillir des données comparables sur les conditions de vie et les revenus des ménages à travers les pays de l'Union Européenne. Cette enquête fournit des informations sur la distribution des revenus, les inégalités économiques et les conditions de vie, avec des données démographiques détaillées telles que l'âge, le sexe, la nationalité, la situation familiale, ainsi que des données socio-économiques incluant le niveau de vie, le revenu, et les dépenses des ménages. Elle permet aussi d'évaluer les conditions de logement et l'accès aux services sociaux, fournissant ainsi une vue d'ensemble des aspects socio-économiques des ménages européens.
 - EU-LFS (European Union Labour Force Survey) est une enquête longitudinale qui suit les dynamiques du marché du travail au sein de l'Union Européenne. Elle fournit des données détaillées sur la participation des individus au marché du travail, leur statut d'emploi, le type de contrat, le secteur d'activité, ainsi que des informations sur les conditions de travail, telles que le salaire, les horaires et la sécurité de l'emploi. L'EU-LFS permet également d'analyser les transitions entre différents états d'emploi, comme les passages de l'emploi au chômage ou à l'inactivité, et d'étudier les caractéristiques des travailleurs dans divers contextes européens.
 - HETUS (Harmonised European Time Use Surveys) est une enquête européenne qui recueille des données sur la façon dont les individus utilisent leur temps tout au long de la journée. Elle offre une perspective sur la répartition du temps entre diverses activités telles que le travail, les loisirs, les soins aux autres, et les tâches domestiques. Les données recueillies incluent des informations sur les habitudes quotidiennes, les conditions de travail, et l'impact des responsabilités familiales sur la participation au marché du travail. L'enquête HETUS permet de comprendre les variations dans l'utilisation du temps en fonction de différents facteurs socio-économiques et personnels à travers les pays européens.
- **Format** : Ces données sont structurées dans des formats numériques standard pour les bases de données statistiques (.CSV (Comma-Separated Values) ; .XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). Durant leur exploitation, des formats fermés seront donc utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume de données est potentiellement important, couvrant de nombreuses variables sur des périodes temporelles étendues, représentant potentiellement plusieurs gigaoctets (GO) de données.

Tâche 2.2 : Handicap et emploi : quelles discriminations ? (coord. C. Bouchet, LIEPP, et M. Mercat-Bruns, LISE)

1a. Comment des nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 2.2 : Handicap et emploi : quelles discriminations (coord. C. Bouchet, LIEPP and M. Mercat-Bruns, LISE), nous produirons des données qualitatives en recueillant des entretiens semi-dirigés auprès de personnes en situation de handicap et réutiliserons des données préexistantes provenant d'enquêtes statistiques et d'archives pour analyser la manière dont les personnes handicapées identifient une discrimination à leur encontre et comprendre la manière dont

elles choisissent d'y réagir, voire de l'anticiper. Les sources archivistiques restent à déterminer pour le moment. L'utilisation des données quantitatives et qualitatives existantes s'appuie notamment sur un partenariat spécifique en cours d'élaboration avec le Défenseur des droits. Il est envisagé que ces résultats soient éclairés par une approche comparatiste de la jurisprudence française et américaine.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives produites

- **Type de données** : Données qualitatives recueillies par entretiens semi-dirigés auprès de personnes en situation de handicap.
- **Format** : Transcriptions d'entretiens (fichiers texte, .docx, .pdf), notes de terrain (fichiers texte, ex. .docx, .pdf), enregistrements audio et/ou vidéo (fichiers numériques, ex. .mp3, .mp4), sous des formats fermés durant leur exploitation puis convertis sous formats libres (TXT (Text) ; MKV ; MP3) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre d'entretiens réalisés et de leur durée. Par exemple, si 50 à 100 entretiens sont menés, avec une moyenne de 1 à 2 heures par entretien, les transcriptions pourraient représenter plusieurs centaines de pages (estimé entre 500 à 1 000 pages de texte). Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (GO), en fonction de la qualité d'enregistrement et du format de stockage utilisé. Les notes de terrain pourraient représenter environ 50 à 100 pages de texte, selon la durée et la fréquence des observations.

2. Les données quantitatives réutilisées

- **Type de données** : Données quantitatives existantes issues de deux bases de données:

- L'enquête Accès aux droits du Défenseur des droits a été créée afin de disposer de connaissances fiables sur les discriminations, les droits de l'enfant, la déontologie des forces de sécurité et les relations avec les services publics. Cette enquête statistique vise à établir un état des lieux inédit des situations qui relèvent de ses missions. Elle collecte des informations précises sur le profil social et démographique des personnes afin de mieux caractériser les groupes sociaux concernés par ces différentes situations. Pour chaque thème, face aux situations de manquement aux droits, il est demandé aux personnes si elles connaissent les recours et si, selon elles, ce type d'atteinte aux droits est fréquent dans la société française. Avant de recueillir leur expérience propre, il leur est également demandé si personnellement elles ont déjà été témoin de telles situations. Pour chaque expérience individuelle rapportée, elle recueille, enfin, des éléments sur les recours au droit mobilisés (ou non) face aux situations vécues.

- L'enquête "Autonomie 2022" est une enquête française centrée sur les conditions de vie et d'autonomie des personnes en situation de handicap ou de dépendance. Elle a été élaborée pour fournir des informations détaillées sur divers aspects de la vie des personnes en perte d'autonomie. Cette enquête regroupe des données démographiques telles que l'âge, le sexe, le statut marital, la composition du ménage et le niveau d'éducation. Elle intègre également des données socio-économiques comme le statut d'emploi, le type de contrat, le revenu et les conditions de travail. En outre, elle collecte des informations sur la santé et la dépendance, incluant l'état de santé, les types et niveaux de handicap, les besoins d'assistance, ainsi que le recours aux services de soins et d'aide à domicile.

- **Format** : Ces données sont structurées dans des formats numériques standard pour les bases de données statistiques (.CSV (Comma-Separated Values) ; .XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). Durant leur exploitation, des formats fermés seront donc utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume de données est potentiellement important, couvrant de nombreuses variables sur des périodes temporelles étendues, représentant potentiellement plusieurs gigaoctets (GO) de données.

3. Les données qualitatives réutilisées

- **Types de données** : Rapports, documents administratifs, directives, articles académiques, livres, et rapports de recherche, décisions de justice sur les litiges, jurisprudence.
- **Format** : Documents sous formats texte (.docx, .pdf). Durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (TXT (Text)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre de documents analysés. On peut s'attendre à la gestion de plusieurs centaines de pages de documents textuels (entre 200 à 500 pages).

Tâche 2.3 : Accès à l'emploi et autonomisation économique des jeunes handicapés (coord. J-F Trani, LISE)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la tâche 2.3, portant sur l'accès à l'emploi et à l'autonomisation économique des jeunes en situation de handicap (coord. J-F Trani, LISE), nous réutiliserons des données provenant d'enquêtes statistiques pour décrire la situation française des jeunes en situation de handicap hors et dans l'emploi. La base de données VQS (Vie quotidienne et santé), EU-SILC (European Union Statistics on Income and Living Conditions), l'enquête sur l'emploi, le chômage et l'inactivité (appelée Enquête Emploi en continu - EEC) et les enquêtes Génération 2010 et 2021 du CEREQ (Centre d'études et de recherches sur les qualifications) pourront être mobilisées.

A partir de là, une comparaison avec un pays à revenu intermédiaire, l'Afrique du Sud, sera mise en place à travers la production de données originales. Ces données seront mixtes, avec une part quantitative, portant sur 200 jeunes personnes handicapées âgées de 18 à 35 ans et sélectionnées aléatoirement parmi un panel de ménages en Afrique du Sud et une part qualitative, avec la mise en place d'ateliers de co-conception d'un modèle en groupe (CMG) auprès de 60 jeunes personnes handicapées âgées de 18 à 35 ans en France et en Afrique du Sud présentant différents types de handicap afin de modéliser leurs parcours. Les critères et modalités d'établissement du panel restent à définir à ce stade. En effet, en ce qui concerne l'Afrique du Sud notamment, l'idée d'une enquête randomisée sera difficile, au regard des profils des jeunes (en situation de handicap, en maintien ou en insertion professionnelle). Il est envisagé d'utiliser des intermédiaires, notamment des associations et des entreprises locales. Ces éléments seront précisés ensuite.

Cette approche permettra une compréhension approfondie des expériences vécues par les jeunes en situation de handicap, en France et en Afrique du Sud, tout en exploitant des données quantitatives déjà existantes en France.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives produites

- **Types de données** : données qualitatives recueillies lors d'ateliers de co-conception d'un modèle en groupe (CMG) auprès de 60 jeunes handicapés âgés de 18 à 35 ans en France et en Afrique du Sud.
- **Format** : Transcriptions d'ateliers (fichiers texte, .docx, .pdf), notes de terrain (fichiers texte, ex. .docx, .pdf), enregistrements audio et/ou vidéo (fichiers numériques, ex. .mp3, .mp4), sous des formats fermés durant leur exploitation puis convertis sous formats libres (TXT (Text) ; MKVV ; MP3) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre d'ateliers réalisés et de leur durée. Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (GO), en fonction de la qualité d'enregistrement et du format de stockage utilisé. Les notes de terrain pourraient représenter environ 50 à 100 pages de texte, selon la durée et la fréquence des observations. Une compression pourra être à prévoir. Par ailleurs, la question du floutage et/ou du droit d'image sera abordée dans les prochains mois.

2. Les données quantitatives produites

- **Types de données** : étude quantitative par questionnaire portant sur 200 jeunes personnes handicapées âgées de 18 à 35 ans et sélectionnées aléatoirement parmi un panel de ménages en Afrique du Sud.
- **Format** : Ces données sont structurées dans des formats numériques standard pour les bases de données statistiques (.CSV (Comma-Separated Values) ; .XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). Durant leur exploitation, des formats fermés seront donc utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume de données est potentiellement important, en fonction du nombre de réponses et des variables prises en compte, représentant potentiellement plusieurs gigaoctets (GO) de données.

3. Les données quantitatives réutilisées

- **Types de données** : données quantitatives collectées à partir des enquêtes statistiques pour décrire la situation française des jeunes en situation de handicap hors et dans l'emploi. La base de données VQS (Vie quotidienne et santé), EU-SILC (European Union Statistics on Income and Living Conditions), l'enquête sur l'emploi, le chômage et l'inactivité (appelée Enquête Emploi en continu - EEC) et les enquêtes Génération 2010 et 2021 du CEREQ (Centre d'études et de recherches sur les qualifications).
- L'enquête Vie quotidienne et santé (VQS) est réalisée en France par l'Institut national de la statistique et des études économiques (INSEE). Cette enquête vise à recueillir des informations détaillées sur les conditions de vie des ménages et leur état de santé, en se concentrant sur l'impact de la santé sur la vie quotidienne des individus.
- L'enquête EU-SILC (European Union Statistics on Income and Living Conditions) est centrée sur la collecte d'informations relatives aux revenus, à la pauvreté, à l'exclusion sociale et aux conditions de vie des ménages dans l'Union européenne. Cette enquête a pour objectif de fournir des statistiques comparables entre les différents États membres de l'UE.
- L'enquête Génération 2010 et 2021 du CEREQ (Centre d'études et de recherches sur les qualifications) vise à étudier l'accès à l'emploi des jeunes à l'issue de leur formation initiale. Une « Génération » est constituée par des jeunes ayant en commun d'être sortis du système éducatif une même année, quel que soit le niveau ou le domaine de formation atteint. L'enquête permet de reconstituer les parcours des jeunes au cours de leurs trois premières années de vie active et de les analyser au regard notamment de la trajectoire scolaire et des diplômes obtenus, mais aussi de leur environnement familial et géographique.

- L'enquête sur l'emploi, le chômage et l'inactivité (appelée Enquête Emploi en continu -EEC) est une enquête statistique réalisée par l'Institut national de la statistique et des études économiques (INSEE) en France. Cette enquête a pour objectif de fournir des informations détaillées et continues sur le marché du travail de manière structurelle et conjoncturelle. C'est la seule source fournissant une mesure des concepts d'activité, de chômage et d'emploi tels qu'ils sont définis par le Bureau international du travail (BIT). Elle correspond à l'enquête sur les forces de travail (EFT) qui est la déclinaison française de l'enquête « Labour Force Survey » (LFS) définie au niveau européen.

- **Format** : Fichiers de données statistiques (.CSV (Comma-Separated Values)) ou (.XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). De même, durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé** : Plusieurs fichiers de données statistiques, représentant potentiellement plusieurs mégaoctets (Mo) à quelques gigaoctets (Go), selon la taille des bases de données produites.

Tâche 3.1 - volet quantitatif : Mise en œuvre, appropriation et effets des politiques de quotas (Y. L'horty, ERUDITE)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la tâche 3.1 Mise en œuvre, appropriation et effets des politiques de quotas, le volet quantitatif assuré par Y. L'horty (ERUDITE) permettra de pallier le manque d'évaluation en matière d'obligation d'emploi des travailleurs handicapés (OETH), pourtant pierre angulaire des politiques d'emploi liées au handicap en France. A ce titre, la tâche 3.1 produira une revue systématique de la littérature sur les politiques internationales de quotas d'emploi, dans le double objectif de "policy mapping" (en relation avec la tâche 1.1) et d'un bilan des évaluations existantes de la politique française d'OETH.

A partir de là, une évaluation quantitative inédite sera produite à partir de méthodes de testing (envoi de candidatures spontanées de personnes en situation de handicap auprès d'entreprises), en comparant deux populations d'entreprises (10 000 au total) selon leur éligibilité ou leur absence d'éligibilité à l'obligation.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données quantitatives produites

- **Type de données** : Données quantitatives produites à partir d'une méthode de testing et d'une base de données adhoc des accords handicap d'entreprises
- **Format** : Fichiers de données statistiques (.CSV (Comma-Separated Values)) ou (.XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). De même, durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé** : Pour le testing, 10 000 observations sont prévues, représentant potentiellement plusieurs gigaoctets (GO) de données. Pour la base de données adhoc des accords handicap d'entreprises, les volumes pourraient atteindre plusieurs gigaoctets (GO) en fonction de l'exhaustivité de la base.

Tâche 3.1 - volet qualitatif : Mise en œuvre, appropriation et effets des politiques de quotas (C. Spieser, LISE)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la tâche 3.1 Mise en œuvre, appropriation et effets des politiques de quotas, le volet qualitatif assuré par C. Spieser (LISE) permettra de pallier le manque d'évaluation en matière d'obligation d'emploi des travailleurs handicapés (OETH), pourtant pierre angulaire des politiques d'emploi liées au handicap en France. A ce titre, la tâche

3.1 produira une revue systématique de la littérature sur les politiques internationales de quotas d'emploi, dans le double objectif de "policy mapping" (en relation avec la tâche 1.1) et d'un bilan des évaluations existantes de la politique française d'OETH.

A partir de là, une évaluation qualitative est envisagée à partir d'un corpus diversifié de données. Des documents visuels issus de l'entreprise et la communication institutionnelle, montrant les personnes handicapées au travail (supports de communication RH, annonces de recrutement ciblées, supports de communication RSE, etc.) seront analysés. Une collecte de données Web, des entretiens et des observations permettront d'approcher les « courtiers » de l'OETH (consultants, ONG...). Des entretiens semi-directifs mutualisés du WP2 du projet Prespol seront analysés sous l'angle de la réception du système de quotas par les travailleurs handicapés. Une enquête en ligne et sur papier sur la mise en œuvre du système de quotas, notamment le droit à un poste et à des conditions de travail adaptés, sera mis en place. Enfin, des entretiens avec des professionnels des ressources humaines et du handicap au sein des entreprises seront menés.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives produites

- **Type de données** : données qualitatives via des images ; des fichiers ou des captures d'écran issus de sites web ; des entretiens ; des observations ; une enquête en ligne et sur papier
- **Format** : Transcriptions d'entretiens (fichiers texte .docx, .pdf), notes de terrain (fichiers texte .docx, .pdf), enregistrements audio et/ou vidéo (fichiers numériques, ex. .mp3, .mp4), images (JPG ; PNG), fichiers ou captures d'écran issus de sites web (HTML ; JPG), résultats du questionnaire (fichiers texte .docx, .pdf et CSV (Comma-Separated Values)) ou (.XLSX (Excel)). La plupart de ces fichiers seront sous des formats fermés durant leur exploitation puis convertis sous formats libres (TXT (Text) ; MKV ; MP3 ; JPG ; PNG ; CSV) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre d'entretiens réalisés et de leur durée. Par exemple, si 50 à 100 entretiens sont menés, avec une moyenne de 1 à 2 heures par entretien, les transcriptions pourraient représenter plusieurs centaines de pages (estimé entre 500 à 1 000 pages de texte). Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (GO), en fonction de la qualité d'enregistrement et du format de stockage utilisé. Les notes de terrain pourraient représenter environ 50 à 100 pages de texte, selon la durée et la fréquence des observations. Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (GO), en fonction de la qualité d'enregistrement et du format de stockage utilisé. Les notes de terrain pourraient représenter environ 50 à 100 pages de texte, selon la durée et la fréquence des observations. Le volume d'images dépendra du nombre et de la qualité des supports analysés. Le volume de fichiers ou captures d'écran issus de sites web dépendra du nombre et de la qualité des supports analysés.

2. Les données qualitatives réutilisées

- **Type de données** : Données qualitatives recueillies par entretiens semi-dirigés auprès de personnes en situation de handicap dans le W2, tâche 2.2 du projet PRESPOL
- **Format** : Transcriptions d'entretiens (fichiers texte, .docx, .pdf), notes de terrain (fichiers texte, ex. .docx, .pdf), enregistrements audio et/ou vidéo (fichiers numériques, ex. .mp3, .mp4), sous des formats fermés durant leur exploitation puis convertis sous formats libres (TXT (Text) ; MKV ; MP3) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre d'entretiens réalisés et de leur durée. Par exemple, si 50 à 100 entretiens sont menés, avec une moyenne de 1 à 2 heures par entretien, les transcriptions pourraient représenter plusieurs centaines de pages (estimé entre 500 à 1 000 pages de texte). Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (GO), en fonction de la qualité d'enregistrement et du format de stockage utilisé. Les notes de terrain pourraient représenter environ 50 à 100 pages de texte, selon la durée et la fréquence des observations.

Tâche 3.2 : Emploi accompagné (coord. F. Jaffrès, EHESP)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 3.2. Emploi accompagné (coord. F. Jaffrès, EHESP), une enquête qualitative combinera des données produites et réutilisées. D'une part, des méthodes ethnographiques et d'analyse politique pour mieux comprendre la mise en place du dispositif de l'emploi accompagné, très développé à l'international mais assez peu en France, seront utilisées. En lien avec les tâches 2.2 et 2.3, cette étude comprendra des entretiens biographiques avec des personnes en situation de handicap utilisant le dispositif d'emploi accompagné. D'autre part, la tâche comprendra des données réutilisées, quantitatives et qualitatives, à partir des travaux de l'ANSA (Agence nationale des solidarités actives). Un

partenariat spécifique est en cours d'élaboration.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives produites

- **Types de données** : données qualitatives recueillies par des entretiens avec des personnes en situation de handicap recourant à l'emploi accompagné ; observations de groupes de travail professionnels dédiés à l'emploi accompagné et d'entretiens professionnel-personne accompagnée.
- **Format** : Transcriptions d'entretiens (fichiers texte, .docx, .pdf), notes de terrain (fichiers texte, ex. .docx, .pdf), enregistrements audio et/ou vidéo (fichiers numériques, ex. .mp3, .mp4), sous des formats fermés durant leur exploitation puis convertis sous formats libres (TXT (Text) ; MKVV ; MP3) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre d'entretiens réalisés et de leur durée. Par exemple, si 50 à 100 entretiens sont menés, avec une moyenne de 1 à 2 heures par entretien, les transcriptions pourraient représenter plusieurs centaines de pages (estimé entre 500 à 1 000 pages de texte). Les fichiers audio ou vidéo pourraient atteindre plusieurs gigaoctets (GO), en fonction de la qualité d'enregistrement et du format de stockage utilisé. Les notes de terrain pourraient représenter environ 50 à 100 pages de texte, selon la durée et la fréquence des observations.

2. Les données qualitatives réutilisées

- **Types de données** : données qualitatives issues de rapports, guidelines, documents administratifs, directives, articles académiques, livres, rapports de recherche, données législatives.
- **Format** : Documents sous formats texte (.docx, .pdf). Durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (TXT (Text)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume dépendra du nombre de documents analysés. On peut s'attendre à la gestion de plusieurs centaines de pages de documents textuels (entre 200 à 500 pages).

3. Les données quantitatives réutilisées

- **Type de données** : Données quantitatives existantes issues de la base de données Evaluation de l'emploi accompagné par l'ANSA (Agence nationale des solidarités actives). Dès 2017, le suivi et l'évaluation quantitative du déploiement de l'emploi accompagné ont été confiés à l'Ansa. Cette évaluation s'est traduite par la remontée de données quantitatives de la part des Dispositifs d'Emploi accompagné pendant 3 ans et par la collecte de données qualitatives à travers des visites sur site et de nombreuses rencontres avec des professionnels, des personnes accompagnées, des employeurs et des partenaires.
- **Format** : Données structurées dans des formats numériques standard pour les bases de données statistiques (.CSV (Comma-Separated Values)) ou (.XLSX (Excel)), ou des formats spécifiques à des logiciels statistiques (.DTA pour le logiciel Stata). De même, durant leur exploitation, des formats fermés seront utilisés puis convertis sous formats libres (CSV (Comma-separated values)) pour le dépôt, le cas échéant.
- **Volume estimé** : Le volume de données est potentiellement important, couvrant de nombreuses variables sur des périodes temporelles étendues, représentant potentiellement plusieurs gigaoctets (GO) de données également.

Tâche 4.2 : Médiation scientifique (coord. E. Payen, EHESP, et A. Revillard, LIEPP)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 4.2 : médiation scientifique (coord. E. Payen, EHESP, et A. Revillard, LIEPP), nous produirons de nouvelles données qualitatives et quantitatives en mettant en place diverses activités de dissémination scientifique et de médiation des savoirs, en lien avec les partenaires institutionnels et les organisations de la société civile. Cette approche vise à définir les formats de diffusion scientifique les plus pertinents et à promouvoir l'accessibilité des résultats de recherche auprès d'un large public. La porosité entre données de recherche et données instrumentales sera interrogée durant le projet.

D'une part, de nouvelles données qualitatives seront recueillies à travers la production de portraits de travailleurs handicapés et l'organisation d'expositions hybrides. Les portraits, co-écrits avec les participants de la recherche, permettront d'explorer les parcours individuels des travailleurs handicapés, leurs expériences professionnelles et les obstacles spécifiques auxquels ils sont confrontés. Ces portraits seront complétés par une exposition hybride, qui intégrera également un projet photographique pour illustrer visuellement ces expériences. D'autre part, des données quantitatives pourront être collectées à propos de la réception de ces différents formats.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives produites

- **Type de données** : Données qualitatives recueillies à travers des portraits de travailleurs handicapés, des échanges avec des organisations de la société civile, des partenaires institutionnels, et des productions multimédias (blog de recherche, vidéos, expositions, etc.).
- **Format** : Textes de portraits (fichiers texte, ex. .docx, .pdf), documents multimédias (vidéos en .mp4, affiches en .pdf ou .png), articles de blog (HTML), notes d'entretiens ou de réunions (fichiers texte, ex. .docx, .pdf).
- **Volume estimé** : Les cinquante portraits de travailleurs handicapés (4-5 pages chacun) représenteront environ 200 à 250 pages de texte. Les vidéos (présentations, documentaire) pourraient atteindre plusieurs gigaoctets (GO) selon leur durée et qualité. Les publications en ligne (articles de blog, tweets) occuperont un espace de stockage limité (quelques mégaoctets par publication).

2. Les données quantitatives produites

- **Type de données** : Données quantitatives sur l'interaction et l'engagement des publics avec les différents formats de diffusion (par ex., statistiques de consultation du blog, vues des vidéos, interactions sur Twitter).
- **Format** : Données numériques (tableaux de données au format .csv, .xlsx, ou JSON).
- **Volume estimé** : Le volume dépendra de la fréquence et de l'intensité des consultations et interactions, avec un potentiel de quelques dizaines à centaines de kilooctets par mois.

Tâche 4.3 : Formation (coord. A. Revillard, LIEPP)

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Pour la Tâche 4.3 : Training (coord. A. Revillard, LIEPP), nous produirons de nouvelles données en organisant des activités de formation et de développement des compétences, destinées à renforcer l'émergence d'une nouvelle génération de chercheurs spécialisés dans l'autonomisation économique des personnes handicapées. Ainsi, de nouvelles données seront recueillies à travers les activités de formation, telles que les stages de recherche, les conférences et les projets collectifs sur le terrain. Ces activités généreront des notes de formation, des mémoires d'enquêtes collectives, des rapports de stage et des comptes rendus d'ateliers, qui permettront d'évaluer l'impact des formations sur les compétences et les carrières des jeunes chercheurs. D'autre part, des données préexistantes seront réutilisées pour structurer et enrichir les programmes de formation. Ces données incluront l'utilisation des ressources existantes de formations passées, y compris les cours et modules offerts par les institutions partenaires (EHESP, CNAM, Sciences Po), pour développer de nouveaux contenus pédagogiques ainsi qu'une analyse de données et publications académiques disponibles pour renforcer les contenus de formation et informer les participants sur les développements récents dans le domaine de l'autonomisation économique des personnes handicapées. Ces sources restent à définir.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Les données qualitatives produites

- **Type de données** : Notes de formation, documents pédagogiques et de recherche issus des activités de formation (stages, conférences, projets collectifs), comptes rendus de terrain et d'ateliers pédagogiques.
- **Format** : Documents texte (fichiers texte, ex. .docx, .pdf), présentations (fichiers .pptx, .pdf), notes de terrain (fichiers texte, ex. .docx, .pdf).
- **Volume estimé** : Plusieurs centaines de pages de documents texte et de présentation, estimé entre 500 et 1000 pages au total.

2. Les données quantitatives réutilisées

- **Type de données** : Données issues de bases de données institutionnelles, à définir.
- **Format** : Tableaux de données (fichiers .csv, .xlsx).
- **Volume estimé** : Quelques dizaines de kilooctets à quelques mégaoctets selon le nombre de participants et d'évaluations recueillies.

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

1. Modalités de nommage des fichiers

Pour déterminer les modalités de nommage des fichiers du projet PRESPOL, nous prenons en compte plusieurs contraintes importantes dans le nommage des fichiers afin de garantir une gestion efficace et cohérente des données.

- Nous veillons à ce que les noms de fichiers ne soient pas trop longs, en limitant leur longueur à un maximum de 25 caractères. Cependant, en raison de la diversité des données, nous anticipons déjà un dépassement de ce nombre de caractères. Un travail sur ce problème est en cours.
- Nous évitons l'utilisation de caractères spéciaux tels que & * % \$ £ ° { ! @, qui peuvent être réservés à des tâches spécifiques selon les systèmes d'exploitation, ce qui réduit le risque de conflits ou de dysfonctionnements.
- Pour assurer une compatibilité maximale, nous préférons utiliser le caractère souligné (_) ou le tiret (-) au lieu du point (.) ou de l'espace, ces derniers pouvant être interprétés différemment par les systèmes d'exploitation.
- De plus, nous incluons autant que possible des informations descriptives dans les noms de fichiers pour en faciliter l'identification, indépendamment de leur lieu de stockage.
- Les dates incluses dans les noms de fichiers sont formatées de manière consistante pour éviter toute confusion.
- Nous prenons également en compte que les applications et instruments qui accèdent aux fichiers ne sont généralement pas sensibles à la casse, et avons considéré par défaut que des variations comme DATA, Data, et data sont identiques.
- Enfin, nous utilisons des extensions (.xls, .ssd, .txt, etc.) pertinentes pour refléter correctement l'environnement logiciel dans lequel le fichier a été créé.

De manière générale, nous nous sommes accordés sur la nécessité d'éviter de conserver trop de versions intermédiaires d'un même fichier, car cela pourrait entraîner de la confusion, tant pour les utilisateurs actuels que futurs. La nomenclature a été pensée de manière à ce qu'une personne n'ayant pas participé à la création des données ou à leur organisation puisse s'orienter.

Exemples de nomenclature

A date, les conventions de nommage des fichiers ont pris en compte les recommandations ainsi que les pratiques assez hétérogènes des chercheurs et chercheuses. Cette proposition théorique sera amenée à évoluer en fonction de la typologie des données et du respect effectif des recommandations à l'usage.

Structure Générale du Nom du Fichier : PRESPOL_Revillard_20240821_MEET_RECH_INTERP_01.docx

1. Acronyme du projet : "PRESPOL"

2. Nom du responsable de la collecte : Utilisation du nom de famille du chercheur ou de la chercheuse exclusivement.

Exemple : "Revillard" pour "Anne Revillard"

3. Date : Format YYYYMMDD pour garantir la cohérence et faciliter le tri chronologique. La date correspond à la date de collecte. Exemple : 20240820 pour le 20 août 2024.

4. Abréviations adaptées aux données importées et description :

- Formulaire en ligne (FORM_ONL)

Draft : FORM_ONL_DRAFT

Résultats bruts : FORM_ONL_RAW

Mémos : FORM_ONL_MEMO

Interprétation : FORM_ONL_INTERP

- Formulaire papier (FORM_PAP)

Draft : FORM_PAP_DRAFT

Résultats bruts : FORM_PAP_RAW

Mémos : FORM_PAP_MEMO

Interprétation : FORM_PAP_INTERP

- Boîtier Box et clé CASD (BOX_CASD)

Données sécurisées : BOX_CASD_SEC

Exportations : BOX_CASD_EXPT

Interprétation : BOX_CASD_INTERP

Mémos : BOX_CASD_MEMO

- Réseau QUETELET (RQLET)

Données collectées : RQLET_COL

Interprétation : RQLET_INTERP

Rapports : RQLET_RPT

Mémos : RQLET_MEMO

- Réseaux sociaux / Internet (SOC_INT)

Données brutes : SOC_INT_RAW

Interprétation : SOC_INT_INTERP

Mémos : SOC_INT_MEMO

Extraits : SOC_INT_EXTR

- Testing (TEST)

Données de test : TEST_DATA

Résultats : TEST_RSLT

Mémos : TEST_MEMO

Interprétation : TEST_INTERP

- Traitement automatisé des données (AUTODAT)

Données brutes : AUTODAT_RAW

Résultats : AUTODAT_RSLT

Extraits : AUTODAT_EXTR

Interprétation : AUTODAT_INTERP

- Ateliers CMG (CMG_WKS)

Données d'atelier : CMG_WKS_DATA

Résultats : CMG_WKS_RSLT

Mémos : CMG_WKS_MEMO

Interprétation : CMG_WKS_INTERP

- Réunions de travail entre chercheurs (MEET_RECH)

Compte rendu : MEET_RECH_CRT

Notes : MEET_RECH_NOT

Résultats : MEET_RECH_RSLT

Interprétation : MEET_RECH_INTERP

- Entretiens (ENT) :

Audio: ENT_AUDIO

Transcriptions : ENT_TRANS

Résumé : ENT_SUM

Interprétation : ENT_INTERP

Sortie d'analyse de contenu : ENT_AC

- Sources écrites, archives papier ou numériques (SRC_ECR) :

Documents scannés : SRC_ECR_SCAN

Documents numériques : SRC_ECR_DIGI

Archives papier : SRC_ECR_PAP

Interprétation : SRC_ECR_INTERP

- Observations directes (OBS_DIR) :

Notes d'observation : OBS_DIR_NOT

Résultats bruts : OBS_DIR_RAW

Résumé : OBS_DIR_SUM

Interprétation : OBS_DIR_INTERP

- Focus groups (FOC_GRP) :

Enregistrements : FOC_GRP_REC

Transcriptions : FOC_GRP_TRANS

Résumé : FOC_GRP_SUM

Interprétation : FOC_GRP_INTERP

5. Numérotation par ordre d'ajout sur le fichier partagé pour les métadonnées : "01"

6. Extension : Indiquer l'extension du fichier pour refléter le format (ex. .xls, .txt, .pdf....).

2. Choix du standard de métadonnées

Pour le projet PRESPOL, trois types de métadonnées nous semblent essentielles à assurer à court et long termes :

- Les métadonnées descriptives (titre, auteur, sujets, mots-clés, date...)
- Les métadonnées de provenance (origine, traitement...)
- Les métadonnées techniques et administratives (licences...)

Il nous semble opportun de privilégier le standard de métadonnées Dublin Core. Ce standard correspond à un cadre international utilisé pour documenter et gérer les données, en particulier celles provenant des sciences sociales, économiques, et comportementales.

Tout d'abord, il favorise la standardisation et la cohérence des métadonnées. En offrant une structure normalisée pour la description des données, ce standard assure une uniformité qui garantit que les métadonnées sont créées et interprétées de manière homogène. Cette uniformité facilite l'interopérabilité et le partage, essentielle dans le cadre de ce consortium multi parties prenantes.

En matière de documentation complète et détaillée, Dublin Core offre une description approfondie des données, incluant la méthodologie de collecte, les variables, les codes de valeur, et les concepts théoriques sous-jacents. Cette documentation détaillée assure une traçabilité de nos données complexes en répondant à notre impératif d'assurer les trois types de métadonnées identifiées.

En outre, Dublin Core est versatile et peut être utilisé pour gérer à la fois les données quantitatives et qualitatives. Cette flexibilité permet de documenter divers types de données, qu'il s'agisse de grandes enquêtes statistiques ou d'observations qualitatives, répondant ainsi aux besoins variés des chercheurs de PRESPOL.

Enfin, Dublin Core correspond à l'entrepôt Nakala, privilégié par le consortium à date.

3. Sensibilisation aux métadonnées

La mise en place d'un standard au sein du projet Prespol commence par des échanges en plénières lors du COPIL mais également, au fil de l'eau, par une sensibilisation des équipes dans le cadre de rencontres entre l'équipe de coordination et les responsables de tâches et de WP. Ces sessions permettront aux membres de comprendre les principes et l'importance des métadonnées Dublin Core, ainsi que la manière dont elles seront appliquées dans le projet.

Parallèlement, l'équipe de coordination fournira des ressources et des trames avec des exemples concrets. Cela permettra aux équipes d'avoir une référence claire et de mieux appréhender l'intégration du standard dans leur travail quotidien. En effet, il est crucial d'intégrer ces outils dans le flux de travail du projet, en veillant à ce que la documentation des données se fasse de manière continue et non de manière rétrospective.

Avant de généraliser l'utilisation du standard choisi, il est envisagé de procéder à une implémentation pilote. Cette étape consiste à appliquer le standard Dublin Core à un sous-ensemble de données du projet Prespol pour identifier les défis potentiels et ajuster le modèle de métadonnées en conséquence. Une fois le pilote réalisé, il est prévu de recueillir les retours d'expérience des utilisateurs pour évaluer la facilité d'utilisation et l'efficacité du standard. Ces retours permettront d'apporter les ajustements nécessaires avant une mise en œuvre à grande échelle.

4. Mise en place du Dublin Core au sein de PRESPOL

Une fois la sensibilisation réalisée et en fonction des retours de l'implémentation pilote, l'étape suivante consiste à s'approprier les outils de documentation compatibles avec le standard choisi. Le développement d'un modèle de métadonnées est une étape clé pour adapter le standard aux spécificités du projet PRESPOL. Nous créerons un profil de standard spécifique au projet, qui définira quelles sections du standard Dublin Core sont nécessaires (telles que les descriptions d'études, les variables, ou les questions d'enquête) et comment elles seront utilisées. Un tableur excel partagé via Share docs sera mis en place, avec une trame pré-remplie proposée par la coordination et validée collectivement en amont.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Pour garantir la qualité des données utilisées dans le cadre du projet PRESPOL, plusieurs mesures seront mises en place. En fonction des tâches, les chercheurs et chercheuses souhaitent procéder par :

- **Description détaillée** : Les méthodes de collecte, d'analyse et de traitement des données sont décrites.
- **Répliquabilité** : Les chercheurs de PRESPOL évalueront si les résultats peuvent être reproduits avec les mêmes données et méthodes.
- **Accès aux données brutes** : Les chercheurs de PRESPOL pourront avoir accès à vos données brutes.
- **Utilisation de checklists spécifiques aux disciplines** : Les chercheurs de PRESPOL utiliseront des listes de contrôle comme CONSORT pour les essais cliniques, PRISMA pour les revues systématiques, ou COREQ pour les études qualitatives.
- **Validation croisée** : Les résultats pourront être comparés à d'autres sources de données ou études similaires pour vérifier leur cohérence.
- **Pré-enregistrement des études** : Les chercheurs de PRESPOL enregistreront les protocoles d'étude et les plans d'analyse avant de collecter les données.
- **Analyse collective des biais potentiels** : Les chercheurs de PRESPOL pourront discuter collectivement des biais

possibles dans la collecte et l'analyse des données.

- **Tests de robustesse** : Les chercheurs de PRESPOl pourront effectuer des tests supplémentaires pour vérifier que les résultats ne sont pas sensibles à des modifications mineures des méthodes ou des hypothèses.

Seule la double saisie des données, c'est-à-dire la méthode consistant à entrer deux fois les données indépendamment, n'est pas envisagée à ce stade.

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Pour le stockage et la sauvegarde des données et des métadonnées tout au long du processus de recherche, l'environnement Huma-Num sera privilégié. Il s'agit d'une infrastructure numérique destinée à la recherche en sciences humaines et sociales en France. Elle est gérée par le CNRS (Centre National de la Recherche Scientifique) et vise à faciliter l'accès, le stockage et la diffusion de données numériques produites par la recherche dans ces disciplines. Plus précisément, deux outils seront privilégiés :

- **Sharedocs** : un service de gestion et de partage de documents destiné aux chercheurs. Il permet aux utilisateurs de stocker, organiser, partager et collaborer sur des documents numériques. A date, la demande de création d'un compte projet a été demandé pour les éléments relevant de la coordination. D'autres comptes projets seront également également spécifiquement pour les travaux collaboratifs de chercheurs, impliquant la manipulation de données sensibles ou personnelles. Ces comptes permettent un accès réglementé et clair tout au long du projet.
- **Huma-Num Box** : un service de stockage de fichiers plus volumineux, avec un objectif plus spécifique de stockage de données de recherche et d'accès limité, adapté à certaines tâches.

En ce qui concerne les métadonnées, un tableur sera mis en place très tôt dans le projet afin que les responsables de tâches puissent remplir au fur et à mesure les informations relatives à leurs données de recherche. La trame sera structurée à partir des items du Dublin core, standard de métadonnées utilisé pour le dépôt sur Nakala. Le tableur des métadonnées sera accessible aux chercheurs via Sharedocs et sera actualisé au fil de l'eau.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Pour le projet PRESPOl, nous allons mettre en place une stratégie rigoureuse de gestion et de protection des données afin d'assurer leur sécurité, leur intégrité et leur disponibilité tout au long de la recherche.

En premier lieu, le recours à un environnement numérique institutionnel, comme Huma-Num, permet d'assurer la sécurité des données et la protection des données sensibles tout au long du processus de recherche grâce à plusieurs caractéristiques clés :

1. Ces serveurs sont gérés par des institutions académiques et des organismes de recherche, garantissant un niveau élevé de sécurité physique et logique.
2. Ils mettent en place des mécanismes de contrôle d'accès rigoureux. Les chercheurs peuvent définir des permissions spécifiques pour les différents niveaux de données, s'assurant ainsi que seules les personnes autorisées ont accès aux informations sensibles.
3. Les données stockées sont souvent chiffrées, tant au repos qu'en transit. Cela signifie que même en cas d'accès non autorisé, les données restent protégées par des protocoles de chiffrement sophistiqués, empêchant leur lecture ou leur utilisation non autorisée.
4. Ces serveurs sont conçus pour se conformer aux réglementations en vigueur sur la protection des données, telles que le RGPD en Europe. Cela inclut des pratiques de gestion des données sensibles et des procédures pour la gestion des demandes d'accès aux données.
5. L'infrastructure assure des sauvegardes régulières et dispose de mécanismes de redondance pour garantir la continuité et la disponibilité des données. En cas de défaillance du système ou de perte de données, les informations peuvent être récupérées rapidement grâce à ces sauvegardes.
6. Ces serveurs offrent des fonctionnalités d'audit et de traçabilité qui permettent de suivre l'accès et les modifications apportées aux données. Cela aide à détecter toute activité suspecte et à garantir la conformité avec les pratiques de gestion des données.

Ensuite, de manière générale, nous appliquerons la règle des 3-2-1 pour le stockage des données. Cela signifie que chaque donnée sera conservée sous forme de trois copies identiques, stockées sur au moins deux supports différents, tels que le

PC du chercheur ou de la chercheuse, une clé USB, un disque dur externe, ou des serveurs institutionnels adaptés à la gestion de données sensibles, notamment Huma-Num. Les chercheurs seront invités à conserver l'une de ces copies dans un emplacement distinct, hors du bureau habituel, afin de prévenir les risques liés à des événements locaux, comme des incidents matériels ou des catastrophes naturelles. En cas d'incident, la récupération des données pourra s'effectuer sur les supports de secours anticipés dans le cadre du stockage des données. L'équipe de coordination du projet rappellera régulièrement ces règles, s'assurera de leur respect et mettra en avant la nécessité d'une mise à jour des systèmes d'exploitation et ordinateurs.

Enfin, l'authentification pour accéder à l'environnement Huma-Num repose sur le système d'identification appelé HumanID, qui offre une méthode sécurisée pour accéder aux services de la plateforme. Ce système permet aux utilisateurs de se connecter à l'ensemble des services offerts par Huma-Num à l'aide d'un identifiant unique et permet :

- Une authentification fédérée permettant de se connecter via les identifiants de leur institution (comme une université ou un centre de recherche) sans avoir à créer un nouveau compte. Les utilisateurs peuvent accéder à divers services avec un seul identifiant, ce qui limite les risques d'attaques par mots de passe faibles ou réutilisés tout en étant pratique au quotidien.
- Une vérification de l'identité sécurisée facilitant l'authentification pour les utilisateurs via OpenID Connect.

La coordination sensibilisera les chercheurs et chercheuses à l'utilisation de mots de passe à niveau de protection élevé (longue chaîne de caractères, majuscules, chiffre, caractères spéciaux) en incitant par exemple à utiliser le logiciel "1password", qui génère automatiquement des mots de passe à haut niveau de protection.

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Les pratiques de traitement des données seront strictement conformes au Règlement Général sur la Protection des Données (RGPD). Cela inclut la collecte, le stockage et le traitement des données personnelles de manière transparente, légale et équitable.

- **Anonymisation et Pseudonymisation** : Lorsque cela est possible, les données personnelles seront anonymisées ou pseudonymisées pour réduire les risques associés à leur traitement. Cette approche limite l'identification des individus à partir des données traitées.
- **Contrôles d'Accès** : Des mesures strictes seront mises en place pour contrôler l'accès aux données personnelles. Seules les personnes autorisées, ayant besoin d'accéder aux données pour leurs tâches spécifiques, pourront y accéder, et ce, via des systèmes sécurisés comme Huma-Num Box ou certains coffres-forts institutionnels.
- **Sécurité des Données** : Les données personnelles seront stockées et transférées en utilisant des technologies de sécurité avancées, telles que le chiffrement des données en transit et au repos. L'environnement Huma-Num offre des garanties de sécurité renforcées, comme le chiffrement lors de la transmission et du stockage des données.
- **Formation et Sensibilisation** : Les chercheurs et le personnel impliqués seront formés aux bonnes pratiques de gestion des données à caractère personnel et aux exigences légales. Des sessions de sensibilisation seront organisées pour garantir une compréhension et une application appropriée des règles de protection des données.
- **Gestion des Violations de Données** : En cas d'incident, des mesures correctives seront prises rapidement pour minimiser les impacts et informer les autorités compétentes conformément aux exigences légales.
- **Documentation** : La conformité sera assurée par la documentation rigoureuse des processus de traitement des données et des choix de gestion.
- **Consentement Informatif** : Plus précisément, avant la collecte de données personnelles, le consentement des individus sera obtenu selon des modalités choisies par l'ensemble des chercheurs et chercheuses de Prespol. Seules certaines tâches peuvent être concernées par des exceptions pour assurer la réussite de la méthode choisie, comme le testing, cette méthode expérimentale consistant justement à ne pas révéler aux entreprises qu'elles sont testées. Ces exceptions seront précisées et détaillées au cours du projet. Dans le cas de Prespol, une attention particulière est donnée à l'accessibilité cognitive des modalités de consentement.

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les principes de l'accord consortium de PRESPOLE s'alignent sur les pratiques courantes forgées par le Code de la Propriété Intellectuelle qui régit les aspects relatifs aux brevets, droits d'auteur et autres droits de propriété intellectuelle.

L'accord est en cours de rédaction et sera susceptible d'évoluer. A date, sont distingués les résultats et connaissances propres et les résultats et connaissances communes :

1. Résultats et connaissances propres

Les résultats propres, définis comme les résultats générés par établissement de recherche du consortium de PRESPOL de manière exclusive, demeurent la propriété entière de cette Partie. Cette Partie a le droit d'exploiter librement ces Résultats Propres, y compris en déposant des brevets ou en enregistrant d'autres droits de propriété intellectuelle, et ce, à ses propres frais et selon son initiative. Cette liberté d'exploitation est fondamentale pour encourager l'innovation et l'utilisation des résultats générés indépendamment.

Toutefois, l'utilisation de ces Résultats Propres par d'autres chercheurs du consortium PRESPOL est autorisée sous certaines conditions. Le principe de proportionnalité doit être respecté dans les modalités de créditation, telles que la mention de la source, les remerciements, ou l'inclusion en tant que co-auteur, en fonction de l'ampleur du matériau utilisé. Les conditions d'utilisation imposent une demande écrite par mail ainsi qu'un engagement explicite à créditer la source des données employées.

2. Résultats et connaissances communes

L'accès commun aux données de PRESPOL est accordé par principe, bien que des exceptions soient possibles pour certaines tâches plus sensibles. Ainsi, les résultats et connaissances communes intègrent les cas où une personne ayant produit les données accepte de les partager avec un autre chercheur du consortium PRESPOL mais également les cas où plusieurs personnes travaillent collectivement sur un corpus commun. Dans le premier cas, le producteur des données conserve la priorité sur leur exploitation et leur publication. Dans le second, les droits de propriété sont partagés selon un principe de proportionnalité. Autrement dit, chaque Partie est copropriétaire des Résultats Communs en fonction de ses contributions intellectuelles, humaines, matérielles et financières respectives.

Cependant, dans ces deux cas, les conditions d'utilisation incluent une demande écrite par mail, ainsi qu'un engagement à créditer la source des données utilisées. Avant toute exploitation de ces Résultats Communs, les Parties doivent donc prévenir les personnes concernées et se mettre d'accord sur la répartition des droits et des obligations associés à ces résultats. Le principe d'un passage par le comité de pilotage pour le partage de connaissances et résultats communs (et notamment la créditation du producteur et les questions de co-autorat), en fonction de l'importance du corpus utilisé, sera inscrit dans l'accord afin de garder une visibilité des coopérations et protéger les droits des producteurs de données. De manière générale, l'utilisation des connaissances et des résultats est gratuite, non exclusive, non cessible, et sans droit de sous-licence et le principe de proportionnalité s'applique aux modalités de créditation, qui doivent être adaptées à l'ampleur du matériau utilisé.

Seuls les membres du consortium ayant une contribution active et significative à la collecte ou à l'analyse des données seront considérés comme des "contributeurs effectifs". Cette liste est établie et mise à jour par le comité de pilotage (COPIL) à partir de la liste de la version initiale du PGD. Le COPIL procédera à des révisions périodiques des contributions des membres et des accès aux données. Cela permettra d'ajuster l'accès en fonction de l'implication réelle des membres au projet. Ainsi, dans un premier temps, l'accès aux données sera réservé aux membres du COPIL et aux contributeurs effectifs identifiés. Cependant, d'autres membres du consortium pourront demander un accès à certaines données en soumettant une demande écrite au COPIL. Cette demande devra inclure : - Une description de la nature de la recherche et de l'utilisation prévue des données.

- Une justification de la pertinence de l'accès pour la contribution à la recherche. - Les modalités de valorisation du producteur.

L'accord du producteur des données sera toujours requis avant toute attribution d'accès. Le COPIL examinera chaque demande d'accès et prendra une décision basée sur :

- La pertinence de la demande par rapport aux objectifs du projet.
- L'importance théorique et pratique de la contribution du demandeur. - La pertinence des modalités de valorisation du producteur. La décision du COPIL sera validée à majorité simple.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Le projet Prespol est un projet de recherche dédié à la question du handicap. A ce titre, les données collectées et traitées sont à la fois personnelles et sensibles. Dans ce cadre, une réflexion collective nous permet déjà d'identifier certains points :

1. Enjeux éthiques

Au regard du sujet du projet PRESPOL, nous identifions des risques éthiques en termes d'impacts propres aux résultats (usages malveillants, stigmatisation, enjeux politiques...) et aux méthodologies (entretiens personnels et intimes...) pour les participants et la population générale. A ce titre, nous nous engageons à respecter strictement les codes déontologiques pertinents.

Cela inclut le respect de la dignité des participants, la transparence sur les objectifs et méthodes de la recherche, et la soumission du projet à un examen par un comité d'éthique. A ce stade, l'équipe de coordination est en cours d'identification des modalités de cet examen pour l'ensemble du projet. Ces engagements sont conçus pour garantir que notre étude respecte les normes éthiques les plus élevées et protège les droits et le bien-être des participants.

2. Minimisation des Risques

Pour minimiser ces risques, nous avons mis en place plusieurs mesures de précaution. Nous garantissons la confidentialité et l'anonymat des données collectées et choisissons des lieux de collecte sécurisants pour les personnes enquêtées. Ces mesures visent à protéger les participants tout au long de l'étude. A ce stade, le recours à la visio pour les entretiens n'est pas écarté et devra faire l'objet d'une attention particulière afin d'assurer la protection numérique des participants.

Nous veillons également à ce que les résultats soient accessibles et présentés tout au long du projet, de manière à éviter toute stigmatisation ou atteinte aux sensibilités des participants. Ainsi, le WP4 est précisément dédié à ces questions, avec le recrutement d'Emeline Payen, médiatrice scientifique. De même, la gouvernance de PRESPOL permettra un suivi de la recherche en train de se faire via un comité des parties prenantes réunissant des associations du secteur du handicap, représentantes et auto-représentantes, et des acteurs institutionnels. La diffusion des résultats, même instables, est un élément important de la démarche.

3. Consentement Éclairé

Nous avons prévu plusieurs modalités de recueil du consentement pour assurer que les participants sont pleinement informés et consentants. Cela inclut la fourniture d'un feuillet d'information détaillé et la réalisation de présentations orales pour clarifier les aspects de l'étude.

En l'absence de document signé par la personne, la saisie des modalités du consentement par le producteur de la donnée devra être systématique et claire dans un tableur partagé dédié mis à disposition par la coordination pour faciliter l'utilisation des données par le producteur et dans le cadre du partage des données. De plus, un formulaire de consentement écrit à signer sera mis à disposition pour ceux qui préfèrent cette méthode. Chacun de ces supports fera l'objet d'une accessibilisation, notamment cognitive (Facile à lire et à comprendre). Plus concrètement, en matière de gestion des données, il est envisagé de demander l'autorisation pour trois items : être recontacté à propos des résultats de l'étude, être recontacté pour une enquête future ; accepter le versement des données dans un dépôt.

Ces démarches visent à garantir que le consentement est éclairé et obtenu de manière transparente, en s'adaptant aux différentes spécificités des personnes en situation de handicap ou des personnes valides. Ces supports pour l'accord sont utilisés afin d'annoncer et de formaliser les règles concernant les objectifs de la recherche, les modes de collecte et la réutilisation des données. Il est prévu d'échanger des modalités possibles de rétractation pour les personnes enquêtées au cours des différents projets de recherche. De manière générale, la participation est souhaitée sans pression directe ou indirecte exercée par les chercheurs, notamment pour répondre à l'ensemble des questions.

4. Médiation et Intermédiaires

Dans les cas où la tâche nécessite la médiation d'un intermédiaire pour accéder au terrain et aux groupes étudiés, nous veillerons à ce que ces intermédiaires respectent les mêmes normes éthiques que celles appliquées directement par notre équipe. Nous assurerons une médiation transparente, en leur fournissant les informations nécessaires pour garantir qu'ils respectent les exigences éthiques et déontologiques. Le cas échéant, des conventions de partenariat pourront être mises en place.

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

Le partage et le dépôt des données sera géré de manière à respecter les exigences de confidentialité et à garantir une diffusion appropriée en fonction de la nature des données. Les données seront rendues visibles pour permettre aux chercheurs et autres parties intéressées de les consulter tout en mettant en place des mécanismes de contrôle d'accès et les protocoles de sécurité pour assurer que seules les personnes autorisées puissent accéder aux données sensibles et/ou personnelles.

1. Distinction entre données anonymisées et données pseudo-anonymisées

Au sein de PRESPOL, il convient de distinguer deux cas de figure, que nous affinerons au fil du projet :

- Certaines données seront anonymisées pendant et après le projet PRESPOL. Certains producteurs de données pourront altérer ou supprimer tout caractère identifiant à un ensemble de données de telle façon que le sujet des informations personnellement identifiables (IPI) ne puisse plus être identifié directement ou indirectement, que ce soit par le responsable du traitement des IPI seul ou en collaboration avec une quelconque autre partie. Ces données pourront être déposées dans des entrepôts de données, avec un accès largement ouvert. Cette approche permet une large diffusion des données tout en préservant le principe de confidentialité des participants. L'anonymisation interdit également tout démarche ultérieure de ré-identification.
- D'autres données seront pseudo-anonymisées pendant et après le projet PRESPOL. Dans ce cas, la Loi n° 2018-493 du 20 juin 2018, relative à la protection des données personnelles s'applique. En effet, si la pseudonymisation est reconnue comme une mesure de sécurité appropriée pour protéger les données personnelles, elle ne dispense pas des obligations du RGPD. Les données pseudonymisées sont toujours considérées comme des données personnelles, car elles peuvent potentiellement permettre l'identification de la personne physique, et restent donc pleinement soumises à la législation. Ainsi, pour respecter le principe d'accountability, les données pseudo-anonymisées seront collectées pour des finalités légitimes, explicites et déterminées collectivement. La pseudo-anonymisation se prête aux cas qui nécessitent ou permettent un retour en arrière des informations nominatives codées aux informations nominatives primitives, ce qui peut être le cas durant l'analyse, y compris dans le cas de travaux collectifs. Les personnes enquêtées seront informées de cette possibilité. Les modalités de consentement incluront des dispositions spécifiques à cet égard de manière homogénéisée dans le projet.

2. Modalités de partage

Durant l'analyse, les données sensibles et personnelles ne pouvant être anonymisées nécessitent une gestion particulièrement rigoureuse et restrictive. Elles pourront également être déposées dans des entrepôts de données, mais avec un cadre d'accès de diffusion particulier. L'accès à ces données sera restreint et nécessitera un contrat spécifique pour toute réutilisation future. Ce contrat garantira que les conditions d'utilisation des données sensibles sont respectées et que leur utilisation est conforme aux accords établis. Les personnes enquêtées seront informées de cette possibilité.

Au sein de PRESPO, l'environnement Huma-Num sera privilégié. Il s'agit d'une infrastructure numérique destinée à la recherche en sciences humaines et sociales en France. Elle est gérée par le CNRS (Centre National de la Recherche Scientifique) et vise à faciliter l'accès, le stockage et la diffusion de données numériques produites par la recherche dans ces disciplines. Plus précisément, deux outils seront privilégiés :

- Sharedocs : un service de gestion et de partage de documents destiné aux chercheurs. Il permet aux utilisateurs de stocker, organiser, partager et collaborer sur des documents numériques.
- Huma-Num Box : un service de stockage de fichiers plus volumineux, avec un objectif plus spécifique de stockage de données de recherche adapté à certaines tâches.

3.Modalités de dépôt

L'équipe PRESPO a mesuré les avantages et inconvénients de plusieurs solutions, y compris Datasciencespo, indiqué initialement comme dépôt privilégié. Cependant, à date, nous envisageons finalement d'avoir recours à Nakala (environnement Huma-Num). Cet entrepôt présente les avantages suivants :

- **Sécurité et Confidentialité des Données** : Les deux sont particulièrement pertinents pour le projet PRESPO en raison de leurs robustes mécanismes de sécurité et de confidentialité. Ces entrepôts sont conçus pour protéger les données sensibles à travers des processus de gestion rigoureux, ce qui est crucial pour garantir la protection des informations des participants et la conformité aux normes éthiques et légales.
- **Gestion des Données Anonymisées et Sensibles** : Nakala offrent une flexibilité dans le traitement et le stockage des données. Il est possible de donner un accès large et contrôlé, ce qui facilite la diffusion des résultats, tout en préservant la confidentialité via des cadres d'accès spécifiques et des contrats de réutilisation, assurant que l'accès est strictement régulé et que les conditions de partage sont respectées. Nakala permettent de rendre les données visibles tout en contrôlant les options de téléchargement.
- **Conformité aux Normes Déontologiques** : Nakala est aligné avec les exigences déontologiques et légales en matière de gestion des données. Il assure que les pratiques de collecte, de stockage et de partage des données respectent les principes éthiques, ce qui est essentiel pour le projet PRESPO. La conformité avec ces normes contribue à maintenir l'intégrité du projet et à protéger les droits des participants.
- **Connexion avec d'autres projets du PPR autonomie** : Nakala est privilégié dans certains projets du PPR Autonomie. Le versement des données dans cet entrepôt pourrait donc permettre une meilleure mise en relation.
- **Accompagnement au versement des données** : Nous pourrions bénéficier d'un accompagnement pour Nakala, avec des ressources humaines dédiées, au versement des données.
- **Conversion des formats fermés en formats ouverts** : Les habitudes de notre communauté de recherche nous imposent d'utiliser des formats fermés au cours du projet afin d'assurer la complétude des fichiers qui seront partagés. Cependant, Nakala permet une conversion automatique des formats et nous anticipons également une étape dédiée à cette question.
- **Interopérabilité avec d'autres outils de gestion des données** : L'environnement Huma-Num propose plusieurs outils qui seront utilisés dans le cadre de PRESPO, notamment Sharedocs et Huma-Num box durant le traitement. Le fait d'utiliser le même environnement assure une bonne cohérence de l'ensemble des données, du début à la fin du projet, tout en évitant de multiplier les outils pour l'équipe.

Les licences que nous apposerons aux données déposées doivent être réfléchies. Plusieurs options s'offrent à nous et doivent être étudiées collectivement, l'utilisation d'une licence de diffusion étant obligatoire. La licence la plus connue est la Creative Commons qui permet de concéder des droits de réutilisation et de reproduction à ceux qui consultent notre ressource. Plusieurs briques peuvent être ajoutées pour personnaliser cette licence : BY (attribution obligatoire), NC (utilisation non commerciale), ND (pas de modification autorisée), et SA (partage sous les mêmes conditions que la licence originale). Ces éléments viennent enrichir les droits d'auteur par défaut. Par ailleurs, certaines licences sont spécifiquement adaptées aux données, comme la CC-BY 4.0, qui impose une attribution, ou la CC-O, qui place les données dans le domaine public, et la CC-SA, qui impose un partage dans les mêmes conditions. La Licence Etalab, également appelée "Licence ouverte", pour être utilisée pour les données publiques françaises financées par des fonds publics.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

A ce stade, il n'y a pas de budget dédié à des modalités d'archivage, c'est-à-dire de mise en disponibilité des données à très long terme.

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Durant le projet, les chercheurs et chercheuses souhaitent utiliser les logiciels et langages suivants (les versions seront précisées dans l'actualisation du PGD) :

- Atlas.ti : logiciel commercial de gestion et d'analyse qualitative de données. Il nécessite l'achat d'une licence pour son utilisation. Les utilisateurs doivent acheter une licence individuelle ou institutionnelle pour accéder à ses fonctionnalités.
- R : langage de programmation et un environnement pour l'analyse statistique qui est entièrement libre et open source. Il est distribué sous la licence GNU General Public License (GPL), ce qui permet aux utilisateurs de télécharger, utiliser, modifier et redistribuer le logiciel gratuitement.
- Stata : logiciel commercial d'analyse statistique et économétrique. Comme Atlas.ti, il nécessite l'achat d'une licence pour son utilisation. Stata propose différentes versions et niveaux de licence pour les utilisateurs individuels et les institutions.
- SAS : logiciel commercial spécialisé dans l'analyse statistique et la gestion de bases de données. L'utilisation de SAS nécessite l'achat d'une licence, et le logiciel est soumis à des restrictions de distribution et d'utilisation.

Au moment du partage des données, nous pourrons garantir l'accessibilité des données du projet PRESPOLE en privilégiant le dépôt des données agrégées sous formats libres et des métadonnées.

Les outils envisagés pour la collecte, le traitement et la sauvegarde des données font l'objet d'une ligne budgétaire, susceptible d'évoluer en fonction des besoins :

- abonnement CASD (config 2, 3 utilisateurs, 5 ans) - 28 000 €
- abonnement logiciel Atlas.Ti (10 utilisateurs, 5 ans) - 9 000,00 €
- matériel informatique - 4 800,00 €

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Nakala attribue un DOI (Digital Object Identifier) aux ensembles de données qu'il héberge pour faciliter leur identification et leur accès pérenne.

Le processus commence par la soumission des enveloppes de données par les chercheurs à Nakala. Cette étape implique le dépôt des données accompagnées d'une documentation complète, incluant des informations détaillées sur le projet, les méthodologies employées, ainsi que les métadonnées associées. La plateforme examine soigneusement ces données pour vérifier qu'elles respectent les normes de qualité et de conformité requises. Une fois la préparation terminée, Nakala procède à la demande d'un DOI via un service d'enregistrement officiel. Ce service est généralement fourni par des agences comme DataCite ou CrossRef, qui sont responsables de l'attribution et de la gestion des DOI. L'agence de DOI attribue un identifiant unique à l'ensemble de données, garantissant une identification pérenne et un accès direct à la ressource.

Après l'attribution du DOI, ce dernier est enregistré dans le système de gestion des données de Nakala. Le DOI est ensuite associé à l'ensemble de données, permettant de créer un lien permanent vers la ressource. Les utilisateurs peuvent utiliser ce DOI pour accéder directement aux données, que ce soit pour des recherches ou des citations dans des publications scientifiques. Nakala prend en charge la maintenance des enregistrements DOI. Cela inclut la mise à jour des informations et des liens associés si nécessaire, afin d'assurer que les ensembles de données restent accessibles même en cas de changement d'URL ou de modification des fichiers. Cette gestion continue garantit que les données conservent leur visibilité et leur accessibilité au fil du temps.

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

Juliette Pinon (Fondation nationale des sciences politiques), project manager de Prespol, est gestionnaire des données, notamment en matière d'homogénéisation des données et métadonnées au niveau du projet. A ce titre, elle fournira plusieurs documents de support, comme les documents de consentement, la trame de nomenclature ou encore la trame de métadonnées. De même, les éventuelles questions de formation et de sensibilisation pourront être organisées par son intermédiaire.

En revanche chaque responsable de tâche reste garant de l'adéquation des modalités de collecte, d'analyse, de vérification et de partage des données avec celles formalisées dans l'accord de consortium et le présent PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Une subvention de 312 000,00 € a été attribuée à l'établissement coordinateur, la Fondation Nationale des Sciences Politiques (FNSP) afin de recruter une gestionnaire de projet, en charge de la rédaction, de la mise en place et de l'actualisation du PGD.

Une subvention de 41 000 € a été prévue pour assurer l'accessibilité des documents de recueil du consentement et de diffusion des résultats. 153 600,00 € sont fléchés sur le contrat d'une ingénieure d'études en charge de ces questions transversales.