

Présentation de la structure et de l'équipe

Présentation de la structure :

L'Institut national d'études démographiques (Ined) est un établissement public à caractère scientifique et technologique (EPST). Il a pour mission de mener des recherches sur les différentes problématiques démographiques (fécondité et famille, santé et mortalité, migrations, etc.) et d'informer le public et les chercheurs des évolutions en cours. L'Ined compte environ 240 agents (agents titulaires, doctorants, agents contractuels, etc.) et accueille de nombreux chercheurs associés ou invités. L'Ined est au cœur d'un réseau de partenariats nationaux et internationaux de recherche.

Contexte professionnel :

Le service des enquêtes et sondages (SES) est un des services d'appui à la recherche de l'Ined. Ses missions centrales s'articulent autour des activités de collecte de données, quantitatives comme qualitatives, depuis la conception de questionnaires et des plans de sondage jusqu'à la mise en forme finale des fichiers de données. Le service mène également une réflexion méthodologique et une veille sur les protocoles et les techniques innovantes en matière d'enquêtes en sciences sociales. Enfin, il s'occupe de la mise à disposition des enquêtes de l'INED au sein du Réseau Quetelet et de la production de bases de données démographiques agrégées ou contextuelles, au sein de son « Datalab ». Le SES comprend quatre statisticien·ne·s d'enquête, organisés au sein d'un pôle « statistique » du service, dans lequel s'inscrira partiellement ce recrutement.

Le projet LifeObs est un Grand équipement structurant pour la recherche (Equipex+) financé par le Programme d'investissements d'avenir (PIA3), <https://lifeobs.site.ined.fr>. L'enquête Étude des relations familiales et intergénérationnelles (Erfi) en fait partie : elle est la version française de l'enquête européenne menée dans le cadre du Generations and Gender Programme, une infrastructure européenne d'enquêtes longitudinales comparatives sur les comportements familiaux des hommes et des femmes : fécondité, vie de couple, solidarités familiales, conciliation entre vie professionnelle et vie privée (<https://www.ggp-i.org>). Le premier round d'enquêtes a eu lieu dans les années 2000, l'enquête française ayant eu lieu en trois vagues, 2005-2008-2011, <https://erfi.site.ined.fr>. Un deuxième round a été lancé pour les années 2020. L'Ined dirige le volet français du projet ; une enquête « Pilote » a eu lieu en 2022 et la première vague de l'enquête Erfi2 aura lieu d'octobre 2023 à juillet 2024 ; les mêmes personnes seront ensuite recontactées en 2027 et 2030. Ce projet est conduit au sein de l'unité de recherche Familles, fécondité, conjugalité de l'Ined, où la personne recrutée sera intégrée, le travail se faisant au sein de l'équipe du pôle « statistique » du SES.

L'équipe-projet Comète du centre Inria de Saclay est spécialisée dans la sécurité et la protection de la vie privée, elle compte seize chercheurs (<https://www.inria.fr/fr/comete>). L'équipe est le créateur de la d-privacy (une généralisation de la differential privacy), qui est paramétrée par une métrique de distance d . En particulier, la d-privacy a été instanciée à la confidentialité de l'emplacement, où elle est connue sous le nom de geo-indistinguishability, et elle est largement acceptée comme l'état de l'art par la communauté. L'équipe Comète possède également une forte expertise dans la local differential privacy (un modèle dans lequel la differential privacy est appliquée au niveau du

données individuelles), en particulier avec les données catégorielles, ainsi que dans les menaces de confidentialité pour les systèmes d'apprentissage automatique (mesure des fuites de confidentialité et mécanismes de défense). A l'échelle internationale, Comète fait partie du consortium ELSA, qui est un centre d'excellence virtuel regroupant 26 institutions européennes (ex. CISPA, NVIDIA Suisse, EPFL). Par ailleurs, la responsable de l'équipe Comète (Catuscia Palamidessi) dirige le projet ERC (European Research Council) HYPATIA, qui s'inscrit dans le cadre du programme de recherche et d'innovation Horizon 2020 de l'Union européenne. Un autre projet international est CRYPTTECS, qui est un projet MESRI/BMBF dont le consortium comprend diverses entreprises et institutions académiques françaises et allemandes. Au niveau national, Comète participe au challenge Fed-Malin Inria et au volet IPoP (Projet Interdisciplinaire sur la Vie Privée) du PEPR Cybersécurité. De 2016 à 2021, l'équipe Comète a été responsable du projet ANR REPAS, qui portait également sur la protection de la vie privée et dont le consortium comprenait diverses institutions académiques françaises.

Missions du poste

La personne recrutée aura pour mission principale de proposer une méthode pour la constitution d'un fichier anonyme à partir d'une enquête par questionnaire en sciences sociales, et de la mettre en application. Dans un premier temps, un fichier d'apprentissage de l'enquête Erfi sera constitué, permettant une mise à disposition sur internet, en open access, d'un fichier répondant au cahier des charges de la formation aux utilisateurs réalisée dans le cadre de LifeObs. Ce travail nécessitera une revue bibliographique des méthodes dans le domaine, la comparaison de plusieurs de ces méthodes et donnera lieu à la publication d'un ou plusieurs documents scientifiques sur la méthode retenue. Le programme pourra ensuite être développé à d'autres enquêtes de l'Ined pour lesquelles une formation est envisagée, à partir d'un fichier anonyme de l'enquête.

Par ailleurs, l'Ined diffuse des fichiers pseudonymisés, disponibles via le réseau Quetelet- Progedo-diffusion (<https://data.progedo.fr>) et, pour les enquêtes GGP, via une application spécifique (<https://www.ggp-i.org/data>). Une analyse des risques liés à l'anonymisation incomplète de certains fichiers et une analyse des règles de pseudonymisation utilisées pourront conduire à l'évaluation des risques, à la rédaction d'un guide de bonnes pratiques, en relation avec les autres producteurs de données pseudonymisées (Insee pour la France, diffuseurs d'enquêtes européennes), et éventuellement à la constitution de fichiers « mieux pseudonymisés ».

Les relations avec des organismes partenaires ainsi que la participation à des colloques peuvent conduire à des déplacements ponctuels en France et à l'étranger.

Profil recherché

Diplôme souhaité : doctorat (formation en Statistiques ou en Informatique ou en Sciences sociales ayant une forte composante mathématique/statistique/informatique). Une formation aux techniques d'anonymisation serait appréciée.

Le poste est ouvert à des candidat-e-s nouvellement diplômé-es. Une formation complémentaire sur certaines compétences pourra être proposée en début de poste si nécessaire.

Le travail sera réalisé en français ou en anglais. Une bonne connaissance d'une de ces deux langues est nécessaire.

Des expériences de travail dans des instituts nationaux statistiques ou auprès d'équipes de recherche non françaises, en particulier dans le cadre d'enquêtes sociales, seraient appréciées.

Compétences attendues :

- Très bon niveau en anglais écrit et oral (C1 idéalement, B2 minimum) ; connaissance du français appréciée ;
- Connaissances en méthodologie d'anonymisation des données ;
- Connaissances des méthodes d'analyse du risque de ré-identification et des risques associés d'inférence, connaissance des mesures de l'utilité des données pour les analyses statistiques ;

- Connaissance pratiques des langages d'analyse de données (R, python, C++) et des logiciels de gestion de base de données (MySQL...);
- Connaissance du RGDP et du contexte légal ;
- Aptitude à la rédaction à visée pédagogique et à la documentation ;
- Aptitude à la veille méthodologique ;
- Rigueur, bonne organisation, sens des priorités, autonomie ;
- Capacités de travail en équipe.

Informations générales

DATE DE PRISE DE FONCTION	Juillet/Août 2023
DURÉE (CDD OU DÉTACHEMENT)	2 ans
TEMPS DE TRAVAIL :	Temps plein 37h55 par semaine 32 jours de congé annuel et 10 JRTT
ACTIVITÉS TÉLÉTRAVAILLABLES	Oui (2 jours maximum par semaine)
REMUNERATION	Selon la grille de rémunération interne et expérience (minimum 2829€ brut mensuel)
LOCALISATION	Ined (50%) Campus Condorcet 9 cours des Humanités à Aubervilliers Métro ligne 12 : Front populaire Et Centre Inria de Saclay/Est (50%) Campus de l'École Polytechnique Bâtiment Alan Turing 1 rue Honoré d'Estienne d'Orves 91120 Palaiseau

Modalités de la candidature

DATE LIMITE DE CANDIDATURE	25 juin 2023
MODALITÉS DE CANDIDATURE	Envoyer à l'attention de Thomas Merly-Alpa, chef du service des enquêtes et sondages : - CV - lettre de motivation en anglais (max 2 pages) - Nom et coordonnées de 2 personnes de référence Sous la référence « Chargé-e d'études anonymisation données » A : recrutements@ined.fr
PROCESSUS DE RECRUTEMENT	Les candidat-es seront informé-es de leur pré-sélection au plus tard le 30 juin 2023 et seront éventuellement auditionné-es en juillet, pour un recrutement en juillet/août 2023.