

## 2019-01614 - Doctorant F/H HistorIA : Grandes bases de données historiques Fouille de données, exploration et explicabilité

Type de contrat : CDD de la fonction publique  
Niveau de diplôme exigé : Bac + 5 ou équivalent  
Fonction : Doctorant

### A propos du centre ou de la direction fonctionnelle

Le centre de recherche Inria Saclay – Île-de-France, créé en 2008, accueille 450 scientifiques et 60 membres des services d'appui à la recherche. Les scientifiques sont organisés en 28 équipes de recherche dont 23 sont communes avec des partenaires du plateau de Saclay.

Le centre Inria Saclay - Île-de-France est un acteur essentiel de la recherche en sciences du numérique sur le plateau de Saclay. Il porte les valeurs et les projets qui font l'originalité d'Inria dans le paysage de la recherche : l'excellence scientifique, le transfert technologique, les partenariats pluridisciplinaires avec des établissements aux compétences complémentaires aux nôtres, afin de maximiser l'impact scientifique, économique et sociétal d'Inria.

### Contexte et atouts du poste

L'objectif de la thèse est de concevoir des indicateurs et des interfaces visuelles à l'usage d'historien·nes (et plus généralement de chercheuses et chercheurs en sciences sociales) pour l'exploration de données massives issues de sources historiques, avec une forte composante d'analyse de réseaux. La thèse, inscrite en informatique, se fera en collaboration avec des historien·es et sociologues.

### Motivations

Depuis le développement des méthodes big data et leur arrivée dans les sciences sociales, plusieurs initiatives très ambitieuses ont vu le jour se fixant pour objectif de changer la manière de faire de la recherche en histoire. Pourtant, le déploiement de ces nouvelles approches se heurte à de nombreuses réticences des historien·nes, qui craignent parfois d'être dépossédés de leur matériau de recherche lorsqu'il est transformé pour être intégré dans des bases de données, et qui sont souvent sceptiques sur la pertinence d'indicateurs parfois abscons, basés sur des données en trop grand nombre pour être appréhendées à la main.

L'un des objectifs de la thèse est donc d'accorder une grande importance à l'explicabilité des indicateurs mis au point, en établissant un protocole permettant des allers-retours entre production d'éléments d'analyse et exploration interactive. Les difficultés d'interprétation donneront ainsi lieu à des corrections des algorithmes au fur et à mesure et l'ajout d'éléments de visualisation qui leur donnent sens.

### Méthodes

L'analyse des réseaux pourra se faire en s'appuyant sur deux approches : l'énumération de structures élémentaires (appelées schémas, motifs ou graphlets), ou l'identification de groupes denses (clusters ou communautés). Les outils visuels d'exploration pourront s'appuyer sur le « Vistorian », une plateforme web à destination des historien·nes pour la visualisation de réseaux temporels. <https://vistorian.net/>

Pour améliorer la réactivité des interfaces, les algorithmes pourront utiliser le principe d'analyse progressive, paradigme de calcul original, imposant un impératif de très faible temps de latence pour produire des résultats de calcul de manière progressive, commençant très rapidement par une approximation grossière et améliorant la qualité à intervalle régulier au cours temps.

### Mission confiée

#### Missions :

Avec l'aide de Jean-Daniel Fekete, Christophe Prieu, et les chercheurs en sciences sociales, la personne recrutée sera amenée à concevoir des algorithmes de clustering et de calcul de motifs dans des réseaux sociaux qui soient compréhensibles et contrôlables par des chercheurs en sciences sociales sans expertise avancée en informatique.

### Collaborations

- Inria, projet Aviz (interaction et visualisation, data science)
  - Responsable: Jean-Daniel Fekete
- Télécom ParisTech, dept SES (data science, sciences éco et sociales)
  - Responsable: Christophe Prieur
- Université Paris-Diderot, LIED (histoire économique)
  - Responsable: Zacarias Moutoukias
- EHESS, LaDéHIS (histoire sociale)
  - Responsable: Pascal Cristofoli

### Principales activités

Principales activités (5 maximum) : aider à la saisie et la structuration de corpus historiques et sociologiques, concevoir des algorithmes de clustering, concevoir des algorithmes de calcul de motifs

### Informations générales

- **Thème/Domaine :** Interaction et visualisation  
Production, traitement et analyse des données (BAP D)
- **Ville :** GIF-SUR-YVETTE
- **Centre Inria :** CRI Saclay - Île-de-France
- **Date de prise de fonction souhaitée :** 2019-05-01
- **Durée de contrat :** 3 ans
- **Date limite pour postuler :** 2019-09-02

### Contacts

- **Equipe Inria :** AVIZ
- **Directeur de thèse :**  
Fekete Jean-daniel / [jean-daniel.fekete@inria.fr](mailto:jean-daniel.fekete@inria.fr)

### A propos d'Inria

Inria, l'institut national de recherche dédié aux sciences du numérique, promeut l'excellence scientifique et le transfert pour avoir le plus grand impact. Il emploie 2400 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3000 scientifiques pour relever les défis des sciences informatiques et mathématiques, souvent à l'interface d'autres disciplines. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 160 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

### L'essentiel pour réussir

L'étudiant devra avoir une formation en théorie des graphes, et si possible en interaction homme-machine ou visualisation d'information. Une bonne maîtrise de la programmation est souhaitée, idéalement en Python et ses bibliothèques scientifiques (Pandas, NumPy, SciPy) et de manipulation de réseaux (NetworkX). Une bonne connaissance des bibliothèques Web pour la visualisation et l'interaction est aussi souhaitée.

### Consignes pour postuler

#### Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

#### Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

**Attention :** Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

dans des graphes, visualiser le résultat de ces clustering et motifs.

#### Exemples d'activités :

- Analyser les besoins des partenaires
- Proposer des solutions de clustering et de calcul de motifs pour des grands réseaux sociaux
- Développer des interfaces de visualisation et d'exploration pour ces réseaux.
- Former à l'utilisation les principaux partenaires
- Animer une communauté d'utilisateurs
- Présenter l'avancée des travaux aux partenaires

## Compétences

Compétences techniques et niveau requis : Masters 2 ou équivalent (ingénieur)

Langues : Français, Anglais, éventuellement Espagnol et Roumain

Compétences relationnelles : aimer collaborer avec des historiens et des sociologues

Compétences additionnelles appréciées : être intéressé par l'histoire sociale.

## Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

## Rémunération

Salaire mensuel (1ère et 2ème année) : 2.085 euros, (3ème année) : 2.085 euros