

Proposition de stage

ANR Time-Us

1. Présentation rapide de l'institution d'accueil, du contexte institutionnel ou entrepreneurial

L'ANR Time-Us a pour but de reconstruire la rémunération et les budgets-temps des femmes et des hommes travaillant dans le textile en France (plus précisément dans les régions de Lille, Paris, Lyon et Marseille) durant la révolution industrielle (sur une période allant de la fin du XVIIIème siècle à la Première Guerre mondiale). Le programme rassemble des historiens, mais aussi des sociologues et des spécialistes en traitement automatique des langues. Coordonné par Manuela Martini, professeure d'histoire contemporaine à l'université de Lyon 2, il réunit 5 équipes dont le laboratoire ICT (« Identités-Cultures-Territoires ») de Université Paris 7 Diderot qui va accueillir le stagiaire de M2 sous la direction de Liliane Hilaire Pérez, professeure d'histoire moderne.

L'Équipe d'accueil ICT (« Identités-Cultures-Territoires ») est un **laboratoire pluridisciplinaire qui regroupe des historiens (UFR GHES) de différentes périodes historiques (du Moyen Âge jusqu'à la période très contemporaine) et des civilisationnistes (UFR EILA) de différentes aires culturelles (angliciste, germanique et hispanique)**. Son projet scientifique a été construit autour d'une volonté d'ouverture pluridisciplinaire pour développer, grâce à ce dialogue, une approche critique de l'histoire globale, comportant une réflexion sur le monde comme un ensemble d'aires culturelles et historiques. De nombreux partenariats avec l'étranger (conventions, projets internationaux) viennent appuyer cette démarche.

Le laboratoire ICT a particulièrement développé, dans le souci des **problématiques de genre, des approches attentives aux différentes formes de mobilité, aux modalités de circulation des femmes, des hommes et des savoirs au cours de l'histoire, aux dynamiques territoriales associées à ces circulations et aux enjeux socio-économiques, culturels, politiques et épistémologiques des mises en relation que la mobilité permet. Dans ce cadre, le programme ANR Time-Us s'inscrit dans les thématiques développées dans l'Axe 1 Territoire, mobilités et pouvoirs et dans l'Axe 3 Genre.**

Le/la stagiaire sera également amené(e) à collaborer de façon étroite avec l'équipe de traitement automatique des langues d'Inria¹, dont certains membres participent activement au projet, sur les aspects informatiques et numériques du stage. Seront impliqués côté Inria des spécialistes en traitement automatique des langues et en humanités numériques, y

¹ Inria (Institut national de recherche en informatique et en automatique) est un établissement public à caractère scientifique et technologique placé sous la tutelle des ministères de la recherche et de l'industrie. Adresse du centre de Paris : 2 rue Simone Iff, 75012 Paris.

compris des historiens. Cette collaboration se placera dans le contexte de la participation de l'équipe au projet européen Parthenos, bénéficiaire du programme Horizon 2020 de l'Union Européenne. Ce projet vise à renforcer la cohésion de la recherche dans le large secteur des sciences humaines en coordonnant l'activité des infrastructures de recherches européennes, notamment par un travail d'harmonisation des pratiques de recherche, en particulier dans l'utilisation de formats standards pour les données de la recherches, le développement de services mutualisés, etc. Le projet Time-Us pourra ainsi mettre en pratique les résultats obtenus dans le cadre de Parthenos, et en deviendra ainsi un cas d'utilisation. Le rôle du/ de la stagiaire sera à cet égard déterminant.

2. *Description (à la fois synthétique et suffisamment précise pour cerner les enjeux du projet et la mission qui serait dévolue au/à la stagiaire) du stage et de la réalisation sur laquelle il devra déboucher, dont*
 - a. *présentation du ou des livrables techniques envisagés (formats, langages, logiciels impliqués, ...)* ;
 - b. *présentation du contexte et des enjeux scientifiques et/ou professionnels du projet* ;

Objet du stage: **Mise en place d'une chaîne de traitement des documents pour l'ANR Time-Us et déploiement d'un outil de transcription collaborative**

L'un des aspects du projet concerne la gestion de corpus historiques, de la transcription à l'annotation. Le/la stagiaire aura à mettre en place une chaîne de traitement des documents utilisant un logiciel comme Transkribus (qui permettrait la mise en place d'une transcription collaborative des sources) et permettant la conversion vers TEI Boilerplate, Nakalona...ou autre (à définir par le/la stagiaire). Il s'agira de travailler sur un schéma TEI pour l'ensemble du corpus, et à une instanciation de ce schéma pour encoder les monographies de Le Play (<https://rh19.revues.org/334>). Ces monographies ont été acquises par le projet et vont être numérisées.

Les missions du/ de la stagiaire seront donc les suivantes:

- Mise en place d'un guide de bonnes pratiques en vue de la constitution et de l'analyse du corpus :
 - création de métadonnées,
 - annotations à envisager pour le traitement linguistique,
 - repérage des informations pertinentes pour l'analyse historique dans le contexte du projet;
- Développement d'un schéma TEI destiné à l'encodage et à l'analyse du corpus, et d'une instanciation de ce schéma pour l'encodage de certains éléments du corpus (notamment les Monographies de famille de l'École de Le Play);
- Production d'un livrable comparant les différents outils de transcription collaborative disponibles, justifiant le choix de l'outil déployé, présentant les modalités de l'implémentation de l'outil choisi et décrivant les perspectives de développement envisagées;
- Déploiement d'un outil de transcription collaborative qui permettra l'intégration de modules/plugins destinés au traitement automatique des langues, et l'utilisation d'un schéma TEI spécifique au corpus étudié.

Le/la stagiaire travaillera en étroite collaboration avec des historiens et des spécialistes du traitement automatique des langues. Il/elle travaillera également avec des élèves du Master "Archivistique et Monde du Travail" de l'Université Lille 3, qui utiliseront l'outil de transcription collaborative mis en place par le/la stagiaire.

3. *Nom et coordonnées complètes de la ou des personnes contact (auprès de qui prendre des informations supplémentaires, et qui encadrera(en)t directement le/la stagiaire sur place), et d'éventuels référents informatiques sur place (s'ils diffèrent du maître de stage) ;*

À ICT le/la stagiaire sera supervisé(e) par Liliane Hilaire Pérez, professeure d'histoire moderne à l'université Paris Diderot Paris 7, membre titulaire du laboratoire ICT et par Manuela Martini, professeure d'histoire contemporaine à l'université Lyon 2 et membre associé au laboratoire ICT².

À Inria, le/la stagiaire sera supervisé(e) par Benoît Sagot. Il/elle sera également en interaction régulière avec d'une part Marie Puren et Charles Riondet pour les aspects liés aux humanités numériques et au projet Parthenos, et d'autre part Éric de La Clergerie pour les aspects liés au traitement automatique des langues³.

4. *Modalités matérielles d'accueil (possibilités d'hébergement, de restauration, support éventuel des frais de déplacement, ...), durée prévue du stage et montant de la gratification*

La durée du stage est de 4 mois (avril-juillet 2017), le montant mensuel de la gratification prévue est de 560 euros.

² Adresses e-mail: manuela.martini3@gmail.com; liliane.perez@wanadoo.fr

³ Adresses e-mail: marie.puren@inria.fr ; charles.riondet@inria.fr ; eric.de_la_clergerie@inria.fr ; benoit.sagot@inria.fr.