**Appel A Projets Recherche 2020**

**Fédération de Recherche Agorantic**

**«Culture, Patrimoines, Sociétés Numériques »**

|  |  |
| --- | --- |
| **Titre** | **Les élus français dans Wikipédia** |
| **Acronyme** | **WikiRNE** |
| **Nom du/des porteur(s)** | Guillaume MARREL et Pierre JOURLIN |
| **Laboratoires associés** | LBNC : Guillaume Marrel  LIA : Pierre Jourlin, Eric Sanjuan |
| **Budget demandé** | 7 381 € |
| **Résumé**  **Max. 1 000 caractères espaces compris** | WikiRNE associe des chercheurs en sociologie politique et en informatique afin de repérer, décrire et analyser la présence des élus français des assemblées politiques locales, nationales et européenne dans l’encyclopédie numérique contributive *Wikipédia* depuis 2001. Il ‘agit de croiser une version historique du Répertoire national des élus (RNE) compilant depuis 2001, au Ministère de l'Intérieur, l’identité nominative de l’ensemble des représentants du suffrage universel, avec la base *dbpedia* où sont stockées les données actuelles et archivées de l’ensemble des notices de l’encyclopédie. En science politique, le projet vise à objectiver les pratiques éventuellement renouvelées d’écriture biographique ou autobiographique et leurs relations avec la vie démocratique et les contraintes de communication et de notoriété des dynamiques électorales, dans le cadre des sociétés numériques. En informatique, cette question est l’occasion de mobiliser de nouveaux outils du TALN pour l’extraction et le croisement d’informations structurées et non structurées. Le projet explore enfin, de manière plus générale, les usages de *Wikipédia* et ses potentielles exploitations croisant sciences sociales et sciences informatiques. |

|  |
| --- |
| 1. **Contexte, positionnement, objectif(s)/problématique et cohérence par rapport à la thématique « Culture, Patrimoines, Sociétés Numériques »**   Autour des pages politiques de l’encyclopédie numérique contributive *Wikipedia*, le projet WikiRNE rassemble des enjeux de recherche et des chercheurs de sociologie politique et d’informatique dans le cadre de l’axe Culture, Patrimoines, Sociétés numériques.  Du point de vue de la science politique, il s’agit de reconnaître que Wikipédia est devenue non seulement la première base de connaissances des moteurs de recherche aujourd'hui, mais aussi l’une des principales bases biographiques de référence des personnalités publiques considérées à un moment ou à un autre, dans un espace donné, comme “remarquables” et “méritant” une notice individuelle. Bénéficiant par définition d'une certaine notoriété, le personnel politique en général y occupe par hypothèse une place singulière. Initialement réservée aux “grands hommes”, la biographie s’est “démocratisée” à la fin de l’Ancien régime et au début du XIXe siècle, avec la multiplication des supports imprimés et l’essor des dictionnaires biographiques territoriaux, sectoriels, thématiques.... Wikipédia libère sans doute ce type de récit des contraintes éditoriales et des filtres académiques élargissant ainsi le spectre des personnalités susceptibles d’accéder à la postérité biographique.  Le “dictionnaire biographique” est en tant que tel, dans son “format papier”, un projet éditorial de nature politique dans les démocraties naissantes : il participe de l’identification et de la construction sociale des nouvelles élites. La publication d’une notice biographique est à la fois une reconnaissance et un facteur de notoriété. Dans la démocratie représentative contemporaine, l’accès aux positions de pouvoir politique reconnues dans les institutions publiques repose pour la plupart d’entre elles sur l’élection au suffrage universel, qui sanctionne une certaine popularité. Faire campagne, être élu et réélu, c’est se faire remarquer, connaître et reconnaître. L’apparition d’un récit biographique personnel signifie quelque chose pour une personnalité qui participe à la vie publique d’un territoire. Critique ou hagiographique, autobiographique ou confié à un proche, le texte biographique et ses réécritures révèlent différents états de cette notoriété. Comment la biographie accompagne-t-elle alors la carrière politique ?  Le format numérique de Wikipédia autorise une analyse systématique des formes et transformations de la présence nominative des élus de France dans les notices de l’encyclopédie. Pour exploiter l’immensité de son contenu, nous disposons du RNE (Répertoire National des Élus) comme base nominative de référence : la totalité des candidats élus ou nommés sur le territoire national dans toutes les assemblées municipales, intercommunales, départementales, régionales, nationales et européenne, y sont recensés et caractérisés succinctement (nom, prénom, date de naissance, profession déclarée, nuance politique), soit environ un million de personnes réparties sur plus de 500000 mandats, de 2001 à nos jours. Il s’agit donc de croiser cette base nominative avec Wikipédia.  La synchronisation historique est par chance presque parfaite, puisque Wikipédia existe officiellement depuis le 15 janvier 2001. La plateforme archive par ailleurs l’intégralité des enregistrements réalisés sur ses notices et autorise donc un suivi généalogique des controverses et des modifications parfois nombreuses auxquelles l’écriture biographique donne lieu. L’écriture biographique est de fait une pratique sensible pour cette encyclopédie collaborative auto-régulée. Elle fait l’objet de plusieurs recommandations générales relatives aux [conventions de style](https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions_de_style#Biographies), harmonisées dans le [Projet biographie](https://fr.wikipedia.org/wiki/Projet:Biographie) de Wikipédia, mais aussi aux enjeux spécifiques des [biographies de personnes vivantes](https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Biographie_de_personne_vivante) et de [l'autobiographie](https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Autobiographie). |

|  |
| --- |
| 1. **Méthodologie et résultats attendus**   Il s’agit de croiser la base nominative RNE avec Wikipédia pour extraire, lorsqu’elles sont disponibles, toutes les informations biographiques concernant les élus. Un.e élu.e peut disposer d’une notice biographique personnelle. Son peut être mentionné dans le cadre d’une autre notice biographique, thématique ou institutionnelle. Dans certains cas, ces informations sont déjà structurées sous la forme d’infoboites. Certaines sont même déjà extraites sous forme de triplets RDF (Resource Description Framework, voir Web Sémantique et dbpedia), donc facilement exploitables sous forme relationnelle.  Le croisement des données issues de dbpedia, n’est pourtant pas trivial. Par exemple, certains maires de communes de petite taille peuvent être mentionnés dans l’infoboite relative à la commune sans disposer pour autant d’une page dédiée à leur personne. Le RNE contient les informations indispensables à la jointure (nom, prénom, date de naissance), mais pas forcément dbpedia (par ex : pas de date de naissance). Or, compte-tenu de la taille du RNE et de celle du Wikipédia, il faut s’attendre à ce que chaque personne puisse avoir plusieurs homonymes répertoriés. Il faudra donc imaginer des techniques relativement sophistiquées, probabilistes ou à base de contraintes pour réaliser ces croisements. (ex : Michel Gentil, maire de Bey, dans le département de l’Ain depuis 2014, peut-il être le Michel Gentil qui possède une notice Wikipedia, député du Loiret de 1791 à 1799 ?).  Dans d’autres cas, les informations ne sont pas du tout structurées. Il faudra alors avoir recours à des techniques de traitement automatique du langage naturel. Des outils d’annotation syntaxique automatique et d’extraction d’entités nommées sont utilisés au LIA et pourront être mis à profit pour réaliser cette tâche.  Ce projet a pour but de faciliter le travail des chercheur.e.s en sociologie politique et de leur permettre de répondre à un certain nombre de questions sur les usages de Wikipédia en termes de communication politique, dont voici une liste non exhaustive :   * Combien d’élus du RNE disposent d’une notice biographique Wikipédia ? * Qui sont-ils ? Quels sont leurs propriétés sociales et politiques ? Peut-on repérer des variables socio-politiques statistiquement significatives dans la distribution de cette population ? * Qu’étaient-ils au moment de la création de leurs notices ? * Leur notice a-t-elle évolué et comment cette évolution est-elle liée ou non aux événements de la carrière de l’individu ? * Ces élus qui disposent d’une notice sont-ils cités dans d’autres notices biographiques ou non biographiques ? Si oui lesquelles ? A quelle fréquence ? Selon quelle “tonalité” ? * Les élus qui ne disposent pas de notice biographique sont-ils cités dans d’autres notices biographiques ou non biographiques ? Si oui lesquelles ? A quelle fréquence ? * Comment représenter et cartographier ces informations ?   Le projet contribue également à mieux connaître les usages de Wikipédia. Le formidable développement de cette encyclopédie fait l’objet depuis le début des années 2010 d’un réel intérêt scientifique en sciences de l’information principalement [(Barbe, Merzeau, et Schafer 2015)](https://www.zotero.org/google-docs/?6pZN5i). La question de l'exploitation de données biographiques à partir de Wikipédia n'est pas nouvelle [(Viseur 2015)](https://www.zotero.org/google-docs/?QYKdBi).   * Comment les corpus de biographies politiques se développent-ils dans la courte histoire de Wikipédia ? * Quand Wikipédia devient-il un espace de publication de biographies de personnalités politiques vivantes pertinent et recherché ? Par qui et pourquoi ? * Qui sont les auteurs et les correcteurs des notices ? * A quels types de controverses l’écriture biographique politique donne-t-elle lieu ? * Peut-on caractériser et décrire ces controverses ? * Que nous apprennent ces wiki-controverses sur l’écriture biographique politique aujourd’hui ?   Le projet vise une double valorisation, en termes de communications et de publications dans les conférences et les revues disciplinaires et interdisciplinaires concernées, et en terme de diffusion scientifique lors des événements grand-publics. |

|  |
| --- |
| 1. **Dimension interdisciplinaire (champs disciplinaires associés)**   Explorée initialement en préparation du DataPolHack du projet Goow du 28 février 2019, la présence nominative dans l’encyclopédie Wikipédia des élu.e.s politiques français.e.s, recensé.e.s dans le Répertoire National des Élus (RNE), constitue en tant que telle une question de recherche en informatique et en science politique. La dimension interdisciplinaire est liée à l’opportunité que constitue le changement d’échelle, de support et de format que constitue la création de Wikipedia, comme base d’information notamment biographique sur les personnalités dignes d’une certaine publicité. La nature partiellement structurée de l’information biographique et sa quantité autorisent ici, en sociologie politique, des mesures et des traitements automatisés de masse, jusqu’ici relativement confidentiels, sur certaines questions relatives à la construction et à l’entretien de la notoriété des acteurs de l'espace public de la démocratie représentative. |

|  |
| --- |
| 1. **Partenariats extérieurs envisagés**   Le dispositif exploratoire qu’il s’agit ici de financer entre le LBNC et le LIA à vocation en cas de réussite à mobiliser d’autres acteurs académiques et non-académiques. La recherche pourrait donner lieu à la présentation d’un sujet de thèse dans le cadre de l’AAP financement doctoral Agorantic. |

|  |  |  |
| --- | --- | --- |
| **Budget (€)\*** | | |
|  | **Brève description** | **Montant** |
| **Stages\*\*** | M2 informatique 6 mois | 3 480 |
| **Post-doc** | IGE informatique 1.5 mois ETP (INM388) | 3 901 |
| **Budget total** |  | 7 381 |
| **Budget demandé à Agorantic** |  | **7 381** |

\* Veuillez modifier les catégories de dépenses si besoin – ajoutez/supprimez des lignes à votre convenance

\*\*Gratification de stage obligatoire au-delà de 2 mois – prévoir environ 580€ par mois