

## APPEL À PROJETS 2023 – DONNEES & OPEN DATA

FEDERATION DE RECHERCHE AGORANTIC  
«CULTURE, PATRIMOINES, SOCIETES NUMERIQUES »

### Attention :

- Ne pas dépasser 5 pages
- Transmettre le fichier au format PDF intitulé : ACRONYME-AAP-Données-Agorantic-2023.pdf
- Envoyer le fichier à [agorantic@univ-avignon.fr](mailto:agorantic@univ-avignon.fr) avant le 3 janvier 2023.

Titre	Beneficial Ownership Register Interconnection Scheme
Acronyme	BORIS
Nom du/des porteur(s)	Vincent Labatut – Pierre-Henri Morand
Laboratoires associés	LIA - LBNC
Budget demandé	5 300
Résumé <i>Max. 1 000 caractères espaces compris</i>	Le projet BORIS vise à récupérer, structurer et exploiter des données du registre français des bénéficiaires effectifs par croisement avec les bases BRÉF et FOPPA déjà portées par des membres de la FR Agorantic. Dans un contexte juridique qui fait craindre la disparition à court terme de ces données au format ouvert, BORIS vise également à proposer une méthode permettant d'objectiver les coûts de la censure d'une partie des informations contenues dans ces registres afin d'alimenter le débat juridique naissant entre les contraintes de transparence des données publiques et de respect des données personnelles.

### 1. Contexte, positionnement, objectif(s)

Le registre des bénéficiaires effectifs (Beneficial Ownership Register) est un fichier dans lequel sont conservées et mises à disposition des informations sur les bénéficiaires effectifs des entités immatriculées au Registre de Commerce et des Sociétés. Un « bénéficiaire effectif » ou « ultimate beneficial owner », désigne toute personne physique qui, en dernier ressort, possède ou contrôle une structure. Dans le cas des sociétés, une personne est considérée comme bénéficiaire effectif si elle détient, directement ou indirectement, plus de 25 % du capital et/ou des droits de vote.

Ce registre offre un important potentiel de valorisation et d'enrichissement des bases de données constituées actuellement au sein de la FR Agorantic. En particulier, en croisant la BRÉF (Base révisée des élu-es de France) avec le RBE ([Registre des Bénéficiaires Effectifs](#)), un travail d'identification des structures détenues ou contrôlées par les élu-es est rendu possible. De manière similaire, en croisant la base FOPPA (French Open Public Procurement Attribution notices) avec le RBE, c'est l'identification de liens de propriétés entre structures juridiques distinctes qui devient réalisable. Enfin, en croisant BRÉF, FOPPA et RBE, ce sont des liens entre structures adjudicatrices, entreprises titulaires des marchés et dirigeants (des structures publiques et des structures privées) qui peuvent être objectivés.

Une récente décision de la Cours de Justice de l'Union Européenne (Arrêt CJUE C-37/20 et

C-601/20 du 22 novembre 2022) vient de porter un coup d'arrêt à l'ouverture et au maintien, en Open Data, de ce registre. Plusieurs pays ont d'ores et déjà limité l'accessibilité de ces données suite à cette décision ; il y a donc une réelle **urgence scientifique** à réaliser, dès le semestre qui s'ouvre, l'ensemble des investigations et croisement de bases évoqués tant que les interfaces françaises demeurent fonctionnelles.

Par ailleurs, dans le cadre du **programme PAUSE**, le programme de recherche financé par l'ANR DeCoMaP et le LBNC accueilleront pour un semestre, et à compter du 3 janvier 2023, notre collègue juriste ukrainienne Madame IASECHKO. Elle aura ainsi vocation, durant son semestre au sein de l'établissement à contribuer à la tâche 3.2 de l'ANR DeCoMap : *RGPD, données ouvertes et prédiction automatique* en questionnant notamment les tensions contradictoires entre le régime de protection des données personnelles (tel que porté par le RGPD en Europe) et les objectifs de transparences et d'accessibilité portés par les projets d'Open Data. La décision de la CJUE évoquée est l'illustration parfaite de ces fortes tensions. Le travail de croisement de différentes bases de données ouvertes porté par le projet BORIS devient un terrain particulièrement pertinent d'analyse juridique de cette problématique émergente.

Le projet BORIS mobilisera :

Vincent Labatut, MCF HDR en informatique, LIA, Avignon Université

Pierre-Henri Morand, PR en économie, LBNC, Avignon Université

Lucas Potin, doctorant contractuel en informatique, LIA, Avignon Université

Téa Toutounji, doctorante contractuelle en économie, CRED, Université Paris II – Assas

## 2. Questionnement scientifique en rapport avec l'intitulé de l'appel

Le projet BORIS s'inscrit dans un double questionnement scientifique en lien avec l'appel à projet Données & Open Data.

Le premier tend à exploiter le potentiel offert par la gestion de deux bases de données ouvertes d'intérêt national au sein d'Avignon Université en explorant les possibilités d'interconnexion de ces dernières. Il s'agirait de la première tentative de faire vivre cet écosystème de données publiques en les interconnectant via une base externe, et en offrant ainsi un potentiel d'analyse et de recherche absent par nature des bases d'origine. La base FOPPA, constituée à partir des données institutionnelles de la TED (Tender Electronic Daily, version électronique du supplément au Journal Officiel de l'Union Européenne recensant les avis de publicité de de passation des marchés publics européens) en proposant une identification des entités acheteuses et fournisseuses à travers leurs numéros SIRET (absents des bases d'origines) offre une opportunité unique de croisement avec le registre des bénéficiaires effectifs, qui identifie également les entités grâce à leurs numéros SIRET. Le croisement avec la BRÉF est moins direct et nécessite l'identification précise des personnes physiques à la tête (ou membre influent) d'un exécutif local ou national et leur réidentification dans le registre (à travers leurs noms, dates et lieu de naissance).

Le second questionnement scientifique est plus réflexif sur le processus de mise en open data des données publiques. Comme l'avocat général de la CJUE dans ses conclusions au cas évoqué plus haut le questionnait : « Quel est le juste équilibre entre, d'une part, l'exigence de transparence en ce qui concerne les bénéficiaires effectifs et les structures de contrôle des sociétés, qui joue un rôle fondamental dans le cadre de la prévention du blanchiment de capitaux et du financement du terrorisme et, d'autre part, le respect des droits fondamentaux des personnes concernées, à savoir les bénéficiaires effectifs, et, notamment, de leurs droits au respect de la vie privée et à la protection des données à caractère personnel ? ». La mise en balance par le juge des bénéfices (en termes de capacité d'analyse, de détection des pratiques frauduleuses etc...) et des coûts (en termes de

respect de la vie privée) de l'ouverture des données nécessite une objectivation scientifique qui pour l'heure n'a pas été menée.

### 3. Méthodologie

La récupération des données des bénéficiaires effectifs peut s'effectuer, s'agissant des données françaises, à travers l'API administrée par l'INPI (<https://api.gouv.fr/les-api/api-rbe> ou <https://www.inpi.fr/beneficiaires-effectifs>). Un ensemble de données, de profondeur limitée, censurant un certain nombre de champs intégrant des données personnelles (jour de naissance, adresse, lieu de naissance, date à laquelle la personne est devenue bénéficiaire effectif) est accessible au public, le registre complet étant accessible aux entités habilitées mais de manière beaucoup plus restreinte (en nombre de requêtes et en potentiel d'automatisation de la collecte des données). Avec une mise à jour quotidienne, la base INPI offre une volumétrie de près de 3,5 millions d'entreprises ayant déclaré leurs bénéficiaires effectifs.

L'entreprise PAPPERS (<https://www.pappers.fr/api>) propose quant à elle, en accès payant, un accès à une base enrichie à partir de la base officielle.

Le point le plus délicat est la mise en correspondance des individus décrits dans les données obtenues (BRE et/ou Pappers) avec ceux présents dans la BRÉF. Le problème s'est déjà posé lors de la constitution de cette dernière, car celle-ci agrège plusieurs sources distinctes, nécessitant de réaliser le même type d'alignement. La méthode mise en œuvre repose sur l'exploitation de certaines variables caractéristiques présentes dans les différentes sources, notamment les nom, prénom, date de naissance, localisation.

L'estimation du gain ou de la perte dans la qualité de l'interconnexion du registre des bénéficiaires effectifs avec la BRÉF et la FOPPA en fonction des données présentes dans cette base (données INPI grand public vs. données INPI entités habilitées vs. données PAPPERS) constitue un des attendus du projet. Cette étape nécessite d'effectuer un travail d'annotation manuelle, afin de constituer une vérité terrain permettant de mesurer les nombres de mises en correspondance correctes, erronées, et manquantes.

### 4. Résultats attendus et caractère innovant de la recherche

Interconnexion de trois bases BRÉF, FOPPA et RBE.

Publication d'une nouvelle base intégrant les données du registre.

Publication d'un data paper sur l'interconnexion des bases

### 5. Dimension interdisciplinaire

Le projet BORIS est par construction interdisciplinaire. Il nécessite une triple expertise informatique, économique et juridique.

- Informatique à la fois dans le travail de récupération des données du RBE, de croisement et d'interconnexion des bases FOPPA et BRÉF et de construction de métriques sur la performance de l'interconnexion des bases en fonction des champs de données collectés
- Informatique plus prospective dans l'élaboration de métriques de gains / pertes de performance des outils de prédiction automatiques en fonction de la profondeur des données mobilisées
- économique, dans l'analyse des liens de propriétés qu'il permettra de mener

- juridique, dans l'analyse des tensions que font naître les réglementations promouvant l'ouverture des données publiques et celles encadrant le respect des données personnelles

#### 6. Partenariats extérieurs envisagés

- Partenariat avec le laboratoire d'analyse économique du droit (CRED), Université Paris II Assas :  
  
Mme Téa Toutounji, doctorante contractuelle, sous la direction du Pr. Bruno Deffains, travaillant sur « Gouvernance numérique des services publics et corruption », sera associée au projet.

#### 7. Valorisation (si prévue) : déclaration d'invention permettant de valoriser un savoir-faire, une base de données ou un logiciel

Publication de la base en accès libre (soit en l'intégrant à la FOPPA, soit en proposant une nouvelle base)

Budget (€)*		
	Brève description	Montant
Missions	Déplacements Téa Toutounji (Paris/Avignon)	1 000
Consommables, petits matériels**	Accès base payante PAPPERS 200 000 jetons, pour un accès à 2 000 000 résultats de recherche	2 000
Organisation de réunions		400
Stages***, <b>vacations</b> , prestations	Vacations informatiques ou stage 6 mois	3 180
Budget total		<b>6 480</b>
Co financements le cas échéant	ANR DeCoMaP.	1 500
Budget demandé à Agorantic		<b>4 980</b>
Recettes extérieures		

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

\*\* Petit matériel ne dépassant pas les 600€

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