

Dossier Suivi par :  
DUC Cindy  
Tél : 0187696174  
Mél : cindy.duc@insee.fr

Montrouge, le 19 décembre 2018  
N°2018\_32857\_DG75-E430

**Objet : Fiche de lancement - Expérimentation SSP Lab - Enquête Associations**

Veuillez trouver ci-joint ci-dessous la fiche de lancement de l'expérimentation qui vise à attribuer un domaine d'activité aux associations à partir de la description de leur objet social dans le Répertoire National des Associations.

La cheffe du département « synthèses  
sectorielles »

Signé : Elisabeth KREMP

Pour information : DSE – C. Colin, E. Kremp, S. Dumartin, F.X. Dussud, E. Clément, P.  
Schull, C. Alviset, A. Maurand  
DMCSI – S. Lagarde, E. Coudin, S. Maillard

## Fiche de lancement – Expérimentation SSP Lab :

Attribution d'un domaine d'activité aux associations à partir de la description de leur objet social dans le Répertoire National des Associations.

**Sponsors :** Cindy Duc et Florian Lécrivain, division enquêtes thématiques et études transversales, DSS, DSE

**Sujet :** L'enjeu de l'expérimentation est d'attribuer un domaine d'activité aux associations et d'utiliser ces domaines d'activité pour les strates de tirage de l'échantillon de l'enquête associations. Une partie des associations est répertoriée dans Sirène (employeuses, subventionnées...) mais 50 % d'entre elles ont un code APE 9499Z ne permettant pas de déterminer leur domaine d'activité de façon précise. Par ailleurs, les associations non présentes dans le répertoire Sirène sont présentes dans le répertoire national des associations (RNA), géré par le ministère de l'intérieur. Le RNA est exhaustif mais contient une part non négligeable de fausses actives. Dans ce répertoire, un champ "objet social" rempli en clair décrit rapidement les activités de chaque association. L'expérimentation vise à analyser textuellement ce champ pour en prédire l'activité. L'enquête 2014 appariée au RNA servira de jeu d'apprentissage et de jeu de tests (18 000 observations). Il pourra être nécessaire d'utiliser des algorithmes pré-entraînés pour améliorer la qualité de la classification. On pourra tester et comparer différents algorithmes d'apprentissage supervisé : régressions logistiques pénalisées, forêts aléatoires, réseaux de neurones.

**Objectif :** L'objectif est de pouvoir classer chaque association du RNA dans un domaine d'activité de 10 modalités. Ce classement sera utilisé pour effectuer les strates de tirage de l'échantillon.

**Mode de travail :** L'expérimentation sera également le sujet du mémoire de master ENSAI de Florian Lécrivain (ETET-DSE). Le coordinateur de l'expérimentation sera le tuteur de ce mémoire. Cette fiche tient lieu d'engagement d'ETET et du SSP Lab à la réalisation de l'expérimentation et du mémoire.

Les sponsors/experts métiers ainsi que le coordinateur SSP Lab consacreront en moyenne une journée par semaine à l'expérimentation.

**Appui scientifique :** Thomas Merly-Alpa appuiera cette expérimentation en matière de théorie des sondages.

### Groupe impliqué dans l'expérimentation

Sponsor/Experts métiers : Cindy Duc, division ETET, DSE, Florian Lécrivain division ETET, DSE

Pilote/coordonateur : Sophie Maillard (SSP Lab, DMCSI)

### Rôles

Les sponsors/experts métiers définissent avec l'aide du coordinateur les objectifs de l'expérimentation, assistent aux démonstrations des avancées, jugent de la conformité des travaux à leurs attentes et redéfinissent les propriétés au besoin. Ils assurent aussi la constitution de la base d'apprentissage labellisée.

Le coordinateur SSP Lab assiste le sponsor dans ses missions, pilote et organise les sessions de travail. Il a aussi en charge d'établir les bilans d'étape et s'assure qu'une documentation minimale est disponible.

### Mandat

### Ressources

Le SSP Lab met à disposition le lieu pour accueillir les expérimentateurs lors de session de travail en commun, du matériel pour accéder à la plateforme innovation le temps de l'expérimentation (notamment pendant les sessions de travail communes) pour les agents impliqués qui n'en auraient pas. La DSI (Unissi) met à disposition les accès idoines à la



plateforme d'innovation, et proposera son assistance pour le paramétrage des ressources correspondantes.

### **Calendrier et types de rendus**

Le tirage de l'échantillon de l'enquête associations doit avoir lieu au plus tard fin mai 2019, le livrable final doit être restitué mi-mai 2019.

Novembre 2018 : session de prise de contact, définition des besoins, présentation du mode de collaboration, désignation des personnes impliquées dans l'expérimentation.

Décembre 2018 : brainstorming R, package utilisé, méthode d'analyse textuelle

Août 2019 : remise à l'ENSAI du mémoire de master

En cas de résultats infructueux, la méthodologie utilisée lors de l'enquête Associations 2014 sera reconduite avec quelques points d'amélioration (ajout du secteur institutionnel dans les strates de tirage par exemple).

### **Annexe : prévision d'engagement des équipes**

ETET sponsors/experts métiers

- Cindy Duc : 20 % du temps pendant toute la durée de l'expérimentation
- Florian Lécivain : 20 % du temps pendant toute la durée de l'expérimentation

