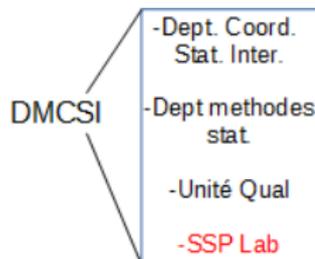


Le SSP Lab: objectifs, missions, fonctionnement

Elise Coudin
responsable SSP Lab, DMCSI, Insee
Pauline Givord
préfiguratrice SSP Lab

Le SSP Lab, c'est qui ?

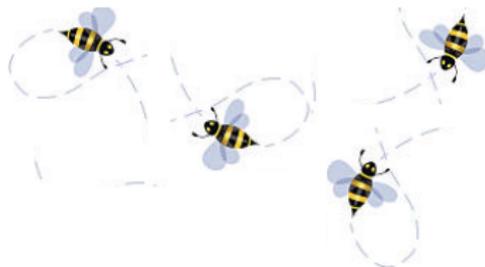
- Nouvelle unité : en place formellement depuis le 1er mai 2018.



- bientôt 8 agents : data scientists, économètres, référent informatique : Yves-Laurent Bénichou, Béatrice Boutchenik, (en sept. François Sémécurbe), Elise Coudin, Julie Djiriguian, Sophie Maillard, Mathilde Poulhes, Simon Quantin, (en sept. Stéphanie Himpens), Benjamin Sakarovitch

Le SSP Lab, pour ?

- pour stimuler l'innovation dans la production statistique, au sein du système statistique publique (Insee, SSM)
- ce n'est évidemment pas l'unique lieu d'innovation
- centre de ressources et d'animation
- travaux en **collaboration** avec les unités métier/SSM
- innovation statistique : autour de la science des données (datasciences)



Le SSP Lab - les origines

- Contexte général:
 - "révolution des données", big data, machine learning, IA...
 - nouvelles données—> nouvelles ressources pour la statistique publique? dans un environnement plus concurrentiel
 - défis techniques et technologiques: adaptation des compétences nécessaires
- Contexte Statistique Publique
 - Eurostat: axe prioritaire vision 2020, certains partenaires (CBS, Istat) déjà sur le pont
 - Insee: expérience données de caisse
 - plusieurs expériences dans les SSM (jobs vacancies Dares, données de santé Drees) - administrateur général des données
 - contexte législatif positif (loi pour une république numérique, loi santé)

De la DMAEE au SSP Lab: nous partîmes à trois; mais par un prompt renfort...

- Création de la division des Méthodes Appliquées de l'Econométrie et de l'Evaluation (DMCSI) en 2012:
 - objectif identifier méthodes innovantes pour la statistique publique, permettre leur appropriation par des applications pertinentes en collaboration, diffuser les connaissances
- Création deux postes de datascientist en 2014 puis 2016
 - expérimentations sur de nouvelles données (ex. partenariat avec Orange, utilisation données internet avec la Conj)
 - nouvelles méthodes (analyse textuelle avec le SSM intérieur)
 - nouveaux modes de travail (équipe mixte DMCSI/SGI deuxième place au hackathon Eurostat en 2017; hackathon les champs de Sirene en janvier 2018)

Le SSPLab : d'Insee 2025 au JO

- Réflexions stratégiques Insee 2025:
 - proposition mise en place d'une structure dédiée à la R&D, unité de veille et d'animation de la recherche
 - Comité de lancement nov 2016-juillet 2017
 - rapport validé par le CD à l'automne 2017: décision de création de l'unité SSP Lab, situé au sein de la DMCSI, en capitalisant sur l'expérience de la division MAEE
- Arrêté du 10 avril 2018 portant modification de l'organisation de l'Insee, avec la création du SSP Lab

Enjeux

- **Enjeu 1** : inciter à, et rendre plus visible l'innovation statistique pour conforter la place de référence de la statistique publique
- **Enjeu 2**: mutualiser les expériences sur l'accueil, le traitement des nouveaux formats de données, gagner en compétences, connaissances, avec réactivité et souplesse

Ce que l'on fait au SSP Lab : expérimenter, collaborer, mutualiser, animer, diffuser

- Domaines d'actions
 - nouvelles sources de données [Sources Méthodes]
 - nouvelles technologies et nouvelles méthodes statistiques des datasciences (yc économétrie) [Sources Méthodes]
 - nouvelles questions/ angles/ compléments
- mener des projets expérimentaux en collaboration avec les services métiers
- assurer une **veille technologique**,
- **diffuser** ces techniques à différents niveaux: formations, documents techniques, ateliers, séminaires, lettre Big Data
- animer ou participer aux réseaux autour de ces sujets innovants, réseaux internes au SSP, ou de pairs (Eurostat, ...) [Réseaux]
- en nouant des partenariats académiques, et avec le privé [Réseaux]

Comment travailler ensemble?

Qu'entend-on par projet expérimental?

Qu'entend-on par projet expérimental?

- sujet "sponsorisé", proposé, porté, par un service métier du SSP
- plutôt court (3-6 mois de travail) ou par étapes avec des rendus différenciés
- approche exploratoire, à "petite" échelle
- test par la preuve
- output: différents formats, de la preuve de concept au prototype expérimental,
- il ne permet pas une mise en production, n'est pas industrialisable en l'état
- il peut ne pas être viable, ne pas être retenu... mais...
- 'droit à l'erreur' propre à l'expérimentation

Comment élaborer un tel projet?

Comment élaborer un tel projet?

- élaboration par allers-retours entre service métier et Lab
- pour une collaboration
- nous contacter avec vos idées au fil de l'eau (recensement des besoins)

Comment sont choisis les projets expérimentaux en association avec le Lab?

Comment sont choisis les projets expérimentaux en association avec le Lab?

- s'ils touchent au moins un domaine d'actions du Lab
- s'ils n'ont pas été traités par ailleurs
- si le Lab a les compétences en interne pour apporter quelque chose (ou s'il peut les acquérir rapidement)
- idéalement projets de l'ordre de 6 mois
- instruire, construire avec souplesse, gestion de stock

Comment travailler ensemble sur un projet expérimental?

Comment travailler ensemble sur un projet expérimental?

- collaboration avec l'unité métier sponsor
- équipe mixte métier, SSP Lab, info, etc..
- engagement souple mais réel des parties [différents modèles d'engagement : 1j par sem, fiche d'expérimentation]
- inspiré de l'agilité : par étape, cycles, avec bilans et rendus réguliers,
- diffusion et partage des résultats et de l'expérience

Peut-on travailler ensemble "hors" cadre d'un projet expérimental?

Peut-on travailler ensemble "hors" cadre d'un projet expérimental?

- Oui!
- Cas où le Lab sert juste d'interface de connexion, d'intermédiaire, (partenariat privé, mise en relation avec la recherche) ou bien d'accueil pour travailler sur des données mises sur la plateforme.
- Cas d'une problématique plus ponctuelle (besoin d'échanges) : ateliers collaboratifs
- Cas d'une demande de conseil/soutien méthodologique en datascience/économétrie
- Cas d'une phase très exploratoire : stages, groupe de statistiques appliquées, ...

Où se trouve **l'info** ? Et comment suivre l'avancée des activités (sans forcément y participer) ?

Où se trouve l'info ? Et comment suivre l'avancée des activités (sans forcément y participer)?

- déjà beaucoup d'éléments sur la page intranet/extranet provenant de la DMAEE : documents méthodologiques, groupes de lecture/formations, outils/logiciels/programmes, Big Data (lettres et séminaires), PAT
- et en construction, une page dédiée aux présentations des expérimentations en cours, des bilans/ résultats/ outputs actualisés au fur et à mesure
- intention de diffuser les retours d'expérience au format adapté

Au fait, quelles sont les **formations** à venir ?

Au fait, quelles sont les formations à venir ?

- Méthodes d'analyse textuelle : Benjamin Sakarovitch et Stéphanie Combes (Drees), le 25 juin
- Apprentissage statistique pour la classification : Benjamin Sakarovitch, Stéphanie Combes (Drees), Julie Djiriguian, 10-11 sept
- Méthodes de décomposition : Sophie Maillard, en novembre
- En préparation ... Python pour la datascience
- En discussion... Git pour tout le monde

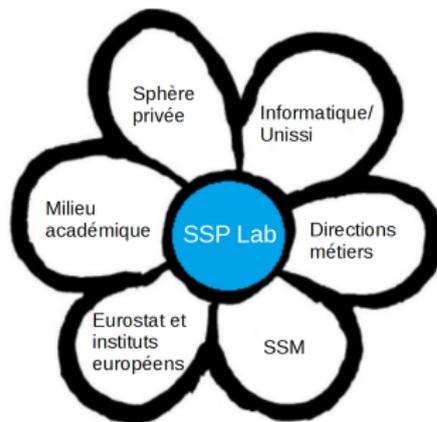
Comment participer au réseau ?

Comment participer au réseau ?

- Réseau ouvert
- Lettre Big Data, intra/extranet
- Yammer (groupe Big Data et datascience : Statistique publique), blog (<https://statoscope.wordpress.com/>)
- Participation aux ateliers, événements collaboratifs, formations, séminaires...

En résumé

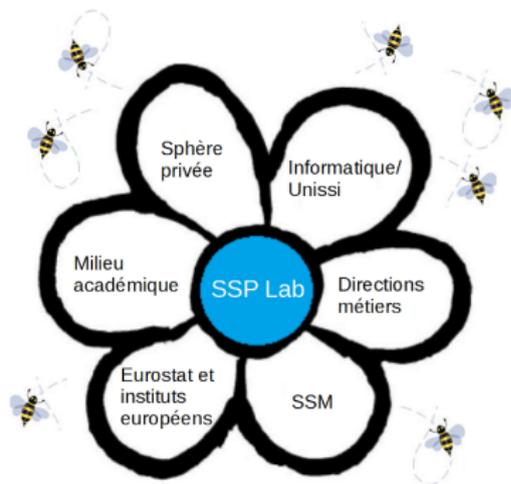
- interface de collaborations internes au SSP et externes (pairs, privé, milieu académique)



- un lieu (salle de travail) et des ressources (ordinateurs avec accès à la plateforme d'innovation de l'Unissi)

En résumé

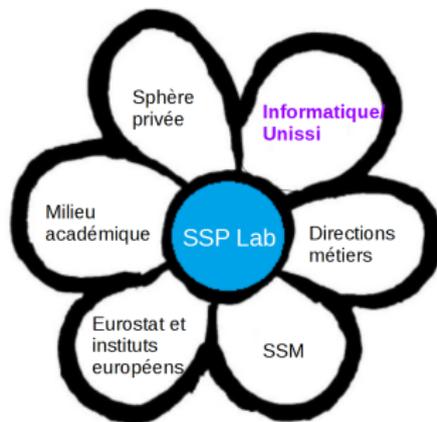
- expérimenter des **pratiques collaboratives de travail** et en **réseau**:
hackathon, équipes mixtes, ateliers, diffusion d'outils collaboratifs...



Programme du séminaire: au fil des projets concrets, illustrations < 10' des missions/aspects/ambitions du Lab

- L'Unissi à l'unisson! [Arnaud Degorre]
- DSN, anomalies et apprentissage statistique [Julie Djiriguan]
- De nouvelles données pour l'analyse des stratégies des élèves de 3ème face à leurs exercices [Sophie Maillard]
- L'insertion du SSP Lab dans les réseaux européens et l'ESSnet Big Data [Benjamin Sakarovitch]
- En France et en Europe : quel apport des données Web pour connaître les offres d'emploi?[Alexis Eidelman]
- Ségrégation sociale, données mobiles et fiscales [Lino Galiana]
- pause
- Hackathon, les champs de Sirene: un événement au sein du SSP [Yves-Laurent Bénichou, Benjamin Sakarovitch]
- Améliorer la chaîne de production du recensement en mobilisant des modes de travail innovants et décloisonnés [Yves-Laurent Bénichou, Benjamin Sakarovitch]
- Linked Open Statistics [Franck Cotton]
- Améliorer l'usage de nos modèles statistiques : Versioning [Mathias André]

La parole à l'Unissi!



Qui contacter?

elise.coudin@insee.fr (resp) ;
mathilde.poulhes@insee.fr (resp-adj 1er juillet);
yves-laurent.benichou@insee.fr (réfèrent info)
beatrice.boutchenik@insee.fr (économètre)
julie.djiriguian@insee.fr (datascientist)
sophie.maillard@insee.fr (économètre/datascientist)
simon.quantin@insee.fr (économètre)
benjamin.sakarovitch@insee.fr (datascientist)

Les projets en cours

Nouvelles sources

- Logs enregistrées d'exercices et stratégies de réponse des élèves [Education]
- Téléphonie mobile, population résidente, mesure de ségrégation [partenariat Orange Labs]
- Accords de branche, quelle protection pour les salariés ? une approche par textmining.
- Données satellitaires et statistique publique (DMS)
- Données des cartes de fidélité dans la grande distribution
- Nouveau panel tous actifs et trajectoires professionnelles salariat non-salariat (DSDS)

Nouvelles méthodes

- "Les champs de Sirene": identification automatique de l'employeur pour enrichir/corriger les informations issues du recensement [DSDS, SGI, DSE]
- Données sociales nominatives, détection d'anomalies et data edition, l'apport du machine learning [DSDS]
- Machine learning pour la prédiction des trajectoires (IPP, PSE)
- Modélisation des effets de pairs en éducation (DEPP)

Documents méthodo en cours

- Méthodes de décomposition pour l'analyse des inégalités
- Modèles de durée dans les études statistiques

Autres études en collaboration: appui méthodologique

- Détermination d'un groupe de comparaison pour les quartiers de la politique de la ville (DDar)
- Effets différenciés des pairs dans la réussite scolaire au lycée (Depp)
- Quantification de l'évitement scolaire sur la ségrégation (Depp)
- L'hétérogénéité de la performance des lycées en France (Dese)
- Rôle des entreprises dans les inégalités salariales entre hommes et femmes (Dsds)
- Discrimination salariale envers les descendants d'immigrés (Dsds)
- Evaluation de la convention d'assurance-chômage 2014 (Dese, Acooss)

Veille et formations

- Machine learning (2j,)
- Analyse textuelle (1j)
- Python
- Méthodes de décomposition
- Versioning, Git (en discussion)

Diffusion et outils de développement de pratique collaboratives

- Hackathons, ateliers collaboratifs
- Diffusion des outils de travail collaboratifs
- Groupe de lecture: frontière entre ML et économétrie pour l'analyse des trajectoires issues de données de panel (projet)
- Ateliers autour du web scraping (projet)

Diffusion et animation de communauté

- intranet, extranet SSM
- lettre Big Data, séminaires Big data
- groupe yammer, blog,
- github

Réseaux

- Eurostat: participation task force Big Data
- Eurostat: ESSnet BigData I, ESSnet City data (flux de mobilité avec données téléphoniques)
- Académique: CREST, IPP, Dauphine