

Hackathon

Les champs de Sirene

séminaire de lancement du SSP lab
05/06/2018

Pourquoi un hackathon pour le SSP ?

Wouai?



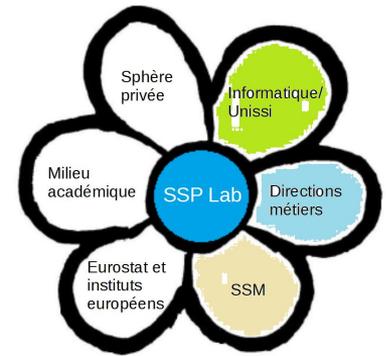
- Animation du réseau SSP
- Formation, montée en compétence sur la *datascience*
- Test du format d'événement et de notre capacité à l'organiser
- Traiter un besoin métier réel adapté au format
- Un événement différent de ceux organisés précédemment : ouvert à tous, non compétitif

Le sujet



- Identifier l'employeur dans les données des bulletins individuels du RP
- Traité actuellement par MCA (Mise en Concordance Automatique) puis par reprise manuelle
- Pour la campagne 2017 :
 - 1 674 000 établissements, 44 % en codage automatique, 56% en reprise manuelle
 - la reprise manuelle occupe 70 personnes pendant 5 mois

Un hackathon pour l'ensemble du SSP



- les 18 et 19 janvier à l'ENSAE
- 61 participants répartis en 12 équipes
- 8 SSM (Intérieur, Justice, Défense, Agriculture, Education nationale, Culture, Environnement, Travail)
- péri-SSM (Pôle emploi, ACOSS, Banque de France, Tracfin, Etalab ...)
- et aussi beaucoup de monde pour la logistique, l'assistance métier et informatique et pour suivre l'avancement des équipes



Avant tout un outil pédagogique :



- 2 journées pour aider les candidats à se préparer :
 - explications détaillées sur le sujet et les données
 - fourniture en avance de phase d'échantillons de jeux de données
 - fourniture d'ordinateurs pour candidats non équipés
 - proposition de 3 pistes et présentation d'outils pour les réaliser
 - aide au regroupement en équipes
- 1 module d'évaluation mis à disposition des participants :
 - retourne le taux de codages réussis, le taux de concordance avec MCA et le compare aux taux des autres équipes

Les retours



Livrables :

- Les 12 équipes ont décrit leurs idées (posters et restitution) et livré leurs codes
- Tout est disponible sur le site [Github](#) de l'Insee, et à reprendre (cf présentation suivante)

Ce que ça a apporté au SSP



- réseau/décloisonnement : des personnes qui se connaissent mieux et des équipes qui ont fonctionné ensemble
- diffusion de connaissances communes au SSP : RP, Sirene
- diffusion des outils/solutions développés par le SSP : API Sirene
- du matériel à exploiter et un “buzz positif” pour continuer (prototypage etc...)

Ce que ça a apporté aux participants



- Personne n'est venu les mains dans les poches
- Même sans compétition, les équipes se sont challengées :-)
- Beaucoup de participants ont profité de l'événement pour trouver du temps pour s'autoformer à python, R ...
- Un bilan de compétences sur les outils informatiques !
- Le hackathon est très prenant : on sort complètement du quotidien, on ne pense qu'à ça même plusieurs jours après
- expérience de dépassement : attention = total focus

Rendez-vous dans un an !

- on recommence ?
- on garde le côté ouvert à tous au sein du SSP
- même style de format
- le sujet : on n'en sait rien pour l'instant
- appel à idées/propositions, qu'est ce qui vous servirait ?
=> le hackathon à votre service



Identification de l'employeur

Les suites

Benjamin Sakarovitch

séminaire de lancement du SSP lab
05/06/2018

Un bel évènement, et maintenant ?

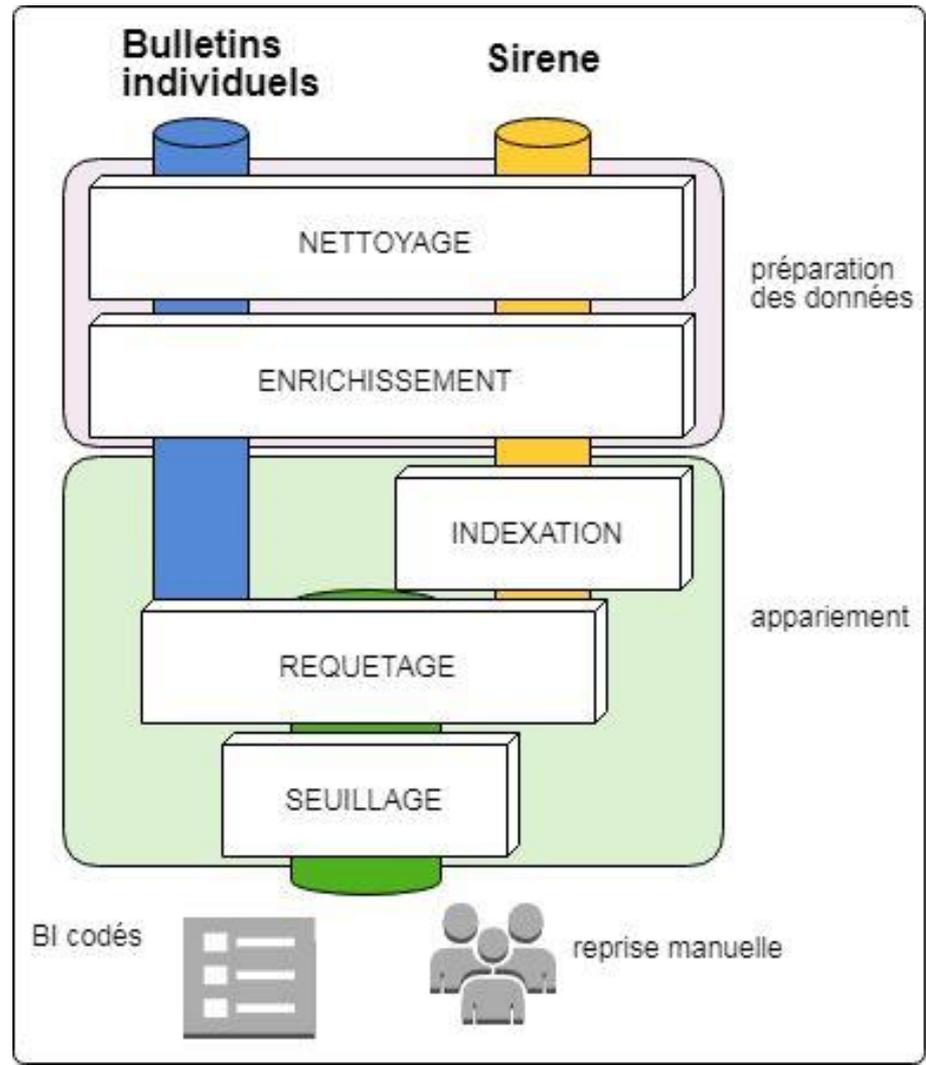
Une fois le hackathon fini, d'où repart-on ?

- fonctionnalités similaires produites par les différentes équipes
- difficulté principale : le passage à l'échelle
- des pistes prometteuses envisagées mais non développées
- du code produit dans un hackathon, pour le résultat rapide et démonstratif

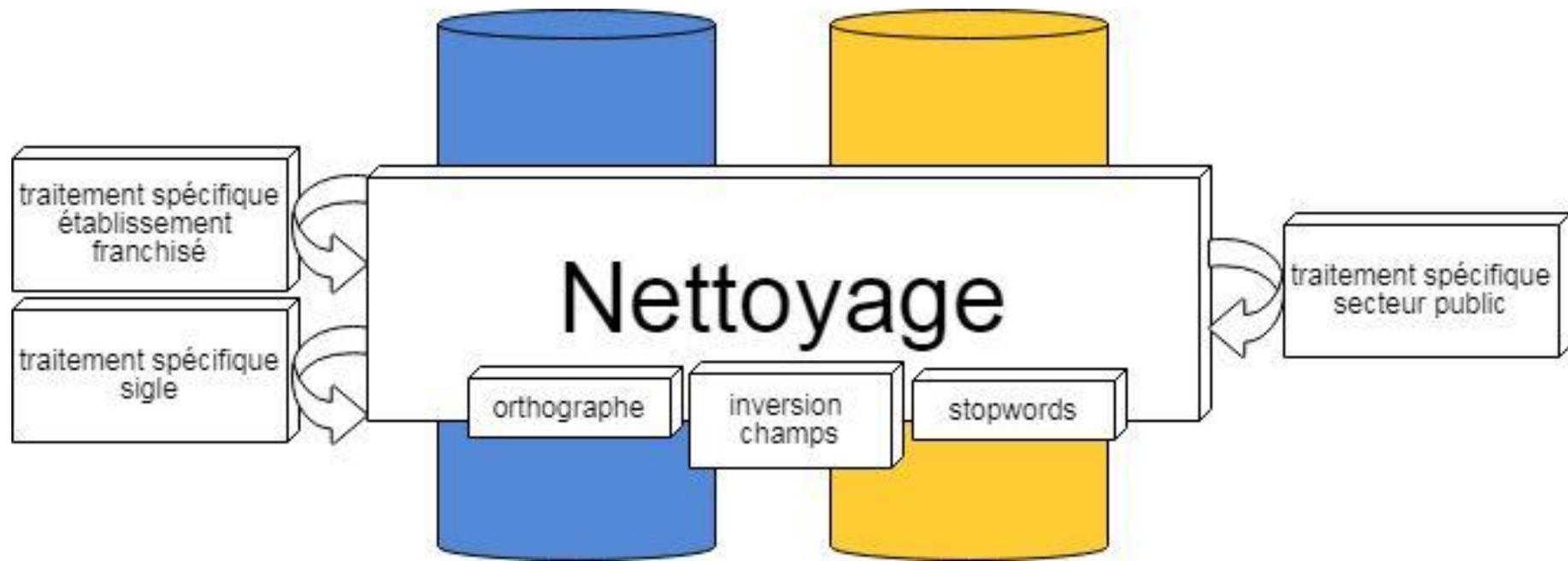
comment intégrer les bonnes idées ensemble ?

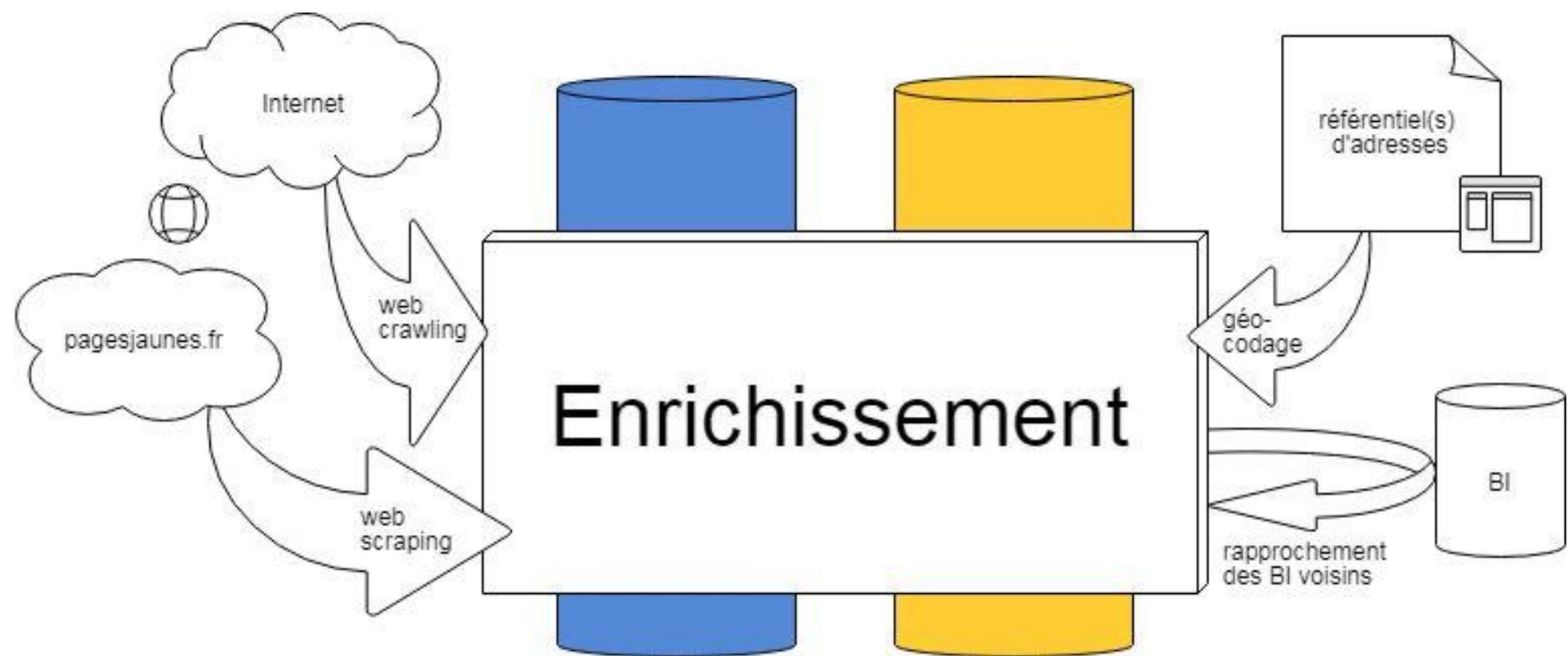
regrouper les idées dans 2 étapes, 5 phases

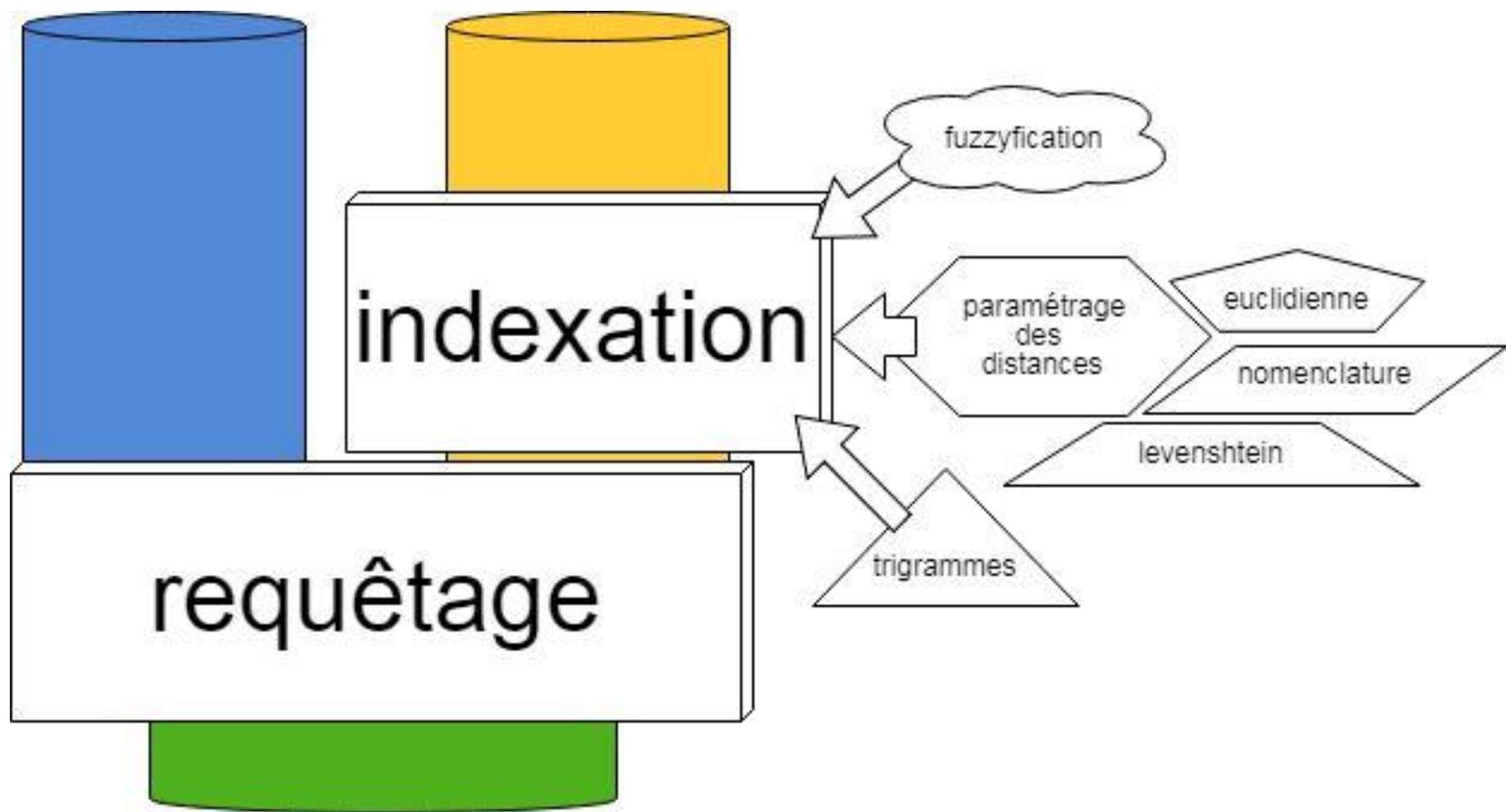
- phases modulaires de préparation de données
- une solution d'indexation/requête pour le passage à l'échelle



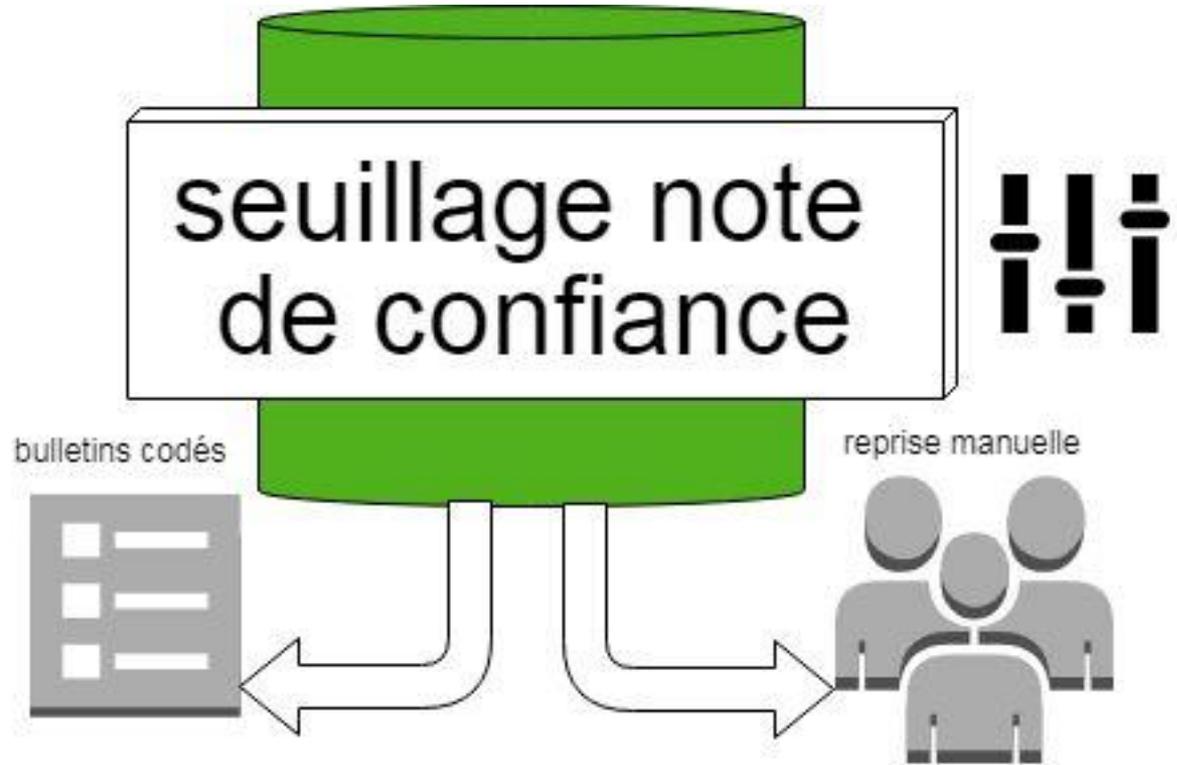
raison sociale
adresse
activité







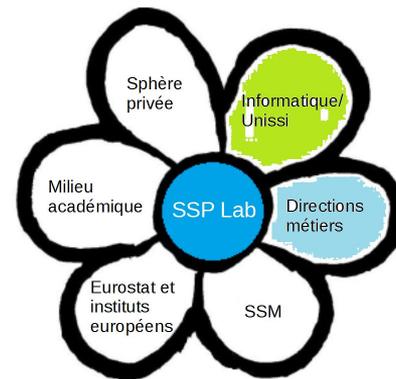
- une étape en dehors des objectifs du hackathon, encore inexploré
- réalisé par MCA
- une étape cruciale



Un projet expérimental

Les acteurs :

- *sponsor* fonctionnel : DSDS
- *équipe* SSP lab, DMS, Unissi et métier DSDS + DSE
- des *satellites* (volontaires, participants au hackathon...)



Le déroulé :

- ❑ *jobs d'été* : 3 ateliers
- ❑ *bootcamp* fin septembre
- ❑ un bilan à 6 mois

Le rendu : un prototype

- ❖ à juger sur ses *performances* (taux de codage, qualité du codage)
- ❖ mais *modulaire*, des améliorations locales possibles