

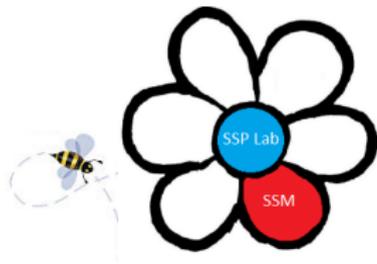
Il n'y a pas que la note qui compte !

De nouvelles données pour l'analyse des stratégies des élèves de
3^{ème} face à leurs exercices

Sophie Maillard

SSP Lab

Séminaire de lancement du SSP Lab - 5 juin 2018



[Sources et Méthodes]

Contexte et motivation

- ▶ Le ministère de l'Éducation nationale étudie actuellement un passage au format numérique des évaluations annuelles des élèves de collège et a conduit des premières expérimentations.
- ▶ Fichiers de *log* contenant toutes les **actions numériques** des élèves : clics de souris, utilisation des outils mis à disposition (calculatrice, compas, règle, etc).
- ▶ Pour les exercices interactifs, ces *logs* constituent une source d'information très riche : en plus du résultat final, ils permettent de reconstruire le **processus** et la **stratégie de réponse** des élèves.

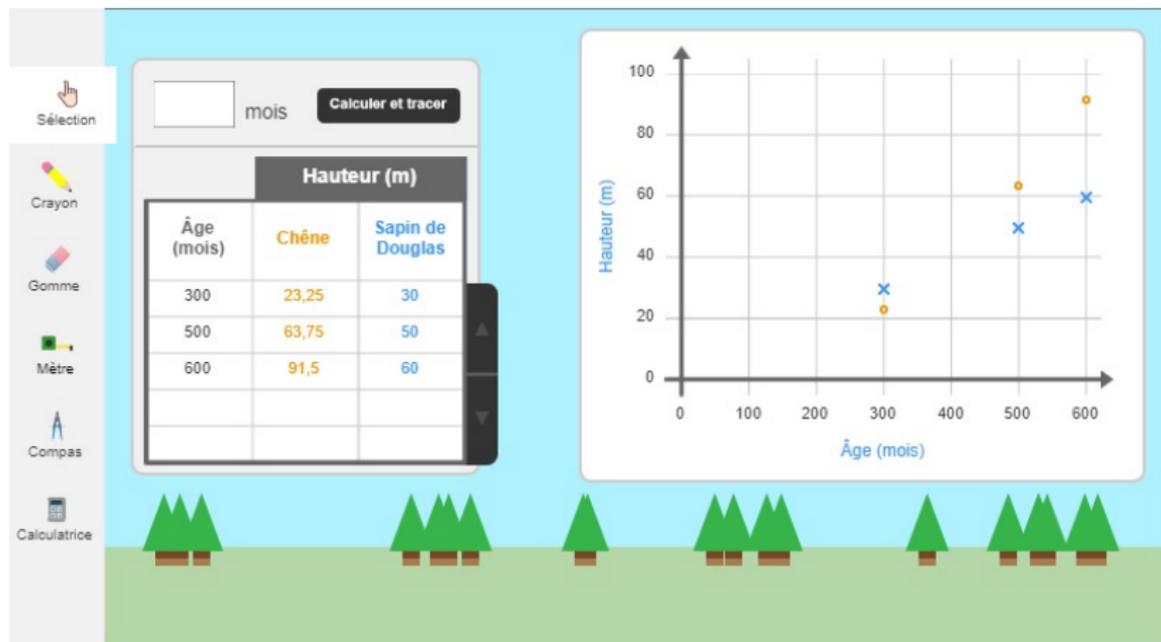
Un travail en coopération Depp - SSP Lab - Unissi

- ▶ Un **projet à l'initiative du Bureau de l'évaluation des élèves à la Depp**.
- ▶ Le SSP Lab et l'Unissi sont mobilisés pour **accompagner l'équipe porteuse du projet**, aussi bien côté informatique que côté méthode statistique.
- ▶ Les principaux objectifs : (1) illustrer l'apport que constituent ces nouvelles données pour une meilleure compréhension pédagogique et (2) aider la montée en compétence de la Depp pour qu'elle puisse traiter les données en interne.

Un exemple : l'exercice de la pousse des arbres

Objectif : trouver à quel âge la taille d'un chêne est égale en moyenne à celle d'un sapin de Douglas.

Principal outil : un tableur permettant de calculer et de tracer sur la courbe la taille correspondant à chaque âge.



Des fichiers peu adaptés à des objectifs d'études

```
root
|-- _id: struct (nullable = true)
|-- componentEvents: array (nullable = true)
|   |-- element: struct (containsNull = true)
|   |   |-- entry: struct (nullable = true)
|   |   |   |-- component_event_id: string (nullable = true)
|   |   |   |-- component_instance_id: string (nullable = true)
|   |   |   |-- component_instance_index: null (nullable = true)
|   |   |   |-- component_type_id: string (nullable = true)
|   |   |-- data: struct (nullable = true)
|   |   |   |--
|   |   |   |-- state: struct (nullable = true)
|   |   |   |   |-- dataList: array (nullable = true)
|   |   |   |   |   |-- element: struct (containsNull = true)
|   |   |   |   |   |   |-- month: integer (nullable = true)
|   |   |   |   |   |   |-- tree1Height: double (nullable = true)
|   |   |   |   |   |   |-- tree2Height: integer (nullable = true)
|-- componentInstances: array (nullable = true)
|-- componentStateCaches: array (nullable = true)
|   |-- element: struct (containsNull = true)
|   |   |-- entry: struct (nullable = true)
|   |   |   |-- component_instance_id: string (nullable = true)
|   |   |   |-- component_type_id: string (nullable = true)
|   |   |-- state: struct (nullable = true)
|   |   |   |-- dataList: array (nullable = true)
|   |   |   |--
|-- device_hash: string (nullable = true)
|-- id_taker: string (nullable = true)
|-- session_hash: string (nullable = true)
|-- session_index: integer (nullable = true)
|-- state_snap_hash: string (nullable = true)
|-- timestamp_end: long (nullable = true)
|-- timestamp_start: long (nullable = true)
|-- user_id: null (nullable = true)
```

- ▶ Un fichier JSON pour chaque exercice × élève
- ▶ Une structure complexe
- ▶ Pas de dictionnaire des variables

La gestion des données

- ▶ Objectif de moyen terme : 800 000 élèves doivent pouvoir passer les tests numériques.
- ▶ Utilisation d'une infrastructure *bac à sable* interne à l'Insee permettant le stockage de données très volumineuses et l'intégration de logiciels spécifiques.
- ▶ Pour les fichiers élèves \times exercice : **MongoDB** - système de gestion de base de données orienté documents, ne nécessitant pas de schéma prédéfini des données. Construction d'une collection par exercice.
- ▶ Pour le fichier agrégé des résultats : **Spark** - système de fichiers distribué et extensible. Une ligne par élèves \times exercice contenant sa réponse finale.

Pistes d'exploitation des données

- ▶ Approche statique
 - ▶ **clustering des élèves** en fonction des outils utilisés, de leur fréquence d'utilisation, du temps passé sur les exercices.
 - ▶ On dispose aussi du sexe et de l'âge des élèves. Les comportements de réponse sont-ils très différents entre garçons et filles ?
- ▶ Approche dynamique
 - ▶ Considérer les **séquences d'actions** des élèves pour y trouver (si elle existe) la méthode.
 - ▶ Distinction parmi les élèves ayant réussi : les élèves efficaces, de ceux qui testent tous les outils, ceux qui hésitent puis trouvent comment procéder.
 - ▶ On utilise des distances entre séquences de caractère pour distinguer les comportements les uns des autres.

En résumé

<i>Enjeux</i>	Faire apparaître les potentialités des données de log et accompagner la Depp vers un passage à l'échelle
<i>Partenaires</i>	Depp et Unissi
<i>Données</i>	Mises à disposition par la Depp, accessibles sur la plateforme innovation
<i>Outils</i>	Mongo DB, Spark

Merci de votre attention !