



L'ÉDUCATION, L'ENSEIGNEMENT, LA FORMATION AUX TEMPS
DU CONTRÔLE SOCIAL FONDÉ SUR LES CLASSIFICATIONS DE
GRANDES MASSES DE DONNÉES :
QUELS ENJEUX ÉTHIQUES EN ÉDUCATION ET EN FORMATION ?

Khansa GHABARA, Université Paris Descartes

22 novembre 2019

LES PLATEFORMES : NOUVELLES UNITÉS DE PRODUCTION

- « **Capitalisme moderne** : importance grandissante du **capital de plateforme**
- Sa reproduction et sa valorisation reposent sur une « **matière première** » particulière, immatérielle : les **données utilisateur** » (Manokha, 2018)



Surveillance
Collecte
Traitement
Production de la valeur

LES PLATEFORMES : NOUVELLES UNITÉS DE PRODUCTION

- Les sociétés à la plus grosse capitalisation boursière au second trimestre 2019

	Microsoft ▲ 1,028,000 ⁶
	Amazon.com ▲ 928,540 ⁸
	Apple Inc. ▲ 911,240 ⁷
	Alphabet Inc. ▼ 751,170 ⁹
	Facebook ▲ 551,490 ¹¹

LES PLATEFORMES ÉDUCATIVES : UNE SITUATION CONFUSE

- Obstacles et incertitudes ➡ Investissements limités : CA des ressources et logiciels éducatifs estimé à 30 M€ par an au maximum (Rapport sur la structuration de la filière du numérique éducatif, Inspection Générale des finances, 2013)
- Face aux contraintes réglementaires, questions autour de la création de valeur hors du modèle publicitaire caractéristique des plateformes à forte capitalisation.

CARTOGRAPHIE DES DONNÉES
EN ÉDUCATION SCOLAIRE :
UN JEU D'ACTEURS

DONNÉES ISSUES DES ENT

Sur les élèves

- Données démographiques
- Devoirs, notes, bilans
- Absences, incidents
- Livret scolaire, parcours
- Traces d'usages des services de l'ENT
- Traces d'usages des outils LMS/de productions pédagogiques intégrés

Sur la gestion de la classe

- Groupes
- Cahier de texte
- Projets de classe

Sur le système éducatif

- Programmes d'enseignements
- Périodes d'enseignement
- Parcours éducatifs
- Diplômes

Sur l'organisation de l'établissement

- Inscriptions
- Emplois du temps
- Dates des évaluations
- Projet d'établissement
- Enseignants, Personnel
- Infrastructures

DONNÉES ISSUES DES SI (RECTORAT)

Sur les élèves

- Données démographiques
- Notes
- Livret scolaire, parcours
- Traces d'usages des services de l'ENT
- Traces d'usages des services Hors ENT
- Journalisation des requêtes/réponses serveurs (filtrage)
- logs des sites visités et des pages vues

Sur le système éducatif

- Programmes d'enseignements
- Périodes d'enseignement
- Parcours éducatifs
- Diplômes

DONNÉES ISSUES D'INFRASTRUCTURES INTERMÉDIAIRES DE GESTION D'ACCÈS

GAR

- Accès aux ressources
- Description des ressources

Proxy

- Traces de navigation

DONNÉES ISSUES DES RESSOURCES ÉDUCATIVES

Traces liées à l'activité enseignante

- Intention pédagogique balisée dans le temps
- Configuration type proposée à l'enseignant

Traces d'apprentissages

- Niveau, Séquences d'activités, Complexité des éléments étudiés
- Traces de navigation, Productions
- Indicateurs calculées à partir des traces
- Profils calculés à partir des traces
- Note globale, Notes spécifiques des activités granulaires, Progression : score/durée
- Conditions d'acquisition, Types de sortie des résultats

Traces d'usages des services intégrés

- Parcours entre les activités granulaires
- Fonctionnalités
- paramètres liés aux objets manipulés : interface, aide, guidage, feedback

Descriptions liées à la ressource

- Description global : titre, résumé, enseignement concerné, ...
- Descriptions spécifiques des activités granulaires
- Métadonnées des images

DONNÉES ÉDUCATIVES : UNE PLACE PRÉDOMINANTE DES SOLUTIONS DE VIE SCOLAIRE

Une quasi-dépendance envers un éditeur de logiciel privé : Rapport de la cours des comptes, juillet 2019



Formation/Education : le top 10 des sites et apps visités en septembre 2019

	Brands/Channels	Audience mensuelle <i>(en visiteurs uniques 2 ans et +)</i>	Couverture mensuelle <i>(en % de la population française de 2 ans et +)</i>
1	IndexEducation.com	4 884 000	7,8 %
2	DigiSchool	3 311 000	5,3 %
3	Le Figaro Etudiant	3 152 000	5,0 %
4	L'Etudiant	3 052 000	4,9 %
5	Pronote	2 990 000	4,8 %
6	EcoleDirecte	2 497 000	4,0 %
7	Education.fr	2 421 000	3,9 %
8	Le Parisien Etudiant	2 151 000	3,4 %
9	Anglaisfacile.com	1 800 000	2,9 %
10	Synonymo.fr	1 703 000	2,7 %

CONTEXTE

- Ministère de l'Éducation Nationale : mission et réflexions en cours sur : « Protéger pour mieux valoriser »
 - GAR : Gestionnaire d'accès aux Ressources
 - Nomination d'un Administrateur ministériel des données
 - Comité Ethique pour les données d'éducation, octobre 2019

QUESTIONS DE RECHERCHE

- Quelles sont les données disponibles ? Dans quelle mesure peuvent-elles faire l'objet d'une analyse ?
- Quelles analyses possibles des données issues :
 - de tâches d'apprentissage disciplinaires ?
 - de l'ENT et des outils généraux de vie scolaire
- De quoi ces données sont-elles la trace?

TERRAIN DE RECHERCHE

- Partenariat en cours de construction :
 - Discussions en cours avec un groupe d'éditeurs de ressources éducatives
 - Implication des acteurs et démarches administratives (Ministère de l'Éducation Nationale, porteurs de projet ENT, chefs d'établissements, enseignants ...)

MÉTHODOLOGIE

- Exploration des données : méthodes, techniques et travaux de référence en EDM et LAK
 - Modèles de prédiction ; découverte de structures, fouille des relations...
 - Analyse automatisée des données textuelles (text mining, analyse du discours)
- Deux tendances :
 - Se rapprocher au plus près du processus d'apprentissage grâce à une analyse fine des actions de l'apprenant
 - Orientation vers les acteurs de l'apprentissage

FOUILLE DES DONNÉES : QUESTIONS ETHIQUES

- La protection de la vie privée : anonymisation oui mais dans quelle mesure pouvons-nous réellement le faire ?
- La propriété et la disponibilité des données : quel cadre clarifiant et facilitant ?
- La généralisation d'un modèle calculé : comment s'assurer que les modèles sont valides pour tous les élèves et pas seulement pour un sous-ensemble d'entre eux ?
- La place de l'erreur : comment appliquer des modèles comportant un taux d'erreur pour la mise en oeuvre d'interventions automatiques ?

RÉFÉRENCES

- Manokha, I. (2018). Le scandale Cambridge Analytica contextualisé : Le capital de plateforme, la surveillance et les données comme nouvelle « marchandise fictive ». Cultures & conflits, (109), 39-59.
<https://doi.org/10.4000/conflits.19779>
- Drachsler, H. & Greller, W. (2016). [Privacy and Analytics – it’s a DELICATE issue. A Checklist to establish trusted Learning Analytics](#). 6th Learning Analytics and Knowledge Conference 2016, April 25-29, 2016, pp. 89-98. Edinburgh, UK. DOI: <http://dx.doi.org/10.1145/2883851.2883893>.
- Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In J. A. Larusson & B. White (Éd.), Learning Analytics (p. 61- 75). Consulté à l'adresse http://link.springer.com/10.1007/978-1-4614-3305-7_4
- Labarthe, H., & Luengo, V. (2016). L'analytique des apprentissages numériques [Research Report]. Consulté à l'adresse LIP6 - Laboratoire d'Informatique de Paris 6 website: <https://hal.archives-ouvertes.fr/hal-01714229>
- Cour des comptes, Rapport public thématique : « LE SERVICE PUBLIC NUMÉRIQUE POUR L'ÉDUCATION : Un concept sans stratégie, un déploiement inachevé », La Documentation Française , juillet 2019, disponible sur www.ccomptes.fr
- Inspection Générale des Finances, Rapport public : « La structuration de la filière du numérique éducatif : un enjeu pédagogique et industriel », juillet 2013, disponible sur www.economie.gouv.fr