



HAL
open science

Données d'apprentissage et données élèves. GTnum GIS2IF "Humanités numériques, entre recherche et formation" #HN_education. Groupes thématiques numériques de la Direction du numérique pour l'éducation (Ministère de l'Éducation nationale et de la Jeunesse)
Eric Bruillard

► **To cite this version:**

Eric Bruillard. Données d'apprentissage et données élèves. GTnum GIS2IF "Humanités numériques, entre recherche et formation" #HN_education. Groupes thématiques numériques de la Direction du numérique pour l'éducation (Ministère de l'Éducation nationale et de la Jeunesse). 2023, 10.5281/zenodo.8278853 . hal-04181599

HAL Id: hal-04181599

<https://hal.science/hal-04181599>

Submitted on 16 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License



Données d'apprentissage et données élèves

Un des aspects des humanités numériques est lié à une activité classique des enseignants : la collecte et le traitement des données sur leurs élèves. La numérisation conduit à divers changements à questionner.

Du fait de la scolarisation, dans des établissements, des données sur les élèves sont collectées par les institutions scolaires, notamment tout ce qui a trait à leur identité (nom, parents...) et à leur place au sein de l'établissement (niveau de classe, groupe, options...) mais aussi à leur attitude (absences, comportements...) et leurs activités et performances scolaires (notes sur les bulletins trimestriels, examens, tests). Certaines données sont communiquées aux parents, les autres demeurant dans l'établissement, remontant parfois au rectorat ou au ministère. Les enseignants gèrent également des données sur les élèves qu'ils conservent pour eux, communiquent parfois oralement et retranscrivent partiellement dans les bulletins. Il ne s'agit pas uniquement de notes, cela peut être des appréciations venant de leur observation du comportement des élèves pendant les cours.

Avec le déploiement de l'informatique et des réseaux numériques, une offre diversifiée de logiciels et de plateformes permet de mettre en œuvre un ensemble d'activités d'apprentissage (Levoine, 2018). Cette offre intègre souvent des modules de gestion et de suivi permettant la collecte et l'analyse des traces d'activité des élèves, et dont l'exploitation est mise au service de l'interprétation des apprentissages en question.

Données d'apprentissage : définition

Les « données scolaires » (produites et traitées par les établissements scolaires) correspondent à des « états » (niveau de classe, groupe, notes, appréciations). Les traces d'activité que l'on peut récupérer quand les élèves utilisent des dispositifs numériques correspondent à des données

d'apprentissage : traces de « comportement » (connexion, navigation, etc.) et moments ou processus d'apprentissage. Ces derniers peuvent être observés par les enseignants et sont plus rarement enregistrés par des dispositifs techniques.

La notion de donnée est complexe, le nom étant mal choisi (Galonnier et al., 2020). Latour préfère ne pas parler de « données », mais d'« obtenues ». La question est plutôt : « Quand est-ce une donnée ? » En effet, elles sont générées, construites et « existent au sein d'un contexte dont elles tirent leur sens, qui provient du point de vue de l'observateur » (Borgman, 2020, p. 40).

Toutes ces données sont considérées comme des données personnelles, nécessitant de respecter un cadre législatif, celui du règlement général européen sur la protection des données personnelles (RGPD ; voir, par exemple, [RGPD : par où commencer](#), [CNIL](#) ou [Protection des données personnelles et assistance](#), [ÉduScol](#)). Notons qu'une anonymisation complète des données est complexe, nécessitant de trouver un compromis acceptable, un degré d'anonymisation dépendant de la quantité de travail nécessaire pour désanonymiser.

Finalité des collectes et traitements

Sur la collecte et le traitement de données d'apprentissage, existe une littérature internationale de recherche abondante, avec deux grandes tendances pour la finalité des traitements :

- > vers la compréhension des processus mis en œuvre par les élèves, grâce à une analyse fine des actions de l'apprenant ;
- > vers les acteurs de l'enseignement pour la mise en œuvre de leviers d'intervention ou de régulation (Ghabara, 2020).

Ainsi, récupérer des données permet d'établir un diagnostic sur les connaissances des élèves et fournit une base pour choisir les interventions

éducatives. C'est ce que font les enseignants sur la base des données qu'ils collectent. Le développement de services numériques contribue à multiplier les données et facilite leur circulation.

Données actuellement collectées et traitées

Zablot *et al.* (2021) ont mené une étude sur les pratiques des fournisseurs de ressources en France (éditeurs scolaires, associations d'enseignants, entreprises développant des technologies éducatives). Il apparaît d'abord que la notion de données d'apprentissage est finalement assez mal partagée. Entre définitions juridiques souvent larges et non contextualisées, définitions techniques généralement très, voire trop, précises, et définitions commerciales souvent insaisissables, les types de données dont on parle s'entremêlent et les discours s'éloignent des questions d'apprentissage.

Les données « élèves » collectées sont peu nombreuses et toujours les mêmes : des données d'identité et de scolarité et quelques données d'apprentissage (le score obtenu par l'élève à chaque activité, le temps passé, des indicateurs de progression). Peu de traces d'activité sont collectées et peu de traitements sont réalisés. L'utilisation de techniques d'intelligence artificielle reste très marginale : pour les entreprises, l'effort nécessaire pour produire des systèmes innovants est jugé très important, sans certitude de retour rapide sur investissement.

Le diagnostic ressort de modèles peu approfondis et les interventions s'appuient sur des modèles simples et robustes (répétition, renforcement). Contrairement aux discours promotionnels, affirmant disposer de techniques (prétendues d'intelligence artificielle) permettant de connaître les besoins des élèves, connaître la situation des élèves (classe/groupe) est en général suffisant. Elle indique les normes attendues (référentiel scolaire) et l'identification des écarts par rapport

à ces normes (via des enquêtes ou des exercices spécifiques) permet de mettre en place un système dit de « remédiation », qui se donne pour objectif de réduire ces écarts.

Vers une gestion locale des données d'apprentissage

Dans son rapport sur l'intelligence artificielle (IA), résultant d'une mission parlementaire confiée par le Premier ministre, Cédric Villani (2018) recommande de renforcer le contrôle des enseignants sur les données d'apprentissage et de « s'assurer que l'IA ne soit pas mobilisée dans des logiques de surveillance ou d'optimisation accrue de performance, mais bien pour augmenter le pouvoir d'agir des enseignants dans l'exercice de leur liberté pédagogique et le dialogue avec les apprenants ».

Trois raisons, au moins, militent pour un traitement des données plus local :

- > la réactivité, rapprocher l'intervention du diagnostic, notamment afin que les élèves prennent conscience de leurs erreurs et modifient leur comportement le plus rapidement possible, afin d'éviter la pérennisation de ces erreurs ;
- > la prise en compte d'autres données plus informelles captées par les enseignants ;
- > la mise en place et le renforcement d'une dynamique collective de la classe.

Les enseignants restent les organisateurs et les décideurs locaux de l'enseignement et des activités menées par les élèves. Des changements importants pourraient s'opérer au prix d'une formation des enseignants à la gestion et à l'utilisation de données d'apprentissage selon les possibilités locales, dans les établissements, de gestion de données et d'alimentation des tableaux de bord pour les enseignants (Boissière et Bruillard, 2021).

Éric Bruillard,
professeur en sciences de l'éducation et de la formation,
université Paris-Cité, EDA

RÉFÉRENCES

- > Boissière, J. & Bruillard, É. (2021). *L'École digitale. Une éducation à construire et à vivre*. Armand Colin.
- > Borgman, C. L. (2020). *Qu'est-ce que le travail scientifique des données ? Big data, little data, no data*. OpenEdition Press.
- > Galonnier, J., Le Courant, S., Pecqueux, A., & Nous, C. (2020). *Ouvrir les données de la recherche ? Tracés*, 19.
- > Ghabara, K. (2020). *Données scolaires élèves : étude de l'existant et des possibilités de recueil et d'analyses* in A. Yessad, S. Jolivet, & C. Michel, Actes des 8^e rencontres des jeunes chercheurs en EIAH 2020, (pp. 48-54).
- > Levoïn, X. (2018). *Mutations dans le paysage des ressources numériques à la lumière des BRNE. Synthèse de l'enquête BRNE 2018*. EDA, université Paris-Descartes.
- > Villani, C. (2018). *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*. Rapport public.
- > Zablot, S., Ghabara, K., & Bruillard, É. (2021). *Collecte et traitement de données d'apprentissage : quelles pratiques des fournisseurs de ressources ?* EDA, université Paris-Descartes.