



HAL
open science

Collecte et traitement de données d'apprentissage : quelles pratiques des fournisseurs de ressources ?

Solene Zablou, Khansa Ghabara, Eric Bruillard

► To cite this version:

Solene Zablou, Khansa Ghabara, Eric Bruillard. Collecte et traitement de données d'apprentissage :
quelles pratiques des fournisseurs de ressources ?. EDA. 2021. hal-03838899

HAL Id: hal-03838899

<https://hal.science/hal-03838899v1>

Submitted on 3 Nov 2022

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.

Collecte et traitement de données d'apprentissage

Quelles pratiques des fournisseurs de ressources ?

Solène Zablou, Khansa Ghabara, Éric Bruillard

[EDA](#) et [ARIPEF](#)

4 novembre 2021

1. Le traitement des données d'apprentissage : une réalité peu documentée

La numérisation de plus en plus déployée au sein de l'espace social en général et à l'École en particulier introduit des dispositifs dans lesquels des données en grande quantité sont produites, stockées et quelquefois exploitées. Dans le secteur de l'éducation, 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.

Ainsi, on peut conserver le temps passé sur une activité, le score obtenu, le succès ou l'échec, le degré d'achèvement, le nombre d'essais effectués, etc. Ces données peuvent être associées ou non à des données de gestion de la scolarité telles que le profil, la classe, le niveau, l'enseignement suivi, etc.

La disponibilité de ces données (et des traces d'activité), la possibilité de les exploiter, d'en collecter et d'en produire, et *in fine* de concevoir des stratégies d'analyse ou d'aide à la décision est susceptible de permettre aux fournisseurs de services éducatifs d'améliorer leur offre. Au cœur de ces possibilités, le vieux rêve de la personnalisation de l'enseignement et des apprentissages, notamment via les parcours, paraît une promesse accessible¹.

L'importance stratégique des données est soulignée par le Ministère en charge de l'Éducation nationale qui déclare « placer les données scolaires au cœur de sa stratégie numérique » ([MENJ, 2018](#)) avec une grande attention portée au respect strict de « la vie privée des élèves et de leur famille, des professeurs et des personnels administratifs » et une devise : « Mieux protéger pour mieux valoriser ». L'éducation constitue un des focus sectoriels du [rapport de Cédric Villani](#) (2018) sur l'intelligence artificielle, qui soutient l'idée de développer la maîtrise de l'apprenant sur ses données d'apprentissage en lien avec son équipe pédagogique et d'aider au développement d'un écosystème Edtech en phase avec les valeurs de notre système éducatif.

Face à ces discours promoteurs et prometteurs, on constate qu'il n'existe pas de publication récente sur les types de collecte et les modes de traitement des données d'apprentissage sur les plateformes offrant des services éducatifs dans l'enseignement scolaire français.

¹ <https://www.educlever.com/ambitions>, <https://www.ac-bordeaux.fr/lalilo-personnalisation-de-l-apprentissage-de-la-lecture-122236>, <https://thinkovery.com/blog/3-questions-evan-friburg-domoscio/>

En outre, nous avons pu éprouver la difficulté à collecter des informations sur les pratiques de collecte des données d'apprentissage et à identifier les projets en cours prévoyant une analyse de telles données². Ce contexte a conduit l'association ARIPEF à lancer une étude pour faire le point sur les activités et les projets autour de la gestion des données d'apprentissage, auprès des acteurs principaux du numérique éducatif en France (éditeurs scolaires, collectifs enseignants producteurs de ressources, entreprises EdTech).

2. Méthode employée : analyse de sites web, tests de produits et entretiens

S'agissant d'obtenir un premier état des lieux des pratiques et des projets autour des données d'apprentissage, nous avons souhaité interroger un échantillon diversifié de producteurs, couvrant largement le spectre, afin de recueillir des points de vue d'entités très différentes et d'avoir une idée des réalisations en cours.

L'analyse de sites web a illustré les discours publics accompagnant l'offre des fournisseurs de service.

Au total, cette enquête a concerné 21 entités produisant ou diffusant des ressources dans l'enseignement scolaires (éditeurs scolaires, entreprises EdTech ou collectifs enseignants producteurs de ressources). Pour les trois quarts d'entre elles, les services proposés sont centrés sur l'enseignement scolaire (le plus souvent en français et en mathématiques). Les autres proposent des services non disciplinaires et leur marché n'est pas uniquement centré sur le scolaire.

Quand cela a été possible, nous avons également testé les services proposés par les entités qui proposent des logiciels et des plateformes mettant en œuvre des activités d'apprentissages et/ou des services associés à leur suivi. Si nous n'avons pu collecter que des éléments partiels à partir des compte-rendu et des tableaux de bord et du test de certains parcours, cela a permis d'obtenir des premiers éléments :

- sur l'organisation des parcours d'apprentissage : organisation des activités, caractéristiques des tâches proposées (répétitions, types de répétition), type de feedback utilisé, informations présentes sur le tableau de bord de l'utilisateur (score, temps passé sur une tâche) ;
- sur les informations transmises aux utilisateurs : score, temps passé sur une tâche, nombre de tâches effectuées.

En parallèle des données collectées sur le Web, nous avons mené des entretiens semi-directifs auprès de 10 entreprises.

À cette fin, nous avons élaboré une grille d'entretien thématique afin d'obtenir des informations complémentaires quant aux stratégies de collecte et d'exploitation des données (voir annexe 1) :

2 En lien avec cette étude, un travail de thèse est mené par Khansa Ghabara au laboratoire EDA. Un des aspects consiste à collecter des données d'apprentissage afin de les analyser (Ghabara, 2020).

1. Projet initial (pour les entreprises créées récemment), projets de développement, positionnement par rapport au recours à l'intelligence artificielle et à l'utilisation plus ou moins massive de données.
2. Données d'apprentissage : quels besoins ? Quels modes de collecte ? Quels obstacles juridiques ou techniques rencontrés (collecte ou traitement des données) ?

La question du rôle du GAR (Gestionnaire d'Accès aux Ressources) a aussi été abordée. Ce service mis en œuvre par le Ministère régule la transmission des données à caractère personnel des élèves et des enseignants dans les ressources disponibles via les ENT (environnements numériques de travail). Les fournisseurs doivent préciser les traitements qu'ils souhaitent faire avec les données qu'ils peuvent collecter ou produire (voir MENJ, 2019, pour un bilan CNIL du déploiement du GAR).

Notons que la période de l'enquête a coïncidé avec la publication de l'appel d'offres ministériel « [Positionnement 2^{nde}](#) », concernant la conception d'une plate-forme de remédiation et d'accompagnement pour l'évaluation et l'accompagnement des élèves de Seconde en mathématiques et en français. L'existence de cet appel d'offres a été mentionnée à plusieurs reprises par les producteurs de plates-formes comme élément sensible les freinant dans la divulgation d'informations au sujet du traitement des données personnelles et d'apprentissage pour l'enquête.

De même, compte-tenu des contraintes liées à la situation sanitaire au printemps 2021, les entretiens, d'une durée de 1 heure à 1 h 30 ont été menés à distance, par l'intermédiaire de logiciels de visioconférence. Ils ont été enregistrés et, bien entendu, nous avons anonymisé les opinions recueillies (voir annexe 3).

L'échantillon de l'enquête ne peut pas être considéré comme représentatif, mais il a été constitué en vue de rendre compte autant que possible de la diversité des logiques de conception et de positionnement. Un éditeur scolaire historique a consacré du temps pour répondre à nos questions fournissant des pistes intéressantes pour mieux situer la position des autres éditeurs scolaires. Par ailleurs, les interlocuteurs ont émis des opinions et des avis correspondant à ce qu'ils pensaient pouvoir dire dans une étude dont ils savaient qu'elle était destinée à être publiée.

Dans la suite de ce rapport, nous allons exposer les résultats auxquels nous sommes parvenus. Nous avons cherché à mettre en évidence les dynamiques d'évolution des services des entreprises et les contraintes rencontrées relativement à la question de la collecte et de l'exploitation des données « élève ».

3. Résultats

Avant de rendre compte des discours des personnes interrogées, il apparaît important de discuter la notion de « donnée d'apprentissage ». Nous l'avons délibérément employée, sans chercher à la préciser dans les entretiens, afin de laisser nos interlocuteurs exprimer leur point de vue selon le sens qu'ils attribuaient eux-mêmes à cette notion (Voir annexe, présentation de l'enquête).

3.1. « Données d'apprentissage », une notion encore floue

La notion de données d'apprentissage est finalement assez mal partagée par les divers interlocuteurs. 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 de l'objet de cette recherche qui est l'apprentissage. En ce sens, pour ce qui relève de la collecte de données, nos interlocuteurs ont le plus souvent insisté sur les démarches utilisées pour l'obtention de données personnelles (ex : nom et prénom de l'élève, de l'enseignant, classe).

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 : (1) vers la compréhension des processus mis en œuvre par les élèves, grâce à une analyse fine des actions de l'apprenant, ou (2) vers les acteurs de l'enseignement pour la mise en œuvre de leviers d'intervention ou de régulation (Ghabara, 2020). En français, la revue [STICEF](#) (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation) publie régulièrement des articles sur le sujet (voir aussi Boissière et Bruillard, 2021).

La notion de donnée est complexe. Galonnier *et al.* (2020), dans l'éditorial d'un numéro de la revue *tracés* consacré aux données de la recherche en sciences humaines et sociales, retracent l'évolution du mot « donnée », particulièrement mal choisi. Ils citent ainsi Latour qui propose de ne jamais parler de « données », mais « d'obtenues ». Les données n'existent pas par elles-mêmes mais sont générées et construites. Les données « existent au sein d'un contexte dont elles tirent leur sens ; ce sens provient aussi du point de vue de l'observateur » (Borgman, 2020).

S'agissant des ressources éducatives dans le contexte scolaire, on peut considérer que l'on peut collecter d'une part des « données scolaires » (manipulées par les établissements scolaires), correspondant à des « états » (niveau de classe, groupe, notes, appréciations) et d'autre part des traces d'activité que l'on peut récupérer quand les élèves utilisent des dispositifs numériques. Il s'agit souvent plus de traces de « comportement » (connexion, navigation, etc.) que d'éléments correspondant à des moments ou des processus d'apprentissage, ce qui a notamment posé question avec l'analyse de données issues des Mooc, qui s'est avérée plutôt décevante (voir Boissière et Bruillard, 2021).

Pour revenir à notre étude, notons que la Cour de justice de l'Union européenne a jugé que les productions d'élèves (et a fortiori celles des enseignants) sont des données à caractère personnel³. C'est ce dernier aspect qui prime pour la gestion et le ministère, dans ses échanges avec les fournisseurs de ressources, différencie plusieurs types de données, une manière de leur donner un statut quant à la logique de protection des individus. Sont ainsi distinguées des données dites

- d'*usage* : données de connexion et données de navigation

³ CJUE, 20 décembre 2017, Peter Nowak c/ data protection Commissionner, n°C-434/16 commenté dans la LIJ du MEN n°202 de mai 2018

- de *personnalisation* : celles liées aux actions de l'utilisateur, en dehors d'une activité explicite de production de documents (ajouter un signet, une note, un surlignage),
- de *production* : ensemble des productions volontaires de documents par l'utilisateur,
- d'*évaluation* : données produites par la ressource ou la plateforme aux fins d'évaluation.

Une telle prise en compte, issue d'une contrainte de meilleure gestion des données personnelles et de leur durée de conservation, ne simplifie pas vraiment une compréhension partagée de la notion de donnée d'apprentissage. Cela permet de comprendre les difficultés rencontrées par nos interlocuteurs de sortir d'un discours sur les données personnelles pour aborder les données d'apprentissage.

Par ailleurs, à partir des entités interrogées et des analyses effectuées, on voit se dessiner une restructuration du paysage éditorial.

3.2. La production de ressources : un domaine en cours de restructuration

L'analyse des projets menés par les différentes entités, éditeurs scolaires historiques, jeunes pousses de la EdTech et collectifs enseignants conduit à repérer une première restructuration autour de quatre catégories de productions ou d'offres :

- de ressources pour les enseignants (banques ou plateformes) : exercices, éléments de cours, outils pour la conception de ressources, vidéos...
- d'une plateforme d'apprentissage pour les élèves : les activités peuvent toutefois être encadrées par les enseignants ou les parents ;
- de solutions techniques pour les plateformes destinées aux enseignants ou aux élèves, des prestations concernant la collecte et l'exploitation de données aux fournisseurs de ressources ;
- de services : initialement pas destinées à un cadre scolaire, mais dont le marché s'est progressivement dirigé vers le scolaire ou le parascolaire.

Ces catégories sont perméables, dépendantes des évolutions des entreprises et de leur marché, et un groupe ou une entité peut intégrer des activités liées à plusieurs de ces catégories (un éditeur scolaire peut en outre produire des éléments pour le marché parascolaire et mettre en œuvre une plateforme). Nous allons présenter quelques caractéristiques de chacune de ces catégories.

3.2.1. Des fournisseurs de ressources pour les enseignants

La première catégorie concerne un peu plus de la moitié de l'échantillon étudié. On y trouve les éditeurs scolaires, des start-up de la EdTech dont le premier marché a coïncidé avec la publication d'appels d'offres du ministère au sujet de la création de banques de ressources et des collectifs enseignants ou des enseignants indépendants qui créent des ressources.

Les éditeurs scolaires et les start-up de la EdTech ont un marché en constante évolution, mais constitué en grande partie par la commande publique. Ils se positionnent sur les appels d'offres en cours et font évoluer leurs offres en fonction. En ce sens, ce sont eux

qui adaptent leurs services et qui ont développé (ou fait appel à des prestataires) des plates-formes d'apprentissage.

Pour ce qui relève des collectifs enseignants, un seul a développé sa plateforme, au départ, pour répondre à une demande de parents. Pour les autres, il n'y a aucune trace d'activité qui soit collectée, puisque les services proposés visent à soutenir le travail de conception de cours par l'enseignant. Seules des données personnelles de l'enseignant sont collectées, c'est-à-dire celles qui permettent de l'identifier et qui servent à la création de son compte.

3.2.2. Des concepteurs de plateformes d'apprentissage pour les élèves

La deuxième catégorie concerne les entités dont le marché de départ était fondé sur la conception de plates-formes pour l'apprentissage scolaire ou parascolaire. Elles collectent des données d'apprentissage, c'est-à-dire les traces d'utilisation laissées par l'élève et des données personnelles, c'est-à-dire celles qui permettent de l'identifier. Dans cette catégorie on retrouve des entreprises et des start-up qui peuvent, tout comme les fournisseurs de ressources, faire appel aux entités de la troisième catégorie

3.2.3. Des fournisseurs de solutions techniques

Dans notre échantillon, nous avons repéré deux entreprises qui proposent des plates-formes et des solutions techniques d'exploitation de données aux entreprises. Pour le cas du marché scolaire, leur développement est conditionné par la signature de partenariats, notamment avec les éditeurs scolaires.

3.2.4. Des fournisseurs de service

Enfin, la dernière catégorie est la plus éclectique en termes de services proposés. Ces entités ne proposent pas de plates-formes destinées à un enseignement disciplinaire. Elles ont été créées pour proposer des services d'appui à la conception de cours pour les enseignants, pour la réalisation de projets (en lycée technique ou professionnel, en classes préparatoires aux grandes écoles, dans les instituts universitaires de technologie...) ou proposent des services d'aide à l'orientation scolaire. Comme pour les deux catégories précédentes, elles peuvent collecter aussi bien des données personnelles que des données d'apprentissage.

Notons que les réponses aux appels d'offres nécessitent souvent la création de consortiums.

Par exemple, dans le cadre du [Partenariat d'innovation et intelligence artificielle \(P2IA\)](#), visant la mise sur le marché d'une plate-forme permettant l'enseignement en primaire des notions dites fondamentales en français et en mathématiques, des alliances ont été nouées entre des éditeurs scolaires, des fournisseurs de solutions techniques et des laboratoires de recherche (souvent associés au moins pour la phase de prototypage).

Voici la liste des six lauréats pour la première phase de recherche et de développement :

- en français : *Kaligo* (LEARN&GO, INSA-IRISA, université de Rennes 2, LP3C, UR1-IRISA), *Lalilo et Navi* (Domoscio, Beneylu, Hachette, DXC, AIDODYS, laboratoire CHArt, laboratoire KDIS, MOBIDYS, STORIPLAYR) ;

- en mathématiques : *ADAPTIV'MATHS* (EvidenceB, APMEP, BLUE FROG ROBOTICS, DEASIGN, INRIA FLOWERS) ISOGRAD, LIP6, SCHOOLBA, Sejer, Nathan), *Mathia* (VMPS/Prof en poche, LUMINAI/ARTFACT, Tralalère, CABRILOG) et *SMART ENSEIGNO* (Educlever, CABRILOG, INRIA, WIMMICS, LUDOTIC).

3.3. *Peu de collecte de données « élèves » et peu de traitements sophistiqués lorsque c'est le cas*

Une partie des entités qui ont participé à nos entretiens ne collecte aucune donnée « élève ». Ces entités ne proposent pas de plates-formes, mais des ressources « indépendantes » dont les contenus sont souvent téléchargeables par les enseignants (vidéos, images, exercices).

3.3.1. **Peu de traces d'activité collectées chez les entités qui proposent des plates-formes**

Pour une majeure partie de notre échantillon, à la lecture des conditions d'utilisation présentées sur les sites web, des entretiens et des tests que nous avons pu effectuer sur les différentes plates-formes, des données « élève » sont collectées, mais pas de traces d'activité, et ces dernières sont toujours les mêmes.

Issues des Systèmes d'Information du ministère, ces informations sont de deux types : les données d'identité (civilité, nom, prénom, identifiant, ...) et les données de scolarité (division (classe), groupe, degré d'enseignement, cycle d'enseignement, enseignement(s) suivi(s)...). Elles permettent de recréer le rattachement élève-enseignant sur la base d'une répartition de groupes et divisions dans un établissement donné.

D_FR3 : pour ce qui vient du GAR dans XXX (nom de la plate-forme), ce qui vient du GAR c'est juste le nom, prénom, classe (...) ça sert à mettre des classes, à ce que le prof retrouve ses élèves dans les classes quoi avec leur nom et prénom pour qu'ils puissent leur affecter des contenus ».

Pour les plateformes non reliées aux Systèmes d'Information du ministère, des comptes « élève » sont créés soit avec les données nominatives de chaque élève, soit avec des identifiants dont l'enseignant gère la création et la répartition dans des groupes.

PDG_FR2 : on en a vraiment pas tant que ça, c'est-à-dire que nous la façon dont on fonctionne on ouvre un compte par école, et après, même si c'est un prof qui s'abonne dans l'école, euh dans ce cas là le gestionnaire n'ouvrira un compte que pour ce prof là, euh donc pour ouvrir un compte c'est l'e-mail du prof, l'e-mail académique, fait partie d'une première donnée donc on a donc on a bien rédigé ça dans nos, enfin vous avez du lire nos privacy enfin la confidentialité, euh un e-mail puis un mot de passe qui se choisit, qui est crypté chez nous quoi, et c'est tout au niveau de, puis le nom de l'école, mais c'est pas personnel ça, par contre la chose qui est peut-être un peu plus sensible c'est, que un professeur qui veut ensuite avec ses élèves, ce qui est le cas avec XXX, gérer sa classe, avoir les résultats de sa classe ou quoi que ce soit, il va créer un compte, ultra simplifié par élève, et ce compte-là pour se connecter y'a juste besoin d'un compte classe, un code à 4 trucs, 4 digits, complètement qui a aucune signification, rien de personnel, et ça ça pourrait être suffisant.

PDG_FR2 : sauf que le prof à un moment donné il est quand même obligé de savoir que pour gérer sa classe Julie a réussi mais que Paul a raté donc on va lui demander, un nom et un prénom, et d'ailleurs je disais avec XXX il faut qu'on mette un prénom parce que des données moins on en a mieux ce serait

Pour ce qui est des données d'apprentissage, comme on en faisait l'hypothèse, les plateformes collectent, pour la plupart, le score obtenu par l'élève à chaque activité, le temps passé sur chaque activité et calculent des indicateurs de progression (sous la forme de pourcentage, de badges ou de points obtenus).

D_FR1 : dans ce cadre-là c'qu'on va récupérer à minima c'est, ce qui nous intéresse d'avoir ça va être on va dire le score de l'élève et le fait qu'il ait réussi ou pas l'activité

Des données d'autres types sont collectées sur les plateformes proposant des services non destinés à des apprentissages procéduraux ou de notions, mais venant en appui à l'organisation des cours ou des projets ou encore à l'orientation. Parmi les nombreuses situations de ce type, on peut citer :

- des présentations techniques des projets souhaités quand il est question d'un service d'aide aux jeunes inscrits en lycées professionnels dans le cadre de la réalisation des Chefs d'Œuvre⁴
- des réponses aux tests d'orientation et des traces de navigation (ex : fiches métiers consultées) notamment quand un agent conversationnel oriente l'élève vers des ressources présentant des métiers qui peuvent correspondre à son profil ;

L'absence de collecte de données d'apprentissage : une prise de position ?

Un fournisseur de service se distingue, car il est le seul qui ne collecte aucune donnée d'apprentissage. Comme pour le cas des collectifs et enseignants indépendants qui produisent des ressources son service est dédié au soutien du travail de conception de cours des enseignants.

Cependant, il collecte des données personnelles des élèves pour la création de comptes et permettre, aux enseignants, d'envoyer et de recevoir des contenus au sein de leurs profils utilisateurs.

Dans son espace, l'enseignant peut créer des ressources et des collections de ressources, et les distribuer aux élèves. Il peut aussi récupérer les productions des élèves. Les fonctionnalités de l'espace des élèves se limitent à la possibilité de compléter les ressources reçues et de les

4 Dans le cadre de la préparation au CAP et le Baccalauréat professionnel, la réalisation d'un Chef d'œuvre peut correspondre tout aussi bien à une réflexion de groupe à l'instar des Travaux personnels encadrés (TPE) mis en place dans les lycées généraux et technologiques ou à un projet de création d'entreprise. Informations disponibles à l'adresse : https://cache.media.eduscol.education.fr/file/fam_metiers_et_R_pedagogiques/40/4/Vade-mecum_realisation_chef-d_oeuvre_1081404.pdf

déposer dans les collections correspondantes définies par les enseignants.

Ce parti pris correspond à une volonté, du concepteur de l'entreprise, de ne ni remplacer ni normer le travail de l'enseignant, son marché de départ n'étant pas la production de ressources pour l'éducation :

PDG_FS3 : ça implique un parti pris extrêmement fort sur l'organisation en éducation, à partir du moment où le logiciel va vouloir se substituer et c'est pas mon débat, partiellement ou profondément à l'élève, le concepteur du logiciel va se demander comment, et donc il va forcément intégrer des normes plus profondes, des normes concrètes, sur ce que doit faire un enseignant, à quel rythme il doit le faire, et qu'est-ce qu'un bon enseignant, ensuite les technologies en tous cas la pratique des industriels (...) Facebook, Twitter, Wikipédia ne remplacent pas ce que quelqu'un pourrait faire, en tous cas ne guident pas l'interaction (...) et donc y'a le LMS qui, à ne pas confondre, y'a le réseau social qui dit attention je veux remplacer personne, (...) je vais créer un espace de discussion pour que les gens interagissent entre eux, XXX (nom de l'entreprise) appartient totalement à cette catégorie-là, et à ma connaissance est quasiment unique dans le domaine de l'éducation, voilà pour l'historique par rapport à ça, accidentellement on s'est retrouvé dans le monde de l'éducation parce que les enseignants nous ont utilisés, on a jamais imaginé qu'on allait venir dans le monde de l'éducation en y sachant rien, et très rapidement, de mon point de vue, ne voulant en savoir rien, entre guillemets, c'est-à-dire, en ne voulant rien prescrire.

Les traitements de données sont alors centrés sur la modération de contenu et sur le pilotage des usages à l'échelle de l'établissement ou de l'académie :

- Treerank (inspiré de Pagerank du moteur de recherche Google) est utilisé pour faire de recommandation de ressources. En revanche, nous ne savons pas comment il est utilisé :

PDG_FS3 : « on n'a pas besoin de lui apprendre (a l'algorithmme utilisé) ce que c'est que les maths. Il va trouver un fichier Geogébra dans un compte, le même type de fichier dans un autre compte (...) il fait des rapprochements. »

- des algorithmes de modération des contenus : pour la détection automatique des contenus pornographiques, des insultes et du harcèlement.

3.3.2. Peu de traitements lorsqu'il s'agit de plates-formes liées à des enseignements de contenus disciplinaires

Le croisement des éléments issus de l'analyse web avec ceux issus des tests de plateformes et ceux issus des entretiens nous a permis d'identifier des types de traitements présents au sein des différentes plateformes. Pour chaque plateforme, nous avons cherché à identifier la(les) forme(s) correspondantes de personnalisation et la manière de la mettre en œuvre, ce qui implique de déterminer de quoi le traitement dépend en entrée, ce qu'il renvoie en sortie et avec quelle technique. Cette analyse, nous a permis de mieux affiner l'identification des données nécessaires aux différents traitements même si les informations collectées sur les techniques utilisées ne sont que partielles.

Ces données sont exploitées essentiellement, pour recommander du contenu sur la base de la mesure du degré de maîtrise d'une règle donnée avec des algorithmes à structure conditionnelle simple. Les données recueillies ne sont pas suffisantes pour élaborer des profils approfondis des apprenants et on se contente de mettre à jour des éléments de maîtrise plus ou moins grande dans une liste de compétences à acquérir. Si des projets plus avancés ont été évoqués au cours des entretiens, cela correspond à des projets de recherche et développement mis en veille ou abandonnés car jugés non prioritaires ou par manque de financement.

On constate un manque de discours autour des techniques utilisées pour le traitement des données, qui peut être expliqué de plusieurs manières. D'abord, comme la question de la collecte et du traitement de données personnelles est sensible, les interlocuteurs ressentent le besoin de partager leur expérience et leurs difficultés de développement. Il a souvent été nécessaire de réorienter les questions afin de recueillir des informations sur les données d'apprentissage.

Ensuite, on peut faire l'hypothèse que les fournisseurs qui se consacrent au marché scolaire n'ont pas besoin de beaucoup de données à partir du moment où le service est destiné à l'enseignant qui va se charger de la remédiation éventuelle. Pour l'enseignement professionnel ou pour la gestion de services, on repère d'autres types de traitements.

Des données à partager pour la réalisation de projets

Une entreprise propose un accompagnement à la réalisation de projets pour les jeunes de 13 à 30 ans, il n'y a pas de traitement de données d'apprentissage en tant que telles. Ce sont les données produites par les élèves à des fins d'aide à la réalisation des projets qui sont exploitées. Pour le concepteur de la plateforme, il ne s'agit pas réellement d'intelligence artificielle, mais d'utilisations d'algorithmes qu'il qualifie d'*avancés* :

- la plate-forme utilise *un algorithme de matching* qui détecte, à partir des présentations de projets, les liens entre les termes du projet et les spécialités des professionnels membres de la plate-forme. L'enjeu est de permettre, aux jeunes/élèves d'entrer en relation avec des professionnels spécialisés du domaine visé par le projet. Les données exploitées par cet algorithme sont contrôlées par une équipe spécialisée, chargée de la mise en relation des professionnels avec les jeunes/élèves ;
- le même algorithme est utilisé pour suggérer aux jeunes/élèves d'entrer en relation avec d'autres qui travailleraient sur des projets similaires.

L'enseignant est également considéré comme un médiateur des apprentissages des élèves. En cela, il a à disposition un tableau de bord qui lui permet de suivre la progression de l'élaboration des projets. Cela suppose de collecter régulièrement des données et de les organiser, mais nous n'avons pas pu obtenir davantage d'informations à ce sujet.

3.3.3. Normes et collecte de données

Pour ce qui relève de l'exploitation des données d'apprentissage, les normes SCORM et xAPI sont majoritairement citées même si les limites de chacune d'entre elles sont souvent mentionnées.

S'agissant de SCORM, bien que systématiquement critiquée pour la pauvreté de son modèle de donnée et la complexité de son implémentation, elle reste assez largement utilisée par soucis de compatibilité avec d'autres plateformes ou de conformité avec les exigences des appels d'offres auxquels la majorité des éditeurs répondent.

PDG_FR2 : si moi j'ai compris la question en fait on a rien d'innovant tout ce qu'on fait on l'a développé, c'est pas sur des formats OpenSource, mais c'est du JavaScript HTML5 on voulait que ça marche partout ; et donc on utilise les choses et c'est là aussi que on est pas du tout sous les radars de certaines choses des technologies innovantes, la France aime bien soutenir des choses à des années lumières de c'qui se passe dans les classes, ben non on a rien de bien spécial et même en termes de normes je me rappelle vous savez y'avait une mode de LOMFR, de SCORM des trucs comme ça, mais pareil ça nous rapporte pas ; c'est pareil les gens nous demandent est-ce que vous êtes ça ?

D_FR2 : Ouais c'était essentiellement des arguments lorsqu'on répondait à des appels d'offres quoi, mais après sur le domaine applicatif on en a jamais vu la couleur.

Quant à xAPI, le défaut essentiel résiderait dans le manque de vocabulaires pouvant prendre en compte toutes les spécificités d'un contexte d'apprentissage. Cela contraint donc certains éditeurs à développer leurs propres vocabulaires et à s'écarter, finalement, de la norme.

DFP2 : On a utilisé au maximum les vocabulaires existants qui sont créés par ADNET mais ce n'est pas suffisant. On avait besoin de plein de trucs qui n'existaient pas du tout. Du coup on a fait nos propres vocabulaires, nos propres extensions. Un exemple la production de ressources : nous dans notre système, l'élève selon la typologie, peut être en capacité de produire lui-même une ressource qui va être ensuite vue par l'enseignant et qui devient un nouvel objet. Ça, ça n'existait pas dans xAPI, dans les vocabulaires de l'ADNET. On avait besoin d'une extension pour décrire quelle était la ressource produite.

Vers de l'enseignement adaptatif ?

Nous avons recueilli des éléments laissant penser que les traitements opérés peuvent aller au-delà de l'exploitation des données disponibles depuis des logiciels qui utilisent des standards comme SCORM dès lors que les fournisseurs ont d'autres marchés qui concernent essentiellement le parascolaire.

C'est ce qui semble être le cas pour cette entreprise qui propose des parcours d'apprentissage qui se veulent adaptatifs. Avant de commencer un parcours, un test de positionnement est

proposé à l'apprenant. En fonction des résultats obtenus, l'apprenant se voit proposer des questions et des contenus de remédiation définis en fonction des réponses données tout au long du parcours d'apprentissage. Les données collectées incluent également les informations concernant les remédiations automatiques apportées par le système :

PDG_FT2 : et donc votre question c'est de quoi on a besoin ben on a besoin de tout, on a besoin de tout en fait on a besoin de savoir à chaque question qu'est-ce qu'il a répondu, confronter la réponse à la bonne réponse, c'est-ce qu'il a été aidé par le système combien de temps il a mis pour répondre, donc on collecte ça (...) quand je dis qu'on a besoin de tout, de c'qu'il a fait et du temps qu'il a mis pour le faire et est-ce que le système l'a aidé oui... on a besoin de tout ça

Il a été impossible de collecter des informations techniques précises concernant les modalités d'exploitation des données, probablement parce que le système développé appartient à l'entreprise, mais nous avons eu des éléments au sujet des principes de fonctionnement :

PDG_FT2 : le système calcule en temps réel le degré de maîtrise de la règle par l'individu, quand il arrive à 100% on considère qu'il la maîtrise et donc on l'embêtera plus avec cette règle et quand le système voit, au fur et à mesure des itérations, que l'apprenant n'arrive toujours pas à faire l'exercice, alors il va l'aider, donc l'aider ça peut être, l'aide va dépendre du type d'exercice, mais ça peut-être lui balancer une vidéo, vidéos explicative (...) en espérant qu'il va comprendre quelque chose (...) le système est capable d'arrêter le jeu (...) en disant regarde la vidéo et... derrière on lui demande est-ce que tu as compris, il doit répondre oui j'ai compris, donc on lui fait faire des petits exercices, juste pour s'assurer qu'il a compris

Le service proposé par le concepteur se veut *adaptatif*. Ainsi, les réponses de l'apprenant sont évaluées selon des règles. Des exercices sont proposés en fonction de l'évaluation de son degré de maîtrise de chacune des règles et des moyens de remédiation sont proposés dès lors que le système interprète des erreurs comme des difficultés de compréhension. La conception de la plate-forme implique la définition de bases de règles et d'exercices conséquentes.

3.3.4. Des services proposés à destination des enseignants et des parents

Le principal public ciblé par les producteurs de ressources et supports de ressources reste les enseignants, ainsi que les parents dans le cas de services parascolaires. Des tableaux de bord permettent de rendre compte, de manière plus ou moins fine, des activités et de la progression des élèves.

En ce qui concerne le secteur scolaire, ce positionnement reflète un parti pris au sujet du rôle de l'enseignant. Les fonctionnalités proposées au sein de tableaux de bord permettent d'avoir un compte-rendu sur les résultats obtenus par les élèves, le temps passé pour chaque activité, les activités déjà réalisées et celles qu'il reste à effectuer.

Toutefois, les tableaux de bord n'offrent pas exactement les mêmes fonctionnalités. Nous avons rencontré deux grandes catégories.

La première propose aux enseignants le choix de fonctionnalités de remédiation. De manière assez classique lorsqu'il s'agit d'enseignements disciplinaires, ces activités sont réparties selon un référentiel avec des données élèves⁵ :

L'outil auteur intégré permet de créer, modifier et partager des contenus pédagogiques. Selon l'objectif pédagogique, des groupes d'élèves peuvent être constitués. L'enseignant peut leur assigner des parcours spécifiques et suivre la progression de chacun de ses élèves en temps réel. [Enseigno \(Educlever\)](#)

Labomep est une plate-forme permettant aux enseignants de proposer à leurs élèves des exercices de découverte, d'apprentissage, d'entraînement, d'approfondissement et d'en suivre les résultats, ainsi que résumés de cours animés, et beaucoup d'autres ressources (calcul mental, manipulation encadrée d'un logiciel de géométrie, etc.). Sa conception autorise la pédagogie différenciée ; il est en effet possible d'adapter les parcours pédagogiques à chaque élève, en créant des sous-groupes (à partir d'un individu), en utilisant un nombre important de ressources (en constante évolution) et d'en créer de nouvelles, ainsi que de structurer ces ressources selon un ordre et des critères logiques personnalisés. [Labomep \(Sésamath\)](#)

Les services sont centrés sur l'accompagnement des apprentissages des élèves, et à ce titre, les enseignants peuvent assigner des devoirs ou des ressources en fonction des résultats obtenus par les élèves à chaque activité.

La seconde catégorie de plates-formes propose des tests de positionnement en début de parcours. Ces tests et les parcours d'apprentissage sont constitués de séries d'exercices (par exemple sous la forme de QCM) qui sont définis en fonction de référentiels de compétences. A chaque compétence correspond des niveaux de maîtrise attendus.

Sur les tableaux de bord, les enseignants ou les parents peuvent consulter les activités réalisées par les élèves, et leur progression. En revanche, contrairement au premier cas, ils ne peuvent pas proposer eux-mêmes des exercices ou des contenus complémentaires en passant par la plate-forme :

PDG_FP2 : C'est-à-dire bah on a un référentiel. On peut évaluer les gens dessus ; on peut sortir des profils d'acquisition mais après on ne sait plus rien faire parce qu'on n'a pas donné le sens au référentiel. C'est ce que j'appelle moi les relations didactiques c'est-à-dire de dire que cet élément est le pré-requis de celui-là, cet élément est le même que celui-là mais en plus compliqué ; cet élément est un levier de compréhension c'est-à-dire que s'il ne comprend pas ça, peut être que tu peux lui sortir celui-là et éventuellement ça peut le débloquent. Tout ça c'est ce qui donne du sens au référentiel. Une fois que j'ai ça, par contre, les algos peuvent tourner à plein régime et vraiment exploiter le référentiel pour exploiter quelque chose de pertinent.

Ainsi, quel que soit le type de fonctionnalités qui est proposé, les tableaux de bords sont conçus de manière à permettre aux enseignants ou aux parents de conserver, *a minima*, un rôle d'accompagnateur. Lorsque les parcours d'apprentissage sont définis en fonction

⁵ Les deux extraits suivants sont issus de l'analyse des sites web des entités investiguées, ils sont donc publics.

d'algorithmes, la collecte et l'exploitation des données permettent d'alimenter des comptes-rendus aux enseignants.

3.3.5. Les outils d'analyse web : une place encore mineure

L'utilisation des outils d'analyse web est centrée sur la mesure d'audience, non personnelle et non associée au processus d'apprentissage.

D_FR2 : on utilise Analytics mais seulement pour voir la provenance d'un utilisateur, la provenance c'est-à-dire le territoire pas mal de géolocalisation aussi, on l'utilise pas par contre sur la gestion des données personnelles qui sont déjà niveau GAR c'est pas possible

Certaines entreprises optent pour des solutions « maison » et implémentent ainsi des méthodes de collecte qui permettent de repérer les contenus les plus consultés et de proposer des recommandations aux utilisateurs :

D_FP2 : alors nous on a un outil maison qui nous permet justement dans les référentiels d'aller jusqu'à savoir quelles sont les notions qui sont travaillées par les élèves, donc ça c'est intéressant pour nous, ça permet d'orienter aussi la production éditoriale, ça permet d'orienter l'affichage qu'on veut faire du référentiel (...) et donc on a le métier pédagogique de la ressources on connaît tout sur la ressource, les liens entre la notion et la ressource et donc c'est très facile pour nous, même si on y avait pas pensé avant de reventiler les éléments euh tels clients, les élèves travaillent sur telle notion.

3.4. Possibilités et contraintes pour le développement des services

3.4.1. Le GAR : entre opportunités et contraintes

Le sujet du GAR (Gestionnaire d'Accès aux Ressources) est quasi-systématiquement évoqué par les fournisseurs de ressources au cours des entretiens. Les fournisseurs souhaitant distribuer leurs ressources via les ENT, doivent, en effet, suivre un processus obligatoire appelé « conformité applicative » : justifier d'une part les demandes de données d'identité et de scolarité issues des ENT et, d'autre part, déclarer les données produites au sein de leurs plateformes (selon la typologie du ministère décrite en section 3.1) ainsi que les exploitations et les modalités de conservation et de purge qui leur sont associées. Les ressources sont également soumises, avant leur intégration dans le GAR, à une série de tests visant à vérifier leur conformité quant aux exigences liées au RGPD et au code de l'éducation (par exemple l'absence de publicité).

Les discours présentent les avantages associés à la mise en œuvre du GAR : son côté simplificateur et normalisant, la garantie de conformité qu'il assure et l'accès aux données « élèves » et « enseignants » qu'il permet.

D_FR1 : alors si on veut... d'abord rester global sur le sujet j'aurais tendance à dire que, le cadre du GAR nous permet, d'avoir des données personnelles élèves, parce que sinon, on ne peut pas, donc euh d'une certaine manière voilà avant le GAR, il n'était pas question de faire des comptes élèves avec des données personnelles, donc ça nous a permis d'avancer on va dire dans le respect de la légalité, versus nos besoins fonctionnels,

D'autre part, les discours positifs parfois même enthousiastes sont souvent suivi par un discours critique associé à l'impact technique fort qu'aurait le GAR sur les systèmes d'information des fournisseurs de ressources. La remise en cause d'un certain nombre de fonctionnalités⁶ existantes sur les plateformes et la nécessité de revoir le modèle de données sont les deux principaux éléments évoqués par les acteurs interrogés. Les développements sont jugés coûteux et difficilement gérables sur le plan organisationnel quelle que soit la taille de l'entité, économiquement difficiles à assumer pour les petites entités.

PDG_FSI : On a eu un échange avec eux (le ministère) un peu compliqué qui nous a obligé à redévelopper pas mal de choses. Nous, en fait, on gérait les groupes, enfin on gère les groupes parce que les gens qui ne font pas partie de l'ENT, quand ils arrivent, automatiquement créent leurs groupes « Chef d'œuvre de classe » puisque la problématique sur un « Chef d'œuvre » c'est que ça ressemble des professeurs d'inter-disciplines et en plus sur deux ans donc c'est assez spécifique « ... ». Le GAR nous a forcés à enlever cette fonctionnalité-là, nous disant que cela posait un problème parce qu'il y'avait déjà des groupes qui étaient créés dans l'ENT et que du coup, on faisait la même chose qu'un ENT et que ce n'était pas bien. On a enlevé les groupes et je vais être honnête avec vous, c'est un peu la cata « ... ». On a été très surpris. Ça a fait un gros boulot chez nous parce qu'on a dû casser cette chose-là.

PDG_FR1 « Le GAR a fait des évolutions fonctionnelles et nous demande plus qu'avant : par exemple la gestion multi-établissements... C'est un changement qui implique de revoir toute l'archi du produit. C'est techniquement très coûteux en temps et en risque. On ne maîtrise pas les délais. Une vraie incertitude économique ces évolutions. »

D_FR2 : la solution telle qu'elle a été pensée ne peut pas entrer en l'état dans le GAR. Ca veut dire qu'il faut qu'on repense une partie de notre système d'information, et toute une partie de l'ergonomie de l'application, et ça c'est à peu près, moi je l'évalue entre 9 et 12 mois de travail, donc ça veut dire que dans le meilleur des cas, XXX d'aujourd'hui sera, accessible en France en septembre 2022, mais ce qui faut bien voir c'est qu'en septembre 2022, XXX d'aujourd'hui n'aura pêtre plus grand chose à voir parce que effectivement XXX évolue on arrête pas de le faire évoluer, voilà on prend en compte les retours utilisateurs, et donc on est sur un décalage.

3.4.2. À propos d'intelligence artificielle : des réalisations embryonnaires

Même si la majorité des acteurs interrogés ou analysés préfèrent parler d'algorithmique avancée, quelques-uns évoquent l'utilisation de techniques d'intelligence artificielle.

Cependant, pour tous, à l'exception d'un acteur bien implanté sur le secteur de la formation professionnelle en entreprise, il s'agit essentiellement de projets en phase de recherche et développement dont l'industrialisation n'est pas encore effective.

⁶ C'est notamment le cas de la fonctionnalité de publication en dehors de l'établissement pour laquelle le GAR impose le respect du cadre établi par le code de l'éducation, à savoir l'obtention de l'accord du chef d'établissement en sa qualité de responsable éditorial à l'échelle de l'établissement.

Certains déclarent avoir eu des réflexions et effectué des tests par le passé, mais les projets de développement ont avorté, à cause du manque de fiabilité des calculs permettant à la machine de définir des profils d'apprentissage et d'automatiser l'attribution de contenus dédiés à la remédiation :

PDG_FP2 : donc on a essayé, y'a quelques années de croiser les données des apprenants, donc avec les techniques de deep learning, pour essayer de voir si on pouvait en tirer des choses, qu'est-ce qu'on peut en tirer en fait, on pourrait en tirer que... j'dis n'importe quoi mais si une personne maîtrise cette règle, cette règle, cette règle, alors elle a 98% de chances de maîtriser aussi celle-là, et c'est ça qu'on... qui pourrait être intéressant, on a fait tourner les choses, et ça, ça n'a rien donné, ça n'a rien donné d'intéressant, donc voilà on a pas réussi à croiser donc voilà on est pas sur le deep learning, pour ces raisons-là, peut-être qu'on y reviendra, j pense que ce serait intéressant peut-être avec une approche différente, pour essayer de voir comment tirer parti de cette masse d'informations.

Par ailleurs, avoir recours aux techniques d'intelligence artificielle est perçu par la majorité des acteurs comme un coût conséquent au regard de son marché actuel et de la masse de données disponible :

PDG_FR2 : on sait tous que pour que ça marche, il faut des millions de données, faut que vraiment il y ait de la masse, et là alors c'est le paradoxe purement français où ne permettant pas qu'il y ait un marché digne de ce nom, en France, à part certains élus d'appels d'offre, c'est hyper restreint,

La question de la place de l'enseignant est centrale. Ainsi, pour certaines entreprises, l'investissement dans les techniques d'Intelligence artificielle entre en contradiction avec leur positionnement initial qui est celui de la production de ressources pour les enseignants :

PDG_FR2 : parce qu'on est pas dans un truc à se faire racheter euh, faire de l'intelligence artificielle, changer le monde, mais à notre manière maintenant on est quand même des entrepreneurs, ça fait 15 ans qu'on investit on a des employés et tout mais, c'était vraiment des outils pour... aider ; j'étais enseignant hein j'utilisais mes propres outils j'ai vu, dans le regard de mes élèves à quel point c'était, ça changeait, et je me suis dit c'est ça qu'il faut, et puis le ministère a dit moi j'aide pas les profs j'aide les entreprises, à juste titre hein c'était parce que ils s'en sortiraient pas d'aider les profs qui ont des délires dans tous les coins, les profs c'est très intelligent, ça a des délires qui peuvent être très sympas mais si il (le ministère en charge de l'Education nationale) commence à financer ça, il a plus personne dans les classes.

D_FR2 : ouais peut-être dire que notre cœur de métier c'est la création de ressources, nous on est sur de la création de ressources pédagogiques, on est pas une plate-forme de services, on est pas, un ENT, enfin voilà, nous notre cœur de métier c'est créer des ressources pédagogiques, et en premier lieu en fait ces ressources étaient destinées aux enseignants, pour pouvoir être utilisées en situation de cours, en classe ou à travers des dispositifs de e-learning, mais, voilà, notre cœur de métier c'est de créer des ressources, et voilà des ressources qui soient interactives, qui soient visuelles, mais voilà notre cœur de métier c'est ça.

On peut distinguer deux trajectoires pour les entreprises : celles qui vont des contenus vers les services (typiquement les éditeurs scolaires), et celles qui font le trajet inverse, c'est-à-dire celles qui ont des plates-formes de services (comme Tactiléo ou Enseigno) et sont obligées d'aller vers les contenus. Beaucoup cherchent à s'adapter notamment quand il s'agit de se positionner sur les appels d'offre du Ministère et tentent de développer des solutions répondant aux exigences, à savoir des solutions intégrant des techniques d'intelligence artificielle mais axées sur deux volets : un volet de parcours autonomes pour les élèves et un volet enseignant lui permettant de conserver son rôle de médiateur des apprentissages grâce à la modification des parcours recommandés

4. Discussion et perspectives

Ce rapport vise à rendre compte des tendances concernant la collecte et l'exploitation des données d'apprentissage par différentes entités spécialistes de la production de ressources ou de supports de ressources.

Les analyses que nous venons d'exposer nous conduisent à reprendre deux oppositions classiques.

La première concerne l'apprentissage personnalisé. Si, selon Justin Reich⁷, on s'accorde sur son importance et sur le fait qu'il est rendu possible par les nouvelles technologies, ce qu'il recouvre diffère : (1) diagnostiquer individuellement les compétences des élèves et appliquer des algorithmes pour fournir de manière adaptative un contenu stimulant approprié à chacun ou (2) donner à chaque élève des possibilités d'exploration et de création et lui permettre de suivre ses intérêts et ses passions dans diverses directions. D'un côté, adopter un modèle industriel de l'éducation et utiliser la technologie pour aider chaque enfant à maîtriser le contenu académique de la manière la plus efficace possible et, de l'autre, disposer d'une technologie plus en arrière-plan, disponible pour ouvrir l'école sur le monde (Herold, 2021).

Dans notre échantillon, la personnalisation reste assez limitée et les réalisations les plus intéressantes sont plus liées à la conduite de projets ou à l'orientation, qu'à la maîtrise de contenus académiques de la manière la plus efficace possible (Herold, 2021). Par ailleurs, nous n'avons pas inclus Google et Microsoft (ou Facebook) dans notre échantillon. Ils n'auraient sans doute pas répondu à nos questions⁸. Ils n'ont pas de ressources diffusées via les ENT et échappent au GAR. Leur respect des contraintes liées au service public d'éducation (par exemple, l'absence de publicité, que les données soient stockées en Europe, de même que les services proposés) reste en question. Des modèles d'apprentissage adaptatifs sont développés aux États-Unis, notamment le programme Summit poussé par Facebook, mais il est très critiqué (Voir Boninger et al., 2020).

La seconde opposition, reliée à la première, concerne le rôle dévolu aux enseignants. Sont-ils considérés comme les organisateurs et décideurs locaux de l'enseignement et des

7

https://blogs.edweek.org/edweek/edtechresearcher/2012/06/battling_over_the_meaning_of_personalization.html

8 Dans l'enquête menée par Xavier Levoine (2018), plusieurs rendez-vous avec un représentant de Microsoft éducation (France) ont été pris puis annulés.

activités menées par les élèves ou de simples exécutants de choix éducatifs effectués hors de leur contrôle, notamment par un système partiellement automatisé.

Le rapport Villani (déjà cité⁹) sur l'intelligence artificielle défend clairement la première approche. Les élèves ne sont jamais complètement autonomes dans leur activité, les producteurs de ressources mettent à disposition des enseignants ou des parents des tableaux de bord qui permettent de suivre les productions et les progressions des élèves et parfois, de choisir les tâches à faire pour les élèves.

Comme nous en faisons l'hypothèse au début de cette enquête, les différentes entités investiguées, quelle que soit leur taille ne collectent et ne traitent pas de masses importantes de données d'apprentissage. Dans le cadre scolaire, les données d'identité *étendues*, pour la plupart gérées au sein des établissements scolaires, sont largement suffisantes : classe/ groupe. Elles permettent d'avoir les normes attendues (référentiel scolaire). Ensuite, il s'agit de déterminer l'écart par rapport à cette norme (via des enquêtes ou des exercices spécifiques), puis mettre en place un système dit de « remédiation » qui se donne pour objectif de réduire ces écarts. On constitue une sorte de modèle (structuré par des compétences) et on choisit ce que l'on donne à faire dans l'ensemble des ressources que l'on a « en magasin ». Les résultats obtenus et le temps passé semblent suffisants pour assurer une telle gestion. On joue sur l'année scolaire, inutile de stocker des données plus longtemps (c'est aussi une contrainte RGPD, pour permettre la mise en œuvre du « droit à l'oubli »). En gros, cela correspond à ce que font bon nombre d'enseignants alternant évaluations formatives et sommatives, afin d'orienter les activités qu'ils vont faire faire aux élèves.

Certains suggèrent qu'un recours plus important à l'exploitation des données nécessiterait un changement des règles traditionnelles du métier, ce qui pose des questions délicates et ne peut s'accomplir que dans la durée. Nos interlocuteurs se concentrent logiquement sur des segments où ils ont une expérience notable, en travaillant selon un principe de parcimonie dans l'utilisation à bon escient de données d'apprentissage. Ce type de choix pragmatique est sans doute d'ailleurs le plus raisonnable dans les conditions actuelles.

Quant aux techniques d'intelligence artificielle, elles ne semblent pas encore susceptibles de produire, à court terme, des résultats probants dans des contextes ordinaires en France. L'effort nécessaire pour produire des systèmes innovants y est très important, sans certitude de retour rapide sur investissement.

Par ailleurs, les exploitations décrites s'avèrent être le reflet des appels d'offres du ministère en charge de l'Éducation Nationale qui ont un impact fort sur les orientations des différentes entités.

Suite à cette étude, on peut proposer deux recommandations, privilégiant l'aide à l'enseignant plutôt que des formes plus automatisées d'exploitation des données.

La première est de former les enseignants à la gestion et à l'utilisation de données d'apprentissage. C'est au cœur de leur métier et la technologie étend les possibilités de recueil et de traitement, mais il faut être bien formé pour en tirer parti (voir Boissière et

9 « 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. »

Bruillard, 2021). S'agissant des tableaux de bord, ils ne sont faciles ni à concevoir ni à interpréter et leur alimentation n'est pas totalement automatique.

Une seconde préconisation serait que les établissements scolaires puissent disposer d'une personne chargée de la gestion des données scolaires et d'apprentissage : de leur recueil, de leur structuration, d'une partie de leur analyse et de l'alimentation des tableaux de bord pour les enseignants. Si les enseignants ne bénéficient pas d'un soutien humain et technique proche, ils auront des difficultés à utiliser les données d'apprentissage.

Enfin, pour terminer, la gestion des données d'apprentissage n'est pas une nécessité qui s'imposerait à tous. Diversifier les activités, multiplier les types de ressources et de supports de ressources sont tout aussi importants, voire plus. Ainsi, le groupe Hachette semble très actif, pas seulement dans ce qui est lié au numérique : acquisition d'une participation majoritaire dans la société britannique d'applications mobiles Brainbow LTD, éditeur de Peak (2016)¹⁰, prise de contrôle du Livre Scolaire¹¹, lien avec Kwyk¹² pour les mathématiques, rachat d'éditeurs de jeux comme Gigamic¹³, la société Brainbow¹⁴, de Sorry We Are French¹⁵, une société française d'édition de jeux de société, comme l'éditeur Hiboutatillus¹⁶, les maisons d'édition La Plage¹⁷ et Kero¹⁸... et on en oublie certainement. On ne sait pas quelle synergie pourra se construire autour de ces différents apports.

A la mise à disposition de contenus, activité historiquement assurée par les éditeurs scolaires, s'ajoutent des offres de services aux personnes et aux établissements. Comment vont s'articuler ces deux activités ? Quelles entités autonomes ou quels consortiums vont les assurer ? Quelles évolutions du métier d'enseignant vont s'enclencher ? Nombre de questions restent à étudier.

5. Références

5.1. Références citées dans le rapport

Boissière, Joël et Bruillard, Éric (2021). *L'école digitale. Une éducation à construire et à vivre*. Collection Sociologia. Armand Colin, 376 p.

Boninger, F., Molnar, A., et Saldaña, C. (2020). *Big Claims, Little Evidence, Lots of Money: The Reality Behind the Summit Learning Program and the Push to Adopt Digital Personalized Learning Platforms*.
<https://nepc.colorado.edu/publication/summit-2020>

10 https://www.hachette.com/fr/communique_presse/le-groupe-hachette-livre-a-acquis-une-participation-majoritaire-dans-la-societe-britannique-dapplications-mobiles-brainbow-ltd-editeur-de-peak/

11 <https://www.livreshebdo.fr/article/hachette-reprend-lelivrescolairefr>

12 <https://www.enseignants.hachette-education.com/collections/kwyk>

13 <https://www.lsa-conso.fr/hachette-livre-rachete-l-editeur-de-jeux-gigamic,309336>

14 <https://www.idboox.com/economie-du-livre/hachette-livre-achete-la-societe-dapplications-brainbow/>

15 https://www.fusacq.com/buzz/hachette-livre-rachete-sorry-we-are-french-a203705_fr

16 <https://www.lsa-conso.fr/hachette-livre-rachete-hiboutatillus-createur-du-jeu-blanc-manger-coco,384908>

17 <https://www.booksquad.fr/hachette-livre-rachete-la-plage/>

18 <https://www.livreshebdo.fr/article/hachette-rachete-kero>

- Borgman, Christine L. (2020). *Qu'est-ce que le travail scientifique des données ? Big data, little data, no data*. Nouvelle édition [en ligne]. Marseille : OpenEdition Press, 2020 (généré le 26 août 2021). <http://books.openedition.org/oep/14692> (Traduction de Charlotte Matoussowsky)
- Galonnier Juliette, Le Courant Stefan, Pecqueux Anthony et Noûs Camille (2020). Ouvrir les données de la recherche ? », *Tracés. Revue de Sciences humaines* [En ligne], <http://journals.openedition.org/traces/10588>
- Ghabara, Khansa (2020). Données scolaires élèves : étude de l'existant et des possibilités de recueil et d'analyses. *Actes des 8^e RJC EIAH 2020, Environnements Informatiques pour l'Apprentissage Humain*. p.48-54. <https://hal.archives-ouvertes.fr/hal-03208548/document>
- Herold Benjamin (2021). Personalized Learning's Big Test Is Coming This School Year. *Edweek*, July 20, 2021. https://www.edweek.org/technology/personalized-learning-big-test-is-coming-this-school-year/2021/07?utm_source=nl&utm_medium=eml&utm_campaign=eu&M=62016396&U=1473110&UUID=509727908dbf7fe60f8ba2c9674e4b70#promo
- Levain Xavier (2018). *Mutations dans le paysage des ressources numériques à la lumière des BRNE. Synthèse de l'enquête BRNE 2018*. ÉDA, Université Paris Descartes sous la direction d'Éric Bruillard, novembre 2018. http://eda.recherche.parisdescartes.fr/wp-content/uploads/sites/6/2019/03/rapport_2018_v3.pdf
- Merieux, J-M (2019). *Développement du numérique éducatif / Ministère de l'éducation nationale et de la jeunesse*. Publication de la DEPP. https://archives-statistiques-depp.education.gouv.fr/Default/doc/SYRACUSE/45582/developpement-du-numerique-educatif-ministere-de-l-education-nationale-et-de-la-jeunesse?_lg=fr-FR
- MENJ (2019). *Bilan CNIL-GAR après deux ans de déploiement. Principes et orientations de la protection des données à caractère personnel*. En ligne, novembre 2019 https://gar.education.fr/medias/fichier/gar-a4-36pages-20191115_1574023611376-pdf
- Ministère de l'éducation nationale, de la jeunesse et des sports (2018, août). *Le numérique au service de l'école de la confiance*. <https://www.education.gouv.fr/le-numerique-au-service-de-l-ecole-de-la-confiance-308365>
- Reich, J. (2012). *Battling over the Meaning of « Personalization »*. Education Week—EdTech Researcher website: http://blogs.edweek.org/edweek/edtechresearcher/2012/06/battling_over_the_meaning_of_personalization.html?cmp=SOC-SHR-FB

Annexes

1. Texte diffusé aux fournisseurs de ressources

Collecter et exploiter les données d'apprentissage

Par Khansa Ghabara, le 29/03/2021

Contexte

La numérisation de plus en plus déployée au sein de l'espace social en général et à l'École en particulier introduit des dispositifs dans lesquels des données en grande quantité circulent. Dans le secteur de l'éducation, une offre diversifiée de logiciels et de plateformes permet de proposer un ensemble d'activités d'apprentissage¹.

Cette offre intègre de plus en plus souvent des fonctionnalités d'individualisation et de suivi. Les fonctions de conservation des personnalisations, de production de documents utilisant ou pas les contenus proposés par la ressource, de soumission de contenus, d'assignations de tâches occupent une place croissante, quelquefois accompagnées par des fonctions de collecte, conservations, exploitation et analyse des traces liées aux actions des élèves.

Ainsi, l'enseignant peut fournir des prescriptions individualisées à ses élèves ; il est possible de conserver le temps passé sur une activité, le score obtenu, le succès ou l'échec, le degré d'achèvement, le nombre d'essais effectués etc. ; on peut faciliter et structurer les échanges entre enseignant et élèves, voire les échanges entre pairs.

La disponibilité de ces données, la possibilité de les exploiter, d'en collecter et d'en produire, et in fine de produire des stratégies d'analyse ou d'aide à la décision devrait permettre aux fournisseurs de services éducatifs d'améliorer leur offre. Au cœur de ces possibilités, le vieux rêve de la personnalisation de l'enseignement et des apprentissages, notamment via les parcours, devient une réalité accessible².

Mais cette personnalisation s'appuie sur des activités de collecte, de production, d'analyse, etc. de ces corpus de données, par essence personnelles, sensibles et impliquant des faisceaux complexes de responsabilités. Le développement de telles fonctionnalités au sein des ressources ne peut se contenter de la juxtaposition et la stratification de fonctions nouvelles, mais devrait s'appuyer sur une forme de politique d'usage des données.

Les raisons d'une étude exploratoire sur la collecte et le traitement des données d'apprentissage

C'est bien cette politique d'usage des données que l'ARIFE souhaite ouvrir aux échanges et à la réflexion collective, alors même qu'elle constitue pour chaque éditeur l'expression de sa stratégie pour les années à venir.

L'association a donc décidé de lancer une étude exploratoire pour faire le point sur les

activités et les projets autour de la gestion des données d'apprentissage, auprès des acteurs principaux du numérique éducatif en France. Elle a pour principal objectif d'obtenir un premier état des lieux, via l'interrogation des éditeurs scolaires et des entreprises EdTech en éducation sur leurs activités et objectifs en termes d'exploitation des données d'apprentissage (types de données recueillies, types d'analyses, utilisation de normes ou de standards, obstacles à lever, partenariats, etc.).

Une analyse des discours accompagnant l'offre des fournisseurs de service éducatifs va donc être entreprise. Une campagne d'entretiens va être également lancée auprès d'un échantillon diversifié de producteurs de ressources éducatives, éditeurs scolaires traditionnels et nouvelles entreprises apparues avec la généralisation d'Internet.

L'étude devrait se dérouler durant trois mois, à partir de début avril 2021 et donner lieu à un rapport qui sera publié pendant l'été 2021. Elle sera menée par Khansa Ghabara et Solène Zablouk sous la direction d'Éric Bruillard, directeur du laboratoire EDA (université de Paris) et vice-président de l'ARIFE.

À l'issue de ces premiers travaux qui se concentrent sur le contexte français, un prolongement par une étude internationale est envisagé.



Association Recherche et Innovation pour les Politiques d'Éducation et de Formation. (<http://aripef.org/>)

1 Levoine Xavier (2018). Mutations dans le paysage des ressources numériques à la lumière des BRNE. Synthèse de l'enquête BRNE 2018. ÉDA, Université Paris Descartes sous la direction d'Éric Bruillard, novembre 2018.

2 <https://www.educlever.com/ambitions>, <https://www.ac-bordeaux.fr/lalilo-personnalisation-de-l-apprentissage-de-la-lecture-122236>, <https://thinkovery.com/blog/3-questions-evan-friburg-domoscio/>

2. Grille d'entretien

Les questions suivantes constituent la grille pour un entretien mené auprès d'un fournisseur de plate-forme.

Présentation

1- En tant qu'entreprise spécialisée du numérique en éducation, comment vous définissez-vous ?

2- Pouvez-vous nous raconter les origines de votre projet, au moment où vous avez voulu créer votre entreprise ? Quels étaient vos objectifs ?

3- En termes de propositions de services pour l'enseignement, on voit que vous proposez de plus en plus de contenus en dehors de ceux initialement destinés au collègue ? Comment vos services ont-ils évolué ? Comment souhaiteriez-vous vous développer à l'avenir ?

Utilisation des données

4- Quelles sont les données de l'utilisateur récupérées lors de la connexion sur XXX (nom de la plate-forme créée par le fournisseur) ? A quoi servent ces données ?

5- Sur plusieurs banques de ressources accessibles via XXX (nom de la plate-forme), comme XXX ou XXX (exemples de services), nous avons constaté que vous proposez des parcours d'apprentissage. Comment ces parcours sont-ils définis ? Avez-vous besoin des données d'apprentissage pour les définir ?

6- Comment vous positionnez-vous face à des entreprises qui proposent des solutions basées sur l'utilisation de l'intelligence artificielle ? Diriez-vous que vous proposez des services ayant recours à l'IA ?

7- Sur certaines banques de ressources comme celles développées en partenariat avec XXX (éditeur scolaire), vous expliquez que les enseignants ont la possibilité de produire leurs propres contenus ? Jusqu'où peuvent-ils personnaliser les contenus montrés aux élèves ? Pouvez-vous nous montrer un exemple ?

8- Est-ce que les ressources collectent des données d'usage (données de navigation) ? Comment sont-elles utilisées ?

9- Quelles contraintes rencontrez-vous dans l'exploitation des flux de données scolaires des élèves, notamment celles obtenues via les ENT ? Qu'a changé pour vous, le nouveau flux de données scolaires issu le GAR ?

10- Quels normes et standards utilisez-vous pour vos développements ? Quel impact de ces normes & standards sur le recueil et l'exploitation des données au sein de vos plateformes ?

11 - Dans quel but précis utilisez-vous les données issues des outils d'analyse web comme Google Analytics sur XXX (nom de la plate-forme) ? Quel est l'apport de cet outil pour le développement de services individualisés au sein de vos plates-formes ?

3. Sites étudiés

- 3 éditeurs scolaires ;
- 3 collectifs enseignants et 2 enseignants producteurs de ressources ;
- 3 producteurs de banques de ressources, qui sont également producteurs de plates-formes pour l'apprentissage des élèves ;
- 6 producteurs de plates-formes et 1 fournisseur de solutions techniques pour le développement des plates-formes ;

- 3 fournisseurs de services qui n'ont pas de lien avec un enseignement disciplinaire. Il peut s'agir de services permettant aux enseignants de concevoir leurs cours ou de services permettant un appui à la conduite de projets dans le cadre scolaire ou de l'enseignement supérieur (enseignement professionnel, IUT).

4. Structures interrogées

- 1 éditeur scolaire ;
- 1 collectif d'enseignants
- 1 enseignant producteur de manuels scolaires indépendant ;
- 2 producteurs de ressources qui proposent également une plate-forme ;
- 2 entreprises qui proposent des services qui n'ont pas de liens avec un enseignement disciplinaire ;
- 2 producteurs de plateformes
- 1 fournisseur de services techniques pour les producteurs de plates-formes ;
- Dont 8 structures ayant un verbatim dans le texte.

5. Encodage des données

Pour rendre compte des discours tenus par les entités soit à travers leurs sites web, soit lors des entretiens, nous avons opté pour l'encodage suivant :

À titre d'exemple, l'encodage PDG_FR2 correspond à un discours tenu par le concepteur ou le gérant de la deuxième entité spécialisée dans la fourniture de ressources et D_FP1 correspond au discours tenu par la personne chargée de la protection des données pour première des entités spécialisées dans la fourniture de plates-formes.