

Education Data Hub en France

ETUDE DE CAS

DIRECTION DU NUMERIQUE POUR L'EDUCATION -
MINISTÈRE DE L'EDUCATION NATIONALE - FRANCE

Remerciements à Federica Minichiello (Chef de projet DNE - MENESR), Philippe Ajuelos (Administrateur ministériel des données, algorithmes et codes sources - MENESR) et Hugues Labarthe (DAN Adjoint - DRANE - Bretagne) pour leur aide dans la rédaction de cette étude de cas consacrée au projet de l'Education Data Hub.

Auteur: Isabelle Dufrêne (DNE - MENESR)

This case study is part of Agile EDU's Deliverable D4.1. It is the final version further developed and elaborated based upon comments on the first and second versions (D3.3, due in M29).



Co-funded by the
Erasmus+ Programme
of the European Union

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Contents

Résumé	3
Introduction	4
Contexte	6
Le cadre d'analyse	9
Méthode	10
Analyse	11
Gouvernance des données	11
Droits, réglementations, protection de la vie privée	15
Utilisation des données pour l'enseignement et l'apprentissage	17
Des projets qui répondent aux besoins	19
Gouvernance des données	19
Littératie des données	21
Conclusion et recommandations	24
Conclusion	24
Recommandations	25
Annexes	26
Acteur de la gouvernance ministérielle des données - Philippe Ajuelos	26
Chef de projet - Federica Minichiello	26
Acteur institutionnel territorial - Hugues Labarthe	27
Concevoir un Education Data Hub	28
Préfiguration	28
Étude de préfiguration 1 - avril à juillet 2020 - DNE	28
Étude de préfiguration 2 - mai à juillet 2021 - DNE and INRIA	29
Auditions - novembre 2022 à juin 2023	29
Références	30

Résumé

En France, la politique éducative relève de la responsabilité de l'État, qui définit les lois encadrant l'organisation scolaire, les programmes, le recrutement des enseignants et le financement. Toutefois, ces compétences sont partagées avec les collectivités territoriales, et de nombreux acteurs différents produisent et collectent des données.

Afin de permettre une utilisation plus efficace des données issues des activités d'enseignement et d'apprentissage, le ministère de l'Éducation nationale a élaboré une feuille de route pour sa politique de la donnée, dont l'un des volets consistait en la création d'une plateforme des données de l'éducation (Education Data Hub).

Au cœur du concept d'Education Data Hub se trouvait l'objectif d'améliorer la capacité à exploiter l'ensemble des données disponibles. Ces données, produites dans le cadre de la mission de service public de l'éducation, par des entités publiques comme privées, pouvaient être mobilisées pour l'évaluation des effets des politiques publiques, l'amélioration des services existants ou la création de nouveaux services.

Bien que ce projet n'ait finalement pas été mis en œuvre, la phase de préfiguration a permis de mettre en évidence des besoins auxquels des projets de plus petite échelle apportent aujourd'hui des réponses concrètes.

Introduction

Le ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche s'appuie sur différents échelons locaux pour mettre en œuvre sa politique. L'administration du ministère est présente dans chaque région et chaque département : il s'agit des services déconcentrés de l'Éducation nationale. Elle est organisée en 18 régions académiques, 30 académies et 97 directions des services départementaux de l'Éducation nationale (DSDEN).

L'instruction est obligatoire pour les enfants de 3 à 16 ans. Les écoles maternelles accueillent les enfants de 3 à 6 ans, les écoles élémentaires les enfants de 6 à 11 ans, les collèges les élèves de 11 à 15 ans et les lycées ceux de 15 à 18 ans.

Les collectivités territoriales jouent un rôle de plus en plus important dans la gouvernance du système éducatif, en assurant son fonctionnement matériel (construction et entretien des bâtiments scolaires, transports scolaires, fourniture de matériel pédagogique, etc.). Au niveau local, la gestion administrative et matérielle du système éducatif est assurée par : les communes pour les écoles primaires, les départements pour les collèges et les régions pour les lycées.

Les établissements du second degré disposent quant à eux d'une certaine autonomie pour définir des stratégies permettant d'atteindre les objectifs fixés par la politique nationale, dans le cadre du projet d'établissement.

Les environnements numériques de travail (ENT) offrent un point d'entrée unique aux utilisateurs pour accéder aux services et contenus numériques, en fonction de leur profil et de leur niveau d'autorisation. Ils constituent également un espace d'échange et de collaboration entre utilisateurs, ainsi qu'avec d'autres communautés liées à l'école ou à l'établissement.

L'ensemble de ces acteurs collecte des données pouvant être utiles aux enseignants, aux établissements, aux chercheurs, à l'administration et aux entreprises de la EdTech.

Les objectifs de l'Education Data Hub sont multiples et s'adressent à divers acteurs du domaine éducatif, même si l'accent est mis principalement sur la valorisation des données numériques scolaires : collecte, analyse et traitement des données scolaires et des traces d'apprentissage.

Au sein du système éducatif, l'accessibilité et les capacités de traitement de ces données accompagnent la transformation de l'école :

- Au bénéfice de l'élève, en offrant un environnement d'apprentissage personnalisé : évaluation précise de ses points forts et de ses besoins, parcours pédagogiques adaptés, aides ciblées, suggestions d'activités ou de ressources.
- Au bénéfice de l'enseignant, en lui permettant de consacrer davantage de temps à la différenciation pédagogique et en lui fournissant une vision objective des besoins spécifiques de chaque élève.

- Au bénéfice du système éducatif, en exploitant le potentiel de l'analyse massive de données et de l'intelligence artificielle pour mieux évaluer les pratiques et modéliser les évolutions futures des établissements et des différents niveaux d'organisation de l'institution.

Au sein des organismes de recherche, les chercheurs en éducation pourraient mieux comprendre les mécanismes d'apprentissage, ce qui permettrait une évolution éclairée des pratiques. Les données numériques issues des usages numériques des élèves, des enseignants et de l'administration (via les environnements numériques de travail, les solutions de vie scolaire, les ressources éditoriales, les plateformes d'apprentissage, etc.) pourraient être collectées et agrégées afin d'être analysées et exploitées. Cela permettrait de mieux comprendre les processus d'apprentissage, d'aider les enseignants à proposer des ressources et des activités adaptées à chaque élève, ou encore de contribuer à la gestion des établissements et du système éducatif.

Pour les entreprises de la EdTech, ces données offriraient une meilleure compréhension des besoins.

Ces données, qui ne relèvent ni de la gestion (informations administratives, emplois du temps, utilisation des logiciels scolaires, parcours éducatifs), ni de la statistique (produites par la Direction de l'évaluation, de la prospective et de la performance – DEPP – et les services statistiques académiques), sont produites massivement et en continu, principalement par des entreprises privées. Il s'agit de traces d'apprentissage : des données produites par des instruments physiques, des logiciels ou des personnes dans le cadre d'un processus d'apprentissage, et qui en documentent le déroulement : données issues de capteurs, enregistrements d'interactions, ressources consultées et temps passé à les consulter, réponses données aux questions et temps consacré, productions, annotations d'observateurs ; productions d'élèves ou d'enseignants : écrits d'élèves, supports d'activités, ressources pédagogiques conçues par les enseignants.

Les orientations récentes en matière de politique publique insistent sur la nécessité de collecter et d'agréger ces données massives afin d'en faciliter l'usage et de favoriser le développement de nouveaux outils, notamment en lien avec l'intelligence artificielle (stratégie nationale pour l'IA, rapport Bothorel, stratégie européenne pour les données).

Le ministère de l'Éducation nationale, de la Jeunesse et des Sports (MENJS), en partenariat avec l'INRIA et avec le soutien financier du 4e Programme d'investissements d'avenir (PIA 4), dans le cadre de la stratégie d'accélération « Numérique et Éducation », avait décidé de créer une plateforme de données scolaires. Celle-ci visait notamment à mettre les données numériques à disposition de la recherche ainsi que des entreprises EdTech, et à favoriser leur circulation.

Cette plateforme devait permettre la collecte, l'analyse et l'exploitation de données numériques massives (anonymisées ou pseudonymisées) issues des usages et pratiques numériques des élèves, des enseignants et des établissements (traces numériques). Ce projet concernait les chercheurs, afin d'approfondir leur compréhension des mécanismes d'apprentissage et du rôle des environnements scolaires, ainsi que les entreprises développant des solutions innovantes, afin d'enrichir leurs données, développer et tester des algorithmes d'IA, partager leurs données ou évaluer leurs innovations.

Contexte

Le projet de plateforme de données d'éducation s'inscrit dans un contexte plus large d'ouverture des données éducatives et de partage des données issues de l'éducation. Depuis 2011, le ministère de l'Éducation nationale, de la Jeunesse et des Sports (MENJS) contribue activement à la publication des données scolaires sur la plateforme gouvernementale dédiée.

À la suite de la promulgation de la loi pour une République numérique en 2016, qui établit le principe d'ouverture par défaut des données publiques, le ministère a engagé une démarche proactive au sein de l'ensemble de ses services d'administration centrale. Fin 2016, une plateforme ministérielle *Open Data* a été créée, synchronisée avec la plateforme gouvernementale et reliée à celle de l'enseignement supérieur et de la recherche, afin de faciliter la mise à disposition des données et de valoriser les jeux de données les plus pertinents.

C'est dans cette continuité qu'a été pensée la plateforme de données d'éducation (Education Data Hub). Destinée aux chercheurs, aux acteurs de la EdTech et aux parties prenantes de l'éducation nationale, cette plateforme visait à permettre à la communauté éducative dans son ensemble, ainsi qu'à ses partenaires, de construire un écosystème cohérent, guidé par des standards éthiques communs, autour d'un catalogue partagé de données et d'une bibliothèque ouverte d'algorithmes. Elle est conforme au Règlement général sur la protection des données (RGPD) et s'inscrit dans la stratégie nationale pour l'intelligence artificielle, en vue de produire des ressources innovantes.

Comme dans de nombreux pays, la pandémie a conduit le gouvernement à interroger les usages numériques, tant sur les plans techniques que citoyens. Les enjeux liés au partage des données éducatives ont été abordés selon deux axes : l'aspect éthique, au sein du Comité d'éthique des données, et l'aspect prospectif, à travers diverses recommandations, dont le livre blanc publié par l'Institut national de recherche en sciences et technologies du numérique (INRIA).

Le ministère de l'Éducation nationale et de la Jeunesse (MENJ) a contribué à cette réflexion par le biais d'études, de formations et d'une large concertation organisée en novembre 2020 : les États généraux du numérique pour l'éducation.

Cette concertation s'est déroulée en plusieurs phases :

- Consultation des différentes parties prenantes : usagers du service public numérique de l'éducation et acteurs de l'écosystème.
- Retour d'expériences de terrain : 1 300 initiatives ont été sélectionnées à l'échelle nationale pour leur potentiel de généralisation. Un corpus d'études et de recherches a été analysé, des entretiens et des groupes de discussion ont été menés avec des échantillons représentatifs des différents types d'usagers, et des données ont été compilées, collectées à la fois par le ministère et par les entreprises du secteur EdTech.
- Mise en œuvre de la consultation via une plateforme participative accessible à tous les citoyens (670 propositions), organisation d'États généraux du numérique dans différentes académies (21

documents de synthèse), et réalisation d'une enquête auprès de divers usagers : élèves, parents, enseignants, personnels éducatifs et agents (26 839 répondants).

- Organisation de deux journées de débats et d'ateliers, au cours desquelles les propositions issues de la concertation ont été discutées, approfondies et formalisées

40 propositions ont émergé de cette concertation, regroupées en cinq grands axes : Mettre en place de nouvelles formes de gouvernance et de nouveaux outils d'anticipation, Enseigner et apprendre le numérique et avec le numérique, Garantir un accès égal au numérique pour tous / Lutter contre la fracture numérique, Travailler ensemble autrement et promouvoir une culture professionnelle numérique commune. Favoriser le développement d'un numérique responsable et souverain. C'est dans ce dernier axe que s'inscrit la proposition n°37 : créer une plateforme de données d'éducation l'Éducation Data Hub.

La collecte, l'ouverture et l'exploitation systématique des données sont au cœur de la transformation de l'action publique. Leur circulation contribue à renforcer l'efficacité des politiques publiques et constitue également un levier d'innovation pour l'ensemble de la société.

En 2021, sous l'autorité du Premier ministre, chaque ministère a élaboré, avec le soutien de la Direction interministérielle du numérique (DINUM), une feuille de route sur la politique des données, des algorithmes et des codes sources. Ces documents définissent les ambitions et les enjeux des politiques publiques en matière de données, d'algorithmes et de codes sources publics, les actions prioritaires à mettre en œuvre, ainsi que la gouvernance mise en place pour garantir leur exécution.

Depuis 2019, le ministère de l'Éducation nationale et de la Jeunesse a instauré un comité d'éthique des données d'éducation, créé le poste d'administrateur ministériel des données (en septembre 2019), mis en place en 2020 une organisation et une gouvernance de la politique des données, structurée autour d'un Comité stratégique des données rassemblant l'ensemble des directeurs d'administration centrale, et accéléré l'ouverture des données produites par le ministère.

C'est pourquoi le projet de création d'un Education Data Hub s'intégrait naturellement dans cette dynamique, afin de proposer une grande variété de modalités d'accès, d'ouverture et de partage de données, proposées par des producteurs publics ou privés, au service des apprentissages des élèves, de la recherche, et du développement de services innovants, en lien avec la stratégie nationale pour l'intelligence artificielle.

Pour la mise en œuvre de cette proposition, le Health Data Hub, la plateforme des données de santé pouvait servir de modèle de référence.

Lancée en 2019, cette plateforme a pour objectifs de :

- assebmber, organiser et rendre disponibles les données de santé,
- favoriser leur utilisation à des fins de recherche,
- faciliter l'information des patients et l'exercice de leurs droits.

Le Health Data Hub contribue également à l'interopérabilité des données de santé et soutient certains projets de recherche. En raison des similitudes et de la sensibilité des données traitées, l'expérience acquise lors de la mise en œuvre de cette plateforme peut être mobilisée pour construire l'Education Data Hub.

D'autres initiatives, portées, soutenues ou suivies par la Direction du numérique pour l'éducation (DNE) du MENJ, ont également servi de référence pour l'Éducation Data Hub :

- HUBBLE (2015-2021): ce projet visait à créer un observatoire national pour la construction et le partage de processus d'analyse de données massives issues des traces laissées dans les environnements d'apprentissage en ligne, afin d'analyser et d'expliquer les phénomènes d'enseignement et d'apprentissage dans ces environnements. Neuf études de cas ont été menées, du primaire à l'université.

Des outils spécifiques ont été conçus et utilisés de façon itérative : scénarios d'analyse, collecte de traces, tableaux de bord. Des recommandations ont été formulées, notamment sur le format des traces à collecter, un point à anticiper dès la phase de conception d'une plateforme d'apprentissage.

- METAL (2016-2021): ce projet avait pour objectif la conception, le développement et l'évaluation d'un ensemble d'outils de suivi personnalisé, fondés sur les learning analytics, pour proposer des parcours personnalisés en apprentissage de l'écriture en langue étrangère. Les outils de collecte et de suivi ont été co-construits avec des chercheurs, des enseignants et des élèves. Des dispositifs innovants de collecte de données (eye-tracker) et d'interaction (assistant virtuel) ont été expérimentés dans des établissements volontaires.
- LOLA (since 2018): il s'agit d'une plateforme abordant différents champs des learning analytics : données, modèles, applications, évaluations et accompagnement. Mise en œuvre par le laboratoire de recherche LORIA, les données y sont hébergées et l'accès nécessite la création d'un compte afin d'assurer le suivi et la sécurité des jeux de données et des scénarios d'évaluation conçus dans le cadre du projet.
- Le groupe thématique « Analyse des traces d'apprentissage / Learning analytics » (2017-2018) : ce groupe visait à réaliser un état de l'art des learning analytics avec les outils numériques. La synthèse finale a été précédée de quatre bulletins de veille : sur les outils et méthodes, les usages scolaires, les enjeux éthiques, et les apports de la recherche. Il a été animé par les laboratoires de recherche LIP6 et TECHNE, avec le soutien de l'opérateur Canopé.
- Les projets lauréats du Partenariat d'innovation Intelligence Artificielle (depuis 2019) : il s'agit d'un dispositif innovant de commande publique, lancé par le ministère de l'Éducation nationale, visant à développer des solutions fondées sur l'intelligence artificielle, en partenariat avec des entreprises et des laboratoires de recherche. Le dispositif a d'abord concerné l'enseignement en cycles 2 et 3 (6 à 9 ans), avant d'être élargi au secondaire.

L'Education Data Hub a été inscrit dans la stratégie nationale d'accélération Enseignement et numérique, dans la feuille de route 2021-2023 sur la politique des données du MENJ, ainsi que dans la stratégie numérique pour l'éducation 2023-2027.

Même s'il ne sera pas mis en œuvre sous la forme initialement envisagée, les études et ressources mobilisées pendant la phase de préfiguration serviront à répondre aux besoins identifiés, sous des formes plus adaptées aux réalités du terrain et aux évolutions technologiques.

Le cadre d'analyse

L'étude de cas est analysée à partir du cadre d'analyse développé par le projet [Agile EDU literature review on datafication in and of education](#). Ce cadre considère l'usage des données et la datafication comme un écosystème impliquant divers acteurs de l'éducation, chacun ayant ses propres priorités et sa propre vision des défis éducatifs à relever grâce à l'exploitation des données.

Cet écosystème repose sur l'interrelation entre trois piliers : (1) Les droits, les réglementations et la protection de la vie privée (à l'échelle nationale et internationale) ; (2) L'usage des données pour l'enseignement et l'apprentissage (au niveau de la classe et de l'individu) ; (3) La gouvernance des données (au niveau régional et local). Au centre de ces trois piliers se trouvent les données numériques générées par les élèves et les enseignants.

Méthode

La phase de préfiguration de ce projet a été principalement pilotée par Federica Minichiello (alors cheffe de projet à l'Institut national de recherche en sciences et technologies du numérique – INRIA), Nicolas Roussel (directeur du centre Inria de l'Université de Bordeaux) et Philippe Ajuelos (administrateur des données, algorithmes et codes sources au MENESR).

Un rapport a été produit et publié en février 2024 sous le titre : *Plateforme des données de l'éducation : rapport de préfiguration*.

Pour recueillir les informations nécessaires, des entretiens semi-directifs ont été conduits en février 2024 avec Federica Minichiello et Philippe Ajuelos. Pour la seconde phase de cette étude de cas, trois nouveaux entretiens ont été réalisés en novembre 2024 : avec Philippe Ajuelos pour les volets relatifs à la *formation* et à l'*open data*, avec Federica Minichiello pour l'outil *ÉduPilote (Tableau de bord du numérique éducatif)* et avec Hugues Labarthe, délégué régional adjoint au numérique éducatif, pour le projet *Wikidata Challenge*.

Analyse

Le projet aborde les trois piliers identifiés dans le cadre d'analyse Agile EDU, qui seront examinés dans cette section. Cependant, il s'intéresse principalement à la question de la gouvernance des données, qui constitue son axe central.

Gouvernance des données

L'une des premières questions soulevées est celle de l'accès aux données. **Quels sont les obstacles à l'accès aux données ?**

Actuellement, le projet n'a pas encore atteint sa maturité, notamment en ce qui concerne la préparation technologique. Plusieurs projets de plateformes visent à centraliser une grande quantité de données, mais ils se heurtent à des difficultés similaires : personne n'a véritablement accès aux données centralisées, principalement en raison de l'extrême hétérogénéité des systèmes d'information. Certains de ces systèmes sont obsolètes, conçus à une époque où la valeur des données n'était pas prise en compte. Bien que ces données puissent être captées, même en mobilisant d'importants moyens humains, financiers et technologiques, leur exploitation reste complexe.

Un autre frein tient au fait que les données ne sont pas toutes centralisées : elles sont partagées avec d'autres acteurs, comme les collectivités territoriales, les associations ou les entreprises, dans le cadre de relations juridiques et contractuelles diverses.

S'agissant des données sensibles, qu'elles soient réglementairement qualifiées comme telles ou qu'il s'agisse de données individuelles non sensibles mais appelant à la vigilance, elles ne sont pas centralisées. Le ministère ne dispose pas de la gouvernance nécessaire pour assumer cette responsabilité, tant d'un point de vue juridique que dans le cadre de sa mission de service public. Il lui faudrait, pour cela, des ressources humaines, organisationnelles, techniques et financières à la hauteur des besoins pour instaurer une telle gouvernance.

Un autre obstacle identifié au cours de cette phase de préfiguration est la qualité des données. À différents niveaux de l'institution scolaire, certaines données censées être saisies ne le sont pas, en grande partie faute de retour ou de valorisation sur les informations saisies.

La question des outils de collecte de données constitue également un point de vigilance : il est essentiel de réduire autant que possible la charge pesant sur les personnes chargées de la saisie, d'éviter les redondances, et de permettre la réutilisation des données à des fins de vérification. Plus la collecte est simple, plus elle sera efficace. Par exemple, dans le suivi des absences des élèves dans le secondaire, les absences enregistrées en classe déclenchent automatiquement l'envoi d'un SMS aux parents pour les informer. Ce sont des outils simples, compatibles avec le RGPD, qui peuvent grandement alléger le fonctionnement quotidien.

Des outils de vie scolaire existent déjà, bien établis dans le second degré, ainsi que des environnements numériques de travail. (ENT) Dans le premier degré, le taux d'équipement des écoles est passé de 13 % à 33 % en trois ans. Dans certaines régions, les données sont accessibles et partagées entre les acteurs, mais ce n'est pas le cas à l'échelle nationale. Chaque territoire a ses spécificités, non seulement en matière d'équipement, mais aussi de ressources, intégrées et poussées dans les médiathèques des ENT, ce qui représente un investissement important pour les collectivités, qui attendent en retour une valorisation des usages, mais aussi le développement des compétences, notamment numériques. Il est donc essentiel de permettre à chacun de s'approprier ces données, étape par étape, en commençant par l'écoute. Même s'il n'a pas été possible de concrétiser un projet de grande ampleur tel que la plateforme de données, le ministère a fait le choix de développer des cas d'usage ciblés, identifiés lors de la phase de préfiguration, et d'apporter des réponses concrètes aux demandes des usagers.

Enfin, un constat marquant concerne la disparité dans l'accès et l'usage des données, notamment en lien avec les politiques menées par les collectivités. Certaines formulent de nombreuses demandes et disposent d'un volume de données conséquent, tandis que d'autres sollicitent peu les établissements scolaires. L'approche varie fortement entre les directeurs d'école et les chefs d'établissement : certains établissements ne perçoivent pas l'intérêt des enquêtes menées par les collectivités, tandis que d'autres, mieux soutenus localement, développent une politique claire en lien avec les directions d'école. Pourtant, parvenir à un consensus pourrait permettre de mieux valoriser les données disponibles, d'éviter les demandes redondantes et d'alléger la charge de travail de tous.

Si les pratiques de partage et d'échange de données varient d'une région à l'autre, les enquêtes menées par le ministère, quant à elles, font consensus : leur utilité est largement reconnue. Le problème réside plutôt dans la charge de travail qu'elles représentent, en particulier pour les chefs d'établissement, et dans la lourdeur de conduire des enquêtes à l'échelle nationale.

La deuxième question porte sur la propriété, le partage et l'usage des données : **Qui a accès à quelles données ?**

Les données éducatives suscitent un intérêt croissant, en France comme au niveau européen. En France, malgré une certaine forme de décentralisation, les compétences sont clairement réparties dans un système qui reste très centralisé. De l'administration centrale aux académies, en passant par les régions, les collectivités locales, jusqu'aux personnels de terrain et aux chefs d'établissement, de nombreux acteurs s'intéressent à la question. En conséquence, tous ces acteurs détiennent des données et souhaitent accéder à celles des autres. Il faut également compter avec les entreprises de technologie éducative (EdTech), comme les éditeurs de logiciels de gestion de la vie scolaire, les plateformes d'apprentissage (ENT) ou les fournisseurs de ressources pédagogiques déployées dans le système éducatif.

Chacun détient une partie de l'information potentiellement pertinente et cherche à accéder à d'autres jeux de données. Prenons l'exemple du tableau de bord du numérique éducatif : les collectivités disposent de données sur les équipements ou les ressources numériques, mais elles aimeraient connaître les taux d'usage de ces ressources, des informations qu'elles n'ont pas aujourd'hui. Elles seraient, par exemple, intéressées par les données issues de Pix, car elles ont investi dans des

équipements et souhaitent savoir si ces investissements permettent réellement aux élèves de progresser dans leurs compétences numériques. Pix est un service public en ligne permettant d'évaluer les connaissances et compétences numériques. La certification du niveau de maîtrise est obligatoire en fin de collège et en fin de terminale.

Philippe Ajuelos a rejoint le projet en décembre 2021. La plateforme dépendait alors de la validation d'un programme plus vaste : le Programme et Équipement Prioritaire de Recherche (PEPR) intitulé « Enseignement et numérique », relevant de l'enseignement supérieur et de la recherche. Cela lui a permis de mieux comprendre le projet, ses enjeux, et de le défendre dans un cadre plus large. La note de cadrage du projet prévoyait une phase de préfiguration permettant de dialoguer avec les parties prenantes de l'écosystème éducatif, à la fois internes et externes au ministère. Cela incluait les organisations professionnelles, les enseignants, notamment en lien avec leur liberté pédagogique et l'identification de leurs besoins.

Un des apports majeurs de cette phase de préfiguration a été de réfléchir à l'emplacement des données éducatives, à leur utilisation, à l'importance de ces données, et à la qualité des données saisies. Pour que les données soient réellement exploitables, leur qualité est essentielle. Il est donc crucial que ceux qui les saisissent aient un retour sur leur usage. Faute de quoi, même avec les meilleurs algorithmes, les prédictions, recommandations ou prescriptions produites seront erronées. Un bon moyen de sensibiliser à cette qualité est d'assurer un retour d'information, de donner de la visibilité sur les objectifs de la saisie, et de proposer des référentiels comparatifs avec d'autres établissements. Dans le cas des publications en open data, les établissements peuvent aussi signaler des erreurs, ce qui permet de corriger les données et d'améliorer les saisies futures.

Impliquer les utilisateurs dans l'amélioration de la qualité des données est une manière très concrète de les acculturer à ces enjeux. C'est pourquoi l'ensemble des acteurs ont été entendus lors des auditions : personnel de terrain, chercheurs, entreprises et associations. Cela permet de créer un cercle vertueux entre exigence de qualité à la saisie et retour d'usage, tant pour les personnels de direction que pour les autorités qui partagent une grande quantité de données.

Le public visé par le projet est très large: enseignants, chefs d'établissement, personnels non enseignants, collectivités, État, chercheurs, entreprises, parents, associations liées aux collectivités, ou encore acteurs spécialisés dans les questions de données. Cet intérêt s'est traduit par un fort enthousiasme lors des auditions menées à partir de novembre 2022, qui ont finalement permis d'entendre deux fois plus de personnes que prévu, tant les parties prenantes ont manifesté leur volonté de participer.

Chacun de ces publics a formulé ses propres attentes vis-à-vis de la future plateforme. Pour les enseignants, par exemple, l'objectif est d'utiliser les données d'apprentissage pour alléger certaines tâches. Il s'agirait, par exemple, d'utiliser les données collectées pour concevoir un outil de correction des travaux des élèves, capable d'identifier les besoins, de fournir un retour individualisé et de libérer du temps pour mieux accompagner les élèves. Toutefois, cette demande n'a pas été reçue de manière unanime par les autres parties prenantes et doit faire l'objet d'un arbitrage du Comité d'éthique des données. À ce titre, les partenariats d'innovation en intelligence artificielle (P2IA) représentent ce que le ministère peut proposer aux enseignants volontaires. Les onze projets P2IA sont évalués par le GTnum #AppropriIA (« évaluation et accompagnement des processus d'appropriation de l'IA dans les

dispositifs pédagogiques »), à partir des jeux de données qu'ils génèrent. L'objectif est d'analyser les évolutions de pratiques en matière de différenciation, de personnalisation et d'accompagnement pédagogique, à travers une approche scientifique fondée sur les usages, les représentations et l'analyse des traces numériques laissées par les enseignants. <https://edunumrech.hypotheses.org/14376>

Les auditions ont permis de donner la parole à la fois aux soutiens et aux opposants du projet de plateforme.

Les demandes exprimées étant très diverses, une approche spécifique a été mise en place pour organiser ces retours, les comprendre, et les prendre en compte. Au terme de cette phase de préfiguration, 14 besoins ont été identifiés. Des solutions sont en cours d'étude ou déjà mises en œuvre pour y répondre, mais sans passer par la création de la plateforme initialement envisagée.

Quel rôle et quelle place pour les entreprises EdTech? est également une question soulevée par cette étude de cas.

Dans le cadre du projet de plateforme de données, les acteurs de la EdTech ont été interrogés au même titre que les autres parties prenantes : fournisseurs d'ENT (Environnements Numériques de Travail), éditeurs de logiciels de vie scolaire ou encore producteurs de ressources pédagogiques. Un meilleur contrôle des informations utilisées actuellement par les logiciels scolaires représenterait un atout. Bien sûr, il existe un cadre réglementaire (RGPD) et des responsables de traitement des données; cependant, le partage de ces informations serait bénéfique à l'ensemble de la communauté, qu'il s'agisse de données issues du ministère ou d'un éditeur privé.

Selon Federica Minichiello, les données, qu'elles viennent des élèves ou des enseignants, appartiennent d'abord aux personnes concernées. Elle espère que les projets fondés sur les données permettront aux individus de reprendre la main sur leurs propres données. Toutefois, la facilité de circulation des données varie selon la position occupée par les acteurs sur le marché. Cette question fait actuellement l'objet d'un travail mené par le ministère.

Dans ce projet, le ministère jouait un rôle moteur, et les entreprises EdTech avaient un double rôle à jouer : en tant que fournisseurs potentiels de données, mais aussi utilisateurs. En tant qu'utilisateurs, les entreprises EdTech souhaitaient accéder à de nombreuses données produites par le ministère, comme les résultats des évaluations nationales. En revanche, lorsqu'il s'agissait de partager leurs propres données, en particulier avec d'autres entreprises, les choses devenaient plus complexes. Un partage de données entre entreprises suppose une vision à long terme, qui dépasse les intérêts immédiats.

Certaines entreprises ont accepté de partager leurs travaux au sein de groupes de travail, notamment sur les traces d'apprentissage via les standards xAPI. Elles ont fait le choix d'y consacrer du temps, de partager leurs expériences et de mettre temporairement leurs intérêts commerciaux de côté.

En France, il est peu courant de consacrer des ressources à des groupes de travail sans retour sur investissement direct. Le rôle de l'acteur public est donc d'assurer une représentation équilibrée, notamment des différentes associations professionnelles du secteur, et de veiller à la défense de l'intérêt général.

Dans le cas d'usage des Territoires Numériques Éducatifs (TNE), les traces d'apprentissage sont précieuses à analyser. Des solutions comme Lalilo, axée sur l'apprentissage de la lecture, sont largement utilisées. Au cours des expérimentations, certaines questions sont apparues, par exemple sur une éventuelle corrélation entre les profils d'élèves en lecture et en écriture. Cette question reste aujourd'hui ouverte, même si les enseignants peuvent avoir une intuition fondée sur leur expérience. Chaque solution issue du programme P2IA (Partenariat d'Innovation en Intelligence Artificielle) est liée à un laboratoire de recherche, mais ces recherches restent attachées à la solution elle-même, tandis que les données pourraient apporter un regard indépendant, d'être plus précis dans les formats demandés et dans les règles de structuration et d'étiquetage des données. Cependant, chaque entreprise ayant proposé ses propres formats de données, cela souligne l'importance pour le ministère d'être plus précis dans les formats demandés et dans les règles d'étiquetage des données.

Il est essentiel de s'assurer que les informations échangées, même dans le cadre de partenariats réussis, profitent à l'ensemble de la communauté. C'est un point de vigilance majeur. C'est aussi ce qui a conduit à la création d'une association, dont l'objet est de garantir que tous les travaux issus des groupes de travail soient publiés et rendus accessibles de manière ouverte, comme le stipulent ses statuts.

Droits, réglementations, protection de la vie privée

Sur le sujet des droits, de la réglementation et de la vie privée, la première question soulevée par le Hub est la suivante : **Comment les réglementations et législations européennes et nationales sont-elles traduites concrètement dans les pratiques et la mise en œuvre locale au sein des établissements scolaires ?**

L'approche visait à poser les bases de la connaissance, notamment en ce qui concerne l'application et la compréhension des réglementations, tant européennes que nationales. Cela incluait non seulement les textes réglementaires tels que les décrets, lois, règlements européens ou directives, mais aussi ce qu'on appelle le droit positif, c'est-à-dire l'ensemble des dispositions réglementaires, légales ou jurisprudentielles en vigueur. L'analyse portait également sur les considérations éthiques liées aux pratiques et usages des données. Cela concernait notamment les questions suivantes : comment les réglementations sont-elles comprises et appliquées, que met-on en place pour s'assurer que les usages sont conformes à ces règles, quelles sont les implications éthiques de l'utilisation des données, en particulier celles concernant les élèves et les enseignants ?

Ce qui est ressorti de cette réflexion, est une utilisation très restrictive des données, fondée non seulement sur le respect des textes réglementaires, mais aussi sur des principes éthiques. Par exemple, dans le cadre de l'orientation scolaire, deux approches sont possibles : la première consiste à déterminer ce que fera un élève à partir de ses résultats passés ; la seconde consiste à analyser ces résultats à la lumière des objectifs fixés par l'élève, afin d'identifier des leviers de progression. C'est cette deuxième approche qui est conforme à une démarche éthique. Cependant, il existe une méconnaissance réelle du RGPD et une forme de crainte qui y est associée. Il y a, par exemple, une réticence à conserver les données au-delà d'un an, alors même que certaines analyses nécessitent un suivi sur plusieurs années, notamment pour effectuer des comparaisons dans le temps. Le RGPD

n'interdit pas cette conservation prolongée, il exige simplement qu'elle soit justifiée et décrite dans la finalité du traitement. Cela met en évidence une problématique d'acculturation au droit : les règlements sont souvent perçus comme des freins à l'usage des données, alors que c'est surtout le manque de compréhension de ces textes qui crée des blocages.

Une fois les besoins mieux identifiés, il est apparu que la centralisation ou la concentration des données n'était pas forcément indispensable. Le ministère n'a pas nécessairement vocation à assurer la gouvernance dans tous les cas. Cela peut être pertinent pour certains usages, mais pas toujours ce sera pas nécessairement pour d'autres, et dans certains cas, le ministère ne dispose peut-être pas actuellement des moyens nécessaires pour le faire.

La deuxième problématique concerne l'usage des données. **Comment les besoins en matière d'exploitation des données sont-ils définis ?**

Un autre obstacle concerne la formalisation du besoin en matière de traitement des données. Par exemple, pourquoi conserver les résultats des élèves d'une année sur l'autre ? Il s'agit de suivre leur progression, notamment à l'échelle des cohortes, pour pouvoir adapter l'accompagnement et en assurer le suivi. Pourtant, les enseignants ne disposaient pas de ces informations, souvent en raison d'une méconnaissance des règles en vigueur. Lorsque les textes réglementaires sont expliqués, cela suscite fréquemment de la méfiance, voire de la crainte à l'idée d'agir. C'est justement là que les délégués à la protection des données (DPO) ont un rôle pédagogique important à jouer auprès des établissements, afin de les informer sur ce qu'ils ont le droit et la possibilité de faire. En l'absence de sensibilisation, les acteurs préfèrent ne rien faire plutôt que de mal faire, ce qui devient problématique. Lors des auditions, une question a été posée aux agents : « Que feriez-vous si vous pouviez conserver les données ? » Habituellement, leur posture est plutôt restrictive en matière d'usage, mais cette fois, ils ont été placés dans une situation différente, partant de leurs attentes pour envisager ce qui pourrait être possible. Ce retour d'expérience a permis de faire émerger un besoin clair de formation.

Même si, dans d'autres secteurs, comme l'agriculture, le tourisme ou la santé avec le Health Data Hub, et même à l'échelle européenne, d'autres plateformes se mettaient en place, l'éducation a conservé sa spécificité, bien que des points communs soient apparus au fil des échanges. Par exemple, dans le domaine de la santé, les données diffèrent de celles de l'éducation, mais les enjeux sont souvent similaires : accéder à certaines données et pouvoir les partager. Dans les deux cas, les laboratoires de recherche rencontrent les mêmes difficultés en début de projet, avec un temps de préparation considérable avant de pouvoir commencer à travailler.

C'est précisément l'un des rôles que pourrait jouer une plateforme de données d'éducation : réaliser une partie de ce travail préparatoire, afin que les porteurs de projets puissent démarrer avec une base prête à l'emploi, puis se concentrer sur les aspects spécifiques de leurs recherches ou actions.

De nombreux acteurs ont exprimé un souhait d'accès aux données, sans toujours pouvoir dire précisément ce qu'ils en feraient une fois obtenues. En revanche, dans des projets en cours comme ÉduPilote (un tableau de bord partagé entre collectivités, autorités académiques, chefs d'établissement et administration centrale dans le champ du numérique éducatif), ce problème est bien moindre car les besoins sont beaucoup plus ciblés. Dans le cas de la plateforme de données, le champ des possibles

était trop vaste, ce qui rendait difficile l'alignement entre les attentes et les besoins des différents acteurs.

Par ailleurs, l'écosystème n'était pas encore suffisamment mature pour accueillir un projet d'une telle ampleur. L'équipe mobilisée était très réduite (trois personnes), et il était difficile de trouver un équilibre entre les besoins, la qualification des données, la compréhension des parties prenantes et la mise en œuvre concrète du travail.

Les attentes étaient nombreuses quant à ce que le ministère pourrait rendre disponible en matière de données, mais la réciproque n'était pas vraie. Par exemple, les entreprises EdTech n'étaient pas particulièrement enclines à partager leurs données, car celles-ci résultent de leurs investissements et leur partage supposerait de les ouvrir à des chercheurs, enseignants, collectivités locales, voire à d'autres entreprises, parfois concurrentes. Pourtant, la réciprocité est essentielle dans un tel projet.

L'approche orientée service est en revanche bien comprise. Il n'est pas nécessaire de connaître la littératie des données pour la comprendre. L'usage des données répond à un besoin, mais pour répondre à ce besoin, il faut d'abord saisir des données. En remplissant un tableau ou en utilisant une interface, chaque personne participe au partage des données. Et en harmonisant les demandes, il devient inutile de saisir plusieurs fois les mêmes informations.

Utilisation des données pour l'enseignement et l'apprentissage

Comment les compétences numériques des élèves et des enseignants en matière d'usage des données et de compréhension des pratiques de partage sont-elles soutenues ? Cette question est au cœur des préoccupations citoyennes.

Au-delà des enjeux de protection des données qui relèvent du champ d'action du délégué à la protection des données, l'accent est mis sur la formation des enseignants et des personnels de direction, non seulement dans une logique d'amélioration continue du service public d'éducation, mais aussi dans une perspective de citoyenneté numérique. Il s'agit de développer une forme de pouvoir d'agir, permettant de comprendre le monde à travers les données et les technologies — non seulement comme un outil d'interprétation, mais comme un fait social en soi, qui exige un accompagnement. Cette approche s'inscrit pleinement dans la mission de l'Éducation nationale, qui est de préparer les jeunes à devenir des citoyens actifs et responsables, capables de comprendre les données, et potentiellement de développer leurs propres outils ou algorithmes.

Comment les entreprises EdTech veillent-elles à ce que leurs produits contribuent au développement de l'ensemble des compétences clés ?

Pour qu'une technologie soit adoptée, il est essentiel que les enseignants puissent s'appuyer sur des éléments concrets. Tous les cas d'usage identifiés sont pertinents pour accompagner l'évolution des pratiques pédagogiques, notamment grâce à l'étude des traces d'apprentissage issues des solutions, en particulier celles développées dans le cadre du partenariat d'innovation en intelligence artificielle (P2IA). La collaboration avec les laboratoires de recherche est indispensable. La technologie éducative

porte de nombreuses promesses, et les premiers résultats montrent de véritables expériences d'adaptation ou de personnalisation des apprentissages. Cependant, ces résultats initiaux doivent être confirmés par des recherches complémentaires. Le sujet est porteur et potentiellement transformateur. Donner aux enseignants le pouvoir d'adapter leur enseignement à chaque élève constitue une véritable transformation, indépendamment de l'existence d'une plateforme de données. Les données issues de la première vague d'utilisateurs des solutions P2IA commencent à être analysées. Il est intéressant d'observer les traces dont disposent les fournisseurs de solutions, la manière dont chaque entité est structurée, et ce que les data scientists peuvent faire de ces données. Cela illustre à la fois le cadre que le ministère peut définir pour ses futures actions et la réalité de ce qui est effectivement mis en œuvre sur le terrain.

L'arrêt du projet de plateforme de données n'a pas affecté ce chantier essentiel, car les solutions déployées sont censées être adaptatives et offrir une réelle personnalisation aux élèves. Le ministère continue d'investir dans ces dispositifs. Une fois les données collectées, il sera possible de voir ce qui en ressort, de mieux comprendre ce que sont réellement les traces d'apprentissage, et de proposer des améliorations pour les futurs P2IA. L'ouverture des données est indispensable, mais elle doit être accompagnée d'un travail de recherche pour être pleinement exploitable.

L'un des premiers objectifs de l'équipe a été de traduire la feuille de route sur le plan opérationnel, afin que tous les sujets soient compris par l'ensemble des acteurs concernés. Cela passait notamment par une clarification des termes : algorithmes, intelligence artificielle, open source, open data, communs numériques, etc. Pour chacun des neuf axes de la feuille de route 2021-2024, l'équipe a fait le lien entre les [42 actions](#) proposées et la réalité du terrain.

Des projets qui répondent aux besoins

Les besoins identifiés sont pris en compte soit dans des projets déjà lancés, soit dans des projets en cours, soit dans des perspectives pour 2025-2026. Les données seront mobilisées pour répondre à ces besoins, mais en dehors du cadre d'un hub. Il s'agira plutôt de dispositifs tels qu'une plateforme d'échange de certaines données avec les collectivités territoriales, ou encore de projets d'analyse des données d'apprentissage, qui ne nécessitent pas de centralisation des données.

Gouvernance des données

- **Programme IDEE:** Accès aux données de la DEPP (Direction de l'évaluation, de la prospective et de la performance) pour les chercheurs. Le programme IDEE est une infrastructure de recherche conçue pour développer des recherches expérimentales à grande échelle dans le domaine de l'éducation et favoriser l'utilisation de leurs résultats. Il s'adresse à l'ensemble des acteurs de la recherche et de l'éducation, en France comme à l'international.
<https://www.idee-education.fr/>



The screenshot shows the IDEE website with a green header bar containing the text 'Données DEPP'. Above the header, there is a navigation bar with links: 'Le programme IDEE', 'La méthode expérimentale', 'IDEE pour la recherche', 'IDEE pour les acteurs de l'éducation', 'Projets', and 'Actualités'. To the right of the navigation bar is a yellow button labeled 'Nous contacter'. The main content area below the header contains a paragraph of text about the collaboration between IDEE and the DEPP to facilitate research access to administrative data.

Catalogues des données

Ce catalogue donne à voir la richesse des données administratives et d'enquête sur le système éducatif français. Il permet à toute personne, et notamment au personnel de recherche, de prendre connaissance des données existantes à la DEPP. Les données qui y sont documentées concernent entre autres les élèves, les établissements et les personnels de l'éducation. Le catalogue permet également aux chercheurs d'initier une demande d'accès à ces données. Cet accès s'inscrit dans un cadre conventionnel adapté.

[> Accéder au catalogue de données](#)

[> Faire une demande d'accès aux données de la DEPP](#)

depp@logue

- **Facilitation de la circulation des données** entre acteurs publics et privés grâce à l'élaboration en cours de référentiels techniques. <https://eduscol.education.fr/3827/doctrine-technique-du-numerique-pour-l-education>
- **Plateforme Open Data pour l'Éducation, la Jeunesse et les Sports**
Soutenue par le ministère de l'Éducation nationale, cette plateforme met à disposition des données publiques relatives aux politiques menées par les deux ministères. Elle s'inscrit dans la volonté de l'État de proposer une gouvernance des données toujours plus transparente et au service des usagers. Le portail data.education.gouv.fr donne accès à un large éventail de jeux de données produits par les services du ministère, dans des formats variés et accessibles à tous grâce à des visualisations de données personnalisées. Que l'on soit collectivité locale, chercheur, journaliste, acteur de la EdTech, parent, enseignant, élève ou simple citoyen, chacun peut y trouver des données utiles. En 2024, le portail data.education.gouv.fr a enregistré 80 millions d'appels API et 2,2 millions de téléchargements. En 2025, la plateforme intégrera un outil de recherche sémantique pour faciliter l'accès aux informations recherchées par les utilisateurs. <https://data.education.gouv.fr/pages/accueil/>
- **ÉduPilote**: Mise en place d'un tableau de bord partagé pour une meilleure gestion du numérique éducatif. Aujourd'hui, les chefs d'établissement ont une vision assez claire de leur propre établissement, mais peuvent manquer de repères en ce qui concerne le numérique. Par ailleurs, s'ils souhaitent développer des projets numériques, il peut être utile de savoir ce qui se fait dans les établissements voisins, parfois plus avancés sur ces questions, ou encore dans leur territoire, pour s'en inspirer. Le projet ÉduPilote vise à exploiter différentes sources de données pour produire des indicateurs et des visualisations permettant de mieux suivre les investissements et les déploiements en matière de numérique éducatif. Il couvre plusieurs thématiques : usages du numérique, équipements, compétences, formation. Le point de départ pour l'utilisateur repose sur des questions concrètes : Quel est le niveau Pix des élèves ? Comment se situe mon établissement par rapport à un autre quartier ? Quelle est l'allocation des ressources ? Quelle est ma vision du numérique ? Le projet propose la création d'un tableau de bord partagé, avec une lecture transversale du numérique, allant de l'échelle nationale jusqu'au niveau de l'établissement. Le périmètre initial concerne les collèges et lycées publics, en lien avec les collectivités locales et les autorités académiques. L'enjeu est de pouvoir visualiser son propre territoire, tout en ayant une vue sur ce qui se passe à proximité. Les données mobilisées sont déjà ouvertes, mais cet outil permet de les interpréter, de leur donner du sens. Une fois les données « dépliées », elles deviennent des indicateurs exploitables, utiles pour les chercheurs, les collectivités, les enseignants. Un prototype de tableau de bord a été développé avec des groupes pilotes réunissant des collectivités territoriales, des académies et l'administration centrale. L'utilisateur peut choisir une année scolaire, sélectionner des indicateurs de formation, et voir les données s'ajouter progressivement, dans une logique de construction d'indicateurs macro pertinents. Pour accéder à l'outil, les collectivités doivent

accepter de partager leurs données d'équipement. Le partage des données permet une meilleure compréhension du territoire.: Le processus repose sur les étapes suivantes :1. Partage des données - 2. Création d'un compte - 3. Retour d'expérience des utilisateurs ou des groupes de travail - 4. Amélioration continue par itérations

<https://tube-institutionnel.apps.education.fr/w/kRicMhc6qMnYemdrcuV8X8>

- **DoCTE²**: Plateforme d'échange de données entre les régions et le ministère
Dans le cadre de la mise en œuvre de la convention-cadre relative à l'échange des données nécessaires au pilotage du système éducatif, signée à l'automne 2024 entre le ministère de l'Éducation nationale et de la Jeunesse et l'association Régions de France, une plateforme d'échange de données, nommée DoCTE², a été mise en place. Cette plateforme permet aux différents acteurs impliqués de partager des données de manière simple, efficace, intuitive et sécurisée, tout en garantissant un accès restreint aux seules parties prenantes concernées. Elle propose plusieurs formats de partage, tels que des jeux de données, des visualisations interactives, et divers outils facilitant une collaboration optimale. DoCTE² sera utilisée pour l'ensemble des projets valorisant les données, y compris ÉduPilote.

Littératie des données

- **Wikidata Challenge en classe**: une acculturation des élèves par l'expérimentation
Ce projet, mené auprès d'élèves de seconde dans la région Bretagne, vise à favoriser leur acculturation aux données à travers une expérimentation pédagogique basée sur Wikidata. Objectif : permettre à des élèves de seconde de s'approprier une question sociale vivante, en les amenant non seulement à ne plus être uniquement objets de données, mais à les manipuler activement.

Le projet se déroule en plusieurs phases :

- Novembre : formation des formateurs, puis des enseignants
- Phase 1 : positionnement des classes participantes
- Phase 2 : sélection des équipes
- Phase 3 : 12 et 13 mars : challenge au Édulab Pasteur de Rennes
- Phase 4 : retours d'expérience des classes participantes

https://www.wikidata.org/wiki/Wikidata:Challenge_Wikidata_en_classe

Thème : école et patrimoine en Bretagne

1- Définir une problématique

Le **matrimoine** existe, mais il a longtemps été invisible. Que cela signifie-t-il dans une région comme la Bretagne ?

Livrable : identifier les ensembles de données pouvant être mobilisés.

2- Manipuler et analyser des jeux de données

Problématiser une question de société pour susciter l'intérêt d'élèves qui ne sont pas nécessairement familiers avec l'analyse de données.

Bases de données mobilisées : Wikidata – Arts et architecture – Éducation et orientation – Politiques publiques.

3- Créer et enrichir des données

RÉGION ACADEMIQUE BRETAGNE | Délégation régionale académique au numérique éducatif

Phase 2. Manipuler et analyser les données

Où sont les femmes, à l'École, en Bretagne ?

- Construire une problématique
- Identifier, collecter, représenter des données
- Présenter la démarche et ses résultats

LIVRABLE 2. CRÉER UN POSTER

- Étayer la démarche
- Exécuter une requête SPARQL
- Analyser une représentation graphique
- Proposer une conclusion

<https://culturesnumeriques.ac-rennes.fr/wikidata>

Pour ce premier challenge, la région académique de Bretagne, la délégation régionale au numérique éducatif et le délégué académique au numérique ont accompagné le projet, avec le soutien de Wikimedia France (Wikidata) et de la Direction du numérique pour l'éducation (DNE-AMDAC). Cinq équipes de lycéens (22 filles et 5 garçons) ont mené une véritable enquête à partir de données ouvertes, autour d'un thème central : le patrimoine breton. Collecte de données, structuration, requêtes SPARQL, réflexion critique, analyse, rectification, enrichissement, et engagement citoyen : toutes ces dimensions ont nourri le projet, donnant lieu à des productions en ligne riches, documentées et publiées.

<https://porte-plume.app/projet/cc95889b-3af5-4b09-9ac9-07a8e9ba8401/blog?categories=35e87dbc-6281-458f-99a2-5d7a7a0e88ff>

Les projets finalistes :

- Résistantes bretonnes : le groupe Marceau à Quimper, un exemple de courage et de solidarité pendant la Seconde Guerre mondiale.
- Peintres bretonnes : Jeanne Malivel, Mary Piriou... une enquête sur l'omission des artistes femmes dans les récits officiels.
- Conteuses bretonnes : des femmes qui transmettent le patrimoine oral breton, souvent moins documenté que celui de leurs homologues masculins.
- Les femmes dans l'enseignement en Bretagne : un métier longtemps féminisé, mais un héritage encore peu valorisé dans les programmes scolaires.
- La place des femmes célèbres dans l'espace public : une toponymie encore largement masculine. Les élèves interrogent la présence des femmes dans les noms de rues et les monuments.
- Les résultats et jeux de données produits seront prochainement mis en valeur sur le portail data.education.gouv.fr.

- **Pix+Données:** Formation des personnels de l'Éducation nationale via deux parcours Pix dédiés aux données. Les parcours Pix+ ont été lancés en juin 2024. Pix est une plateforme de formation autonome et de certification des compétences numériques à destination des citoyens. Les deux parcours proposés viennent compléter le cadre existant afin de sensibiliser les agents du service public à trois grands enjeux liés aux données : Rechercher et produire des données de qualité, Ouvrir, protéger et partager les données en France, Utiliser et réutiliser les données. Dans Pix, 16 compétences sont évaluées, celles du Cadre de Référence des Compétences Numériques (CRCN), l'adaptation française du cadre européen DigComp. Le cadre est structuré en 5 domaines, dont celui de l'information et des données. Deux compétences spécifiques concernent directement les données : Gérer des données, et Traiter des données.

Par exemple, pour la compétence **Gérer des données** :

Niveau 1 : sauvegarder des fichiers sur l'ordinateur utilisé et les retrouver,

Niveau 5 : comprendre les métadonnées et leur fonctionnement.

Pour la compétence **Traiter des données** :

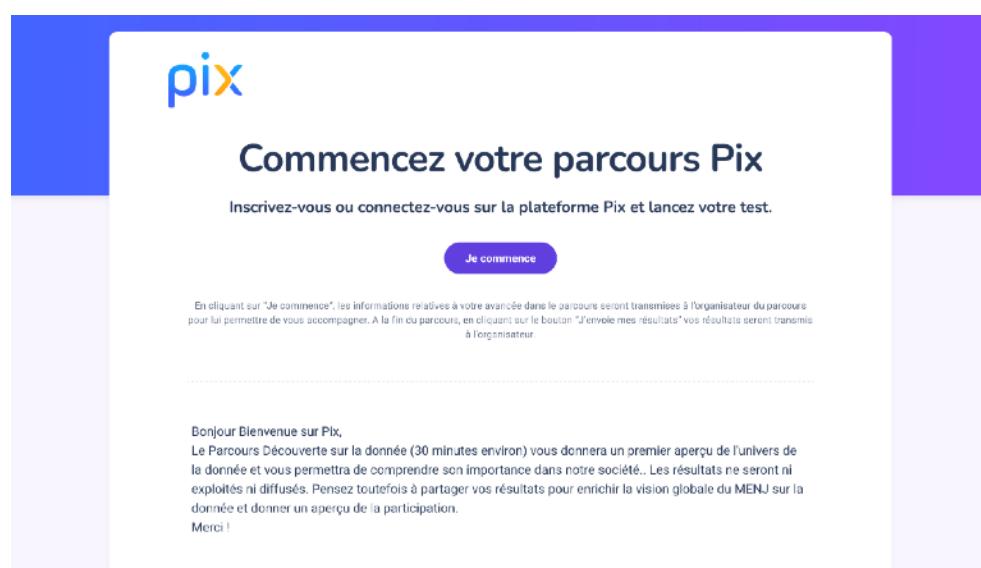
Niveau 1 : sélectionner et relier des informations issues de ressources numériques,

Niveau 5 : automatiser le traitement de données.

Ces compétences sont enrichies par deux nouveaux parcours :

Découverte (30 minutes) : une introduction à l'univers des données,

Panorama (60 minutes) : une exploration plus approfondie, avec une initiation à la manipulation des données.



Commencez votre parcours Pix

Inscrivez-vous ou connectez-vous sur la plateforme Pix et lancez votre test.

En cliquant sur "Je commence", les informations relatives à votre avancée dans le parcours seront transmises à l'organisateur du parcours pour lui permettre de vous accompagner. À la fin du parcours, en cliquant sur le bouton "J'envoie mes résultats" vos résultats seront transmis à l'organisateur.

Bonjour Bienvenue sur Pix,
Le Parcours Découverte sur la donnée (30 minutes environ) vous donnera un premier aperçu de l'univers de la donnée et vous permettra de comprendre son importance dans notre société.. Les résultats ne seront ni exploités ni diffusés. Pensez toutefois à partager vos résultats pour enrichir la vision globale du MENJ sur la donnée et donner un aperçu de la participation.
Merci !

<https://pix.fr/donnees>

Conclusion et recommandations

Conclusion

L'un des aspects innovants de l'initiative autour de la plateforme de données de l'éducation a été de ne pas partir des données disponibles, mais bien des besoins exprimés.

Même si la plateforme n'a pas été finalisée, la méthode de conception a permis de faire émerger de réels besoins. Plusieurs projets sont actuellement en cours pour y répondre, notamment en matière de gouvernance des données et de littératie des données. Un périmètre plus restreint, avec des collectivités volontaires et des partenaires comme Pix, permet une organisation plus simple et des échanges plus fluides et représentatifs. Cela montre que traiter les cas d'usage un à un permet de répondre à des besoins précis dans un cadre bien défini.

L'administrateur ministériel des données, des algorithmes et des codes sources (AMDAC) joue un rôle central pour accompagner et coordonner les acteurs vers une meilleure circulation des données et un usage respectueux des principes éthiques. Au-delà des initiatives déjà en cours, la démarche prospective engagée lors du travail de préfiguration de la plateforme des données de l'éducation se poursuivra sur le long terme, avec des expérimentations menées en collaboration avec les acteurs de la recherche.

Un autre changement important concerne l'orientation. Lors des auditions, l'ensemble des parties prenantes: collectivités, académies, associations, ont évoqué ce sujet. La personnalisation des parcours d'orientation, notamment à travers le programme Avenir de l'Onisep, l'utilisation de Parcoursup ou encore l'accompagnement des lycéens via la plateforme MonProjetSup, illustre la manière dont les données sont mobilisées pour faire évoluer l'orientation. Même si cette évolution n'est pas immédiatement visible d'un point de vue pédagogique, elle est significative sur le plan social. Un point de convergence est apparu avec le Health Data Hub, qui a travaillé sur la santé des enfants. Un cas d'usage a été proposé visant à mettre en place un partage de données liées à la vie scolaire pour alimenter des cohortes, comme cela se fait déjà pour certaines données médicales. Même si sa mise en œuvre se heurte à des contraintes liées à la nature sensible des données de santé, l'approche reste pertinente à condition d'être encadrée. Aujourd'hui, la santé, et notamment la santé mentale des jeunes et des adolescents, est un sujet de préoccupation majeur: croiser certaines informations permettrait à l'école d'adapter ses réponses.

La création d'une plateforme de données de l'éducation n'est pas actuellement à l'ordre du jour. Toutefois, cette question pourrait être réexaminée à court terme en fonction de l'évolution du niveau de maturité, de l'industrialisation des usages et de l'interopérabilité des systèmes.

On observe une véritable montée en compétences du ministère sur ces sujets, que ce soit en tant que commanditaire, à travers la définition d'une doctrine technique, ou sur les enjeux d'interopérabilité. Une prise de conscience et une progression réelles sont également visibles dans le secteur public comme chez les acteurs de la EdTech.

Recommandations

1- Une méthode efficace d'identification des besoins. Un travail mené avec l'ensemble des parties prenantes, en prenant le temps d'organiser des auditions et d'écouter les attentes de chacun, qui permet de faire émerger efficacement des besoins.

L'un des apports majeurs de cette exploration autour d'un possible Education Data Hub a été la volonté de développer des outils permettant une lecture plus orientée « service » des données.

2- De la souplesse pour proposer des réponses ciblées plutôt qu'une solution globale. Une autre nécessité, directement liée à la plateforme de données, concernait l'adaptation du budget du projet, initialement intégré dans un programme de recherche plus large, planifié sur cinq ans. Bien que des financements parallèles aient été obtenus pendant la phase de préfiguration, l'incertitude persistait quant à la mise en œuvre effective de la plateforme. Une clarification budgétaire était indispensable dans l'hypothèse où le projet devrait être maintenu à moyen ou long terme. Malgré les incertitudes, des efforts ont été menés pour démontrer la valeur du projet, le défendre, et mener à bien la phase de préfiguration afin d'en évaluer la faisabilité.

3- La formation à la littératie des données,: un aspect qui apparaît clairement lors de la phase de préfiguration. Il est particulièrement intéressant d'observer l'évolution des parties prenantes, chacune apportant des perspectives et des compétences différentes, tout en contribuant collectivement à faire avancer le projet lorsqu'elles travaillent ensemble.Toutefois, pour que cette dynamique soit effective, il est indispensable de commencer par identifier les données les plus pertinentes au regard de leurs besoins, ainsi que les modalités permettant d'y accéder.

L'efficacité doit constituer un objectif central, en rendant visibles les usages concrets des données, afin d'éviter l'écueil de dispositifs de collecte, tels que certaines enquêtes, que les acteurs ne jugent pas toujours utiles ou dont ils ne perçoivent pas clairement le sens ou la finalité.

Annexes

Acteur de la gouvernance ministérielle des données - Philippe Ajuelos

Philippe Ajuelos Philippe Ajuelos occupe actuellement le poste d'Administrateur des données, des algorithmes et des codes sources (AMDAC) au sein du ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche (MENESR). Ce rôle est multifacette : il agit à la fois comme ambassadeur de la politique de valorisation des données du gouvernement, tout en contribuant à l'élaboration d'une politique spécifique adaptée aux besoins propres de chaque ministère. Depuis décembre 2021, M. Ajuelos s'est appuyé sur l'équipe en place et a concentré ses efforts sur l'adaptation de la feuille de route nationale aux spécificités du ministère. L'objectif principal était de rendre cette feuille de route opérationnelle, afin que tous les acteurs concernés puissent se l'approprier. Cette démarche a débuté par l'instauration d'un langage commun autour de notions clés telles que les algorithmes, l'intelligence artificielle, l'open source, l'open data et les communs numériques.

Par ailleurs, au-delà des questions liées à la protection des données, qui relèvent du délégué à la protection des données, une attention particulière est portée à la sensibilisation des enseignants, et par extension des jeunes générations, à la compréhension du monde à travers les données et les technologies qu'ils utilisent au quotidien. L'objectif de cette éducation est de promouvoir non seulement la souveraineté mais aussi l'autonomie, en formant des citoyens actifs, dotés d'un esprit critique et d'une culture numérique. Il s'agit de former des individus capables non seulement d'utiliser des outils et d'alimenter des algorithmes, mais aussi de comprendre leur fonctionnement, et potentiellement de développer leurs propres outils. Cela inclut la capacité à décrypter les facteurs qui influencent les choix et à exercer leur libre arbitre.

Chef de projet - Federica Minichiello

Federica Minichiello a un parcours en économie, avec une spécialisation dans les institutions publiques et les organisations internationales. Elle a travaillé pour le CIEP, aujourd'hui France Éducation Internationale (FEI), où elle a mis en place un laboratoire d'innovation, consciente de l'importance du lien entre le numérique et le système éducatif dans une structure dédiée aux échanges internationaux. En parallèle, elle a poursuivi sa réflexion sur les données et leurs usages, en s'intéressant particulièrement aux questions de protection et à la recherche d'un équilibre entre valorisation des données et respect des droits, deux dimensions essentielles pour transformer les pratiques à partir des données.

Elle collabore avec, et travaille pour, la Direction du Numérique pour l'Éducation (DNE), notamment dans le cadre d'un projet européen portant sur la création d'espaces de données, un dispositif largement porté par des acteurs privés. Ce projet, Gaia-X, vise à construire une infrastructure européenne de données, fédérée et sécurisée. Elle y a contribué à des réflexions sur les modalités d'échange, de partage et de circulation des données au service de l'innovation. Cette expérience l'a

conduite à rejoindre l'INRIA début 2023, à la demande du ministère, pour piloter l'étude de préfiguration de la plateforme des données de l'éducation.

Depuis janvier 2024, elle travaille à la DNE, où elle est en charge des projets EduPilote et de valorisation des données.

Acteur institutionnel territorial - Hugues Labarthe

Hugues Labarthe est Délégué Académique Adjoint au Numérique (DAN adjoint) pour la région académique de Bretagne. Sa mission consiste à accompagner les établissements scolaires dans leur transformation numérique, en promouvant des pratiques pédagogiques innovantes et en facilitant l'intégration des outils numériques dans l'enseignement.

Avant d'occuper ce poste, Hugues Labarthe était chef de projet à l'Incubateur du Numérique de l'Académie de Créteil, une structure dédiée à l'innovation et à l'expérimentation pédagogiques. À ce titre, il a piloté des projets portant sur la co-conception d'environnements d'apprentissage, l'expérimentation de l'intelligence artificielle dans l'apprentissage adaptatif, la scénarisation pédagogique autour du jeu vidéo, ainsi que l'organisation du hackathon annuel « Du Labo à la Classe ». Il a également contribué à des travaux de recherche menés dans le cadre des Groupes Thématisques Numériques (GTnum).

Pour la Direction du Numérique pour l'Éducation (DNE), il est chargé de la conception, de la mise en œuvre et du pilotage de BRIO, la Bibliothèque de Ressources et d'Idées pour la réalisation des chefs-d'œuvre dans la voie professionnelle. Il œuvre également à intégrer la dimension citoyenne dans les projets numériques. Il a notamment accompagné des initiatives dans lesquelles les élèves exploitent les données ouvertes pour analyser des enjeux locaux ou mondiaux, et il organise le Wikidata Challenge pour l'année scolaire 2024-2025.

Concevoir un Education Data Hub

Préfiguration

Dès la rédaction du document-cadre, une phase de préfiguration était prévue, avec pour objectif de dialoguer avec l'ensemble des parties prenantes de l'écosystème éducatif, de définir clairement l'organisation de la future plateforme, ainsi que d'identifier l'AMDAC – son rôle au sein du ministère de l'Éducation nationale, et ce qui distingue ce ministère dans son rapport aux données.

Après deux phases d'étude préparatoire, la phase de préfiguration a débuté en novembre 2022. Tout au long de ce processus de développement, il n'a jamais été question d'un impératif de concevoir cette plateforme, mais plutôt d'une volonté d'interroger les parties prenantes sur la pertinence d'un tel dispositif. À l'issue de la phase de préfiguration, le ministère a décidé de ne pas créer de plateforme centralisée, mais d'apporter des réponses concrètes aux besoins exprimés lors des phases d'étude et d'auditions..

Étude de préfiguration I - avril à juillet 2020 - DNE

Au cours de cette phase initiale pilotée par la DNE, l'accent a été mis sur le contenu, avec des présentations du projet aux partenaires et la création de prototypes. Elle a permis de confirmer, à travers une dizaine d'entretiens menés auprès de partenaires du ministère (scientifiques, entreprises EdTech, entre autres), le diagnostic de la situation actuelle : des données dispersées, des impératifs de sécurité difficiles à gérer, des échanges peu fréquents, entraînant un manque de valorisation.

Cette étude a également permis de commencer à recenser les données susceptibles d'être mieux exploitées (traces d'apprentissage, données organisationnelles et de vie scolaire, données d'examen et de concours, données d'orientation, données sur les infrastructures numériques, données de gestion, statistiques, etc.) ainsi que les nouveaux usages qui pourraient en découler.

Étude de préfiguration 2 - mai à juillet 2021 - DNE and INRIA

Cela a permis de recentrer le projet autour de quatre axes principaux :

- Accompagner la constitution et la consolidation du patrimoine de données, en promouvant des standards d'interopérabilité, en proposant des clauses juridiques types, ou en produisant des jeux de données;
- Agir comme tiers de confiance dans l'ouverture et le partage de ces données;
- Soutenir l'ensemble de l'écosystème éducatif dans le développement de nouveaux usages des données partagées;
- Assurer le lien avec les élèves, les enseignants et les citoyens, par l'information, une acculturation progressive à l'usage des données, et la co-construction de services.

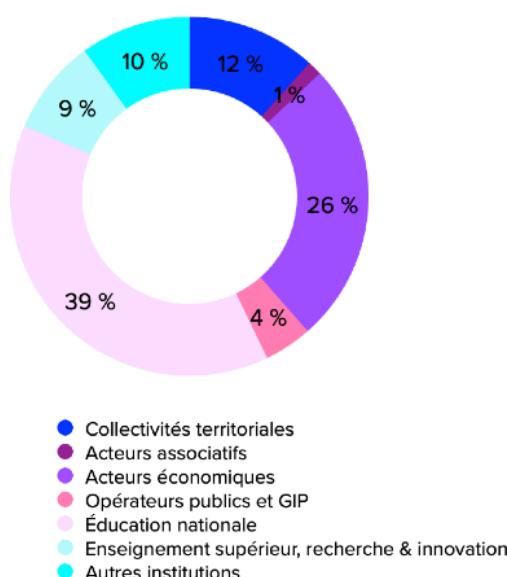
Il est rapidement apparu que, pour disposer de données fiables, il serait nécessaire d'organiser des entretiens à grande échelle.

Auditions - novembre 2022 à juin 2023

At Au départ, il était prévu d'interroger une centaine de personnes, mais au final, **231 personnes ont été auditionnées**, principalement lors de visioconférences spécialement organisées à cet effet, mais aussi parfois en présentiel, en marge d'autres réunions ou événements.

Il ne s'agit pas d'un panel conçu scientifiquement, mais l'objectif était d'être **le plus représentatif possible des parties prenantes collaborant avec le ministère de l'Éducation nationale sur les questions liées aux données**.

Les structures rencontrées se répartissent comme suit:



Références

- Bothorel, Eric, Stéphanie Combes, and Renaud Vedel, *Rapport Pour Une Politique Publique de La Donnée, Des Algorithmes et Des Codes Sources* (Gouvernement, 27 December 2020) <https://www.gouvernement.fr/upload/media/default/0001/01/2020_12_rapport_pour_une_politique_publique_de_la_donnee_-_23.12.2020_.pdf>
- *Doctrine Technique Du Numérique Pour l'éducation VI.1* (Ministère de l'Education nationale et de la Jeunesse, June 2023) <<https://eduscol.education.fr/document/49067/download?attachment>>
- Giraudon, Gérard, Pascal Guitton, Margarida Romero, Didier Roy, and Thierry Viéville, *Éducation et Numérique Défis et Enjeux* (INRIA, December 2020)
- *Loi Pour Une République Numérique* (Gouvernement, 7 October 2016) <<https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000031589829/>>
- Minichiello, Federica, Nicolas Roussel, and Philippe Ajuelos, *Plateforme Des Données de l'éducation: Rapport de Préfiguration* (INRIA - Ministère de l'Education nationale, 22 November 2023) <<https://inria.hal.science/hal-04443624>>
- *Numérique Éducatif Pour Des Territoires Inclusifs, Attractifs et Innovants* (Banque des territoires, November 2020) <https://www.banquedesterritoires.fr/sites/default/files/2020-11/20201102_Banque%20des%20territoires_R%C3%A9flexion%20sur%20le%20num%C3%A9rique%20%C3%A9ducatif.pdf>
- *Pour Une Politique de La Donnée Ambitieuse Au Service de l'École, de La Jeunesse et Des Sports* (Ministère de l'Education nationale, de la Jeunesse et des Sports, 27 April 2021)
- 'Projects4', IDEE <<https://www.idee-education.fr/donnesadmin>> [accessed 6 March 2024]
- *RAPPORT D'ACTIVITE DU COMITE D'ETHIQUE POUR LES DONNEES D'EDUCATION* (Ministère de l'Education nationale, de la Jeunesse et des Sports, September 2021) <<https://www.education.gouv.fr/media/94472/download>>
- Zablot, Solène, Khansa Ghabara, and Éric Bruillard, 'Collecte et traitement de données d'apprentissage Quelles Pratiques Des Fournisseurs de Ressources?', 2021 <<https://hal.science/hal-03838899/document>>



Co-funded by the
Erasmus+ Programme
of the European Union

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.