

Le cadre de l'expérience des données en éducation : gouvernance, représentations et intelligibilité des données dans l'éducation nationale

Anne LEHMANS, professeur en Sciences de l'information et de la communication, Université de Bordeaux, Laboratoire IMS CNRS 5218 équipe RUDII

Camille CAPELLE, maître de conférences en Sciences de l'information et de la communication, Université de Bordeaux, Laboratoire IMS CNRS 5218 équipe RUDII

Résumé : Analyse de la compréhension par les acteurs de l'Éducation nationale et leurs partenaires des enjeux d'organisation, de protection, de mobilisation et de médiation des données. Sur la base des premiers résultats d'une recherche sur la culture des données, les représentations exprimées montrent les difficultés auxquelles se heurtent les acteurs face à leurs missions dans le cadre de la gouvernance des données. En cause, leur intelligibilité des données, dans la mesure où celles-ci renvoient à des objets complexes, à des normes qui sont encore en construction dans leur mise en œuvre, et à des valeurs qui révèlent les contradictions des choix politiques en matière de gestion des données pour l'éducation.

Mots-clés : gouvernance des données, éducation, médiation des données, données personnelles, données ouvertes.

Abstract: Analysis of the understanding of the challenges of organizing, protecting, mobilizing and mediating data by education managers. Based on the first results of a research on data literacy, the representations underline the major difficulties faced by actors in the face of their missions. In question, their understanding of data, insofar as they refer to complex objects, to standards that are still under construction in their implementation, and to values that reveal the contradictions of political choices in terms of data management for education.

Key Words: data governance, education, data mediation, personal data, open data.

Introduction

L'Éducation nationale collecte, produit, utilise et diffuse massivement des données de natures variées. Beaucoup sont des données personnelles, dont certaines sont sensibles, concernant les élèves, leurs familles ou les enseignants. D'autres peuvent faire l'objet d'une mise à disposition publique dans un format ouvert, sur des plateformes d'*Open Data* dont l'une est spécifiquement dédiée à l'éducation (data.education.gouv.fr) et alimentée essentiellement par le ministère lui-même. Toutes ces données dites d'éducation sont gérées par différents acteurs concernés au niveau politique, administratif ou pédagogique. Elles font l'objet, depuis ces dernières années, d'une réflexion globale, sur les plans politique, juridique et éthique, ainsi que sur celui de leur traitement technique. En effet, le contexte actuel tend vers la globalisation des données et l'évolution de la "gouvernementalité algorithmique" qui repose sur la modélisation des comportements et de l'ordonnancement social, à travers notamment les techniques de hiérarchisation, de notation, d'appariement, de profilage et de ciblage (Rouvroy, 2021) qui traversent les usages de l'intelligence artificielle. Ces évolutions sont préoccupantes et rendent indispensable une interrogation sur la façon dont les acteurs qui sont à la source de la diffusion de données et décisionnaires de leurs usages possibles comprennent et interprètent ce qui est en jeu. Ces

acteurs se situent dans des schèmes d'intelligibilité¹ des données qui expliquent les régimes de leur engagement (Thévenot, 2006) et de leurs pratiques de gouvernance dans le contexte de l'éducation. Une tension se dessine entre un mouvement de normalisation, de fermeture et de contrôle d'un côté, dans une perspective de protection et de sécurisation, et un appel à l'ouverture de l'autre, dans une perspective de diffusion, d'usage libre, de partage, d'exploitation des données. Cette tension constitue le cadre de l'expérience des acteurs (Goffman, 1991) et structure l'intelligibilité des données. C'est ce cadre, situé, dans lequel des règles, des normes et des valeurs² sont instituées par les différents acteurs, que nous proposons de décrire et d'explicitier.

Nous analysons ici la façon dont les acteurs qui interviennent dans les choix de gouvernance des données de l'Éducation nationale et dans leur mise en œuvre, cadres, enseignants, juristes notamment, mais aussi leurs partenaires sur le terrain local, comprennent et interprètent les enjeux d'organisation, de protection, de mobilisation et de médiation des données. La gouvernance renvoie au pilotage, le mot partageant la même origine étymologique que la cybernétique (le pilotage des navires). Elle fait aussi référence à l'application de normes destinées à réguler la circulation de l'information et la communication dans des contextes à la fois réticulaires et contraints du point de vue de l'organisation. Le concept de gouvernance peut être associé à des valeurs. Il est question, dans les textes institutionnels, de "bonne gouvernance" (*good governance*), lorsque celle-ci permet la transparence, la responsabilité, l'effectivité, l'équité des politiques et leur conformité à la réglementation³. Les acteurs de la gouvernance assurent la cohérence des interactions entre créateurs, gestionnaires et utilisateurs de données ainsi que leur conformité aux normes.

Cet article porte sur la représentation de la gouvernance des données dans l'éducation⁴. Cette question est traitée à l'aune des représentations de divers acteurs concernés au sein du Ministère de l'éducation nationale, mais aussi de collectivités territoriales, et des discours institutionnels qu'ils mobilisent. Cette contribution vise ainsi, en premier lieu, à identifier les éléments de langage en circulation, les représentations sous-jacentes aux discours sur la gouvernance des données dans l'éducation, et les systèmes de valeurs qui les portent. Nous nous intéressons, en second lieu, aux difficultés majeures auxquelles sont confrontés les acteurs dans le cadre de leurs missions dans leur compréhension des données et les modes de communication associés. Ces difficultés renvoient à des processus complexes du point de vue des connaissances mobilisables, à des normes qui sont encore en construction dans leur mise en place, et à des valeurs qui révèlent les contradictions des choix politiques en matière de gouvernance des données pour l'éducation. Nous présentons ainsi, dans un premier temps, les enjeux de la gouvernance des données dans l'éducation, et, dans un second temps, les acteurs concernés et l'écosystème dans lequel leurs activités se déploient. Enfin, nous analysons les régimes de valeurs qui se dégagent de l'analyse des discours et structurent le cadre de l'intelligibilité des données.

Les données dans l'éducation : enjeux de gouvernance

Les données en éducation concernent tout autant le système dans son organisation que les activités d'enseignement et d'apprentissage. Elles sont hétérogènes, relèvent de plusieurs régimes juridiques qui recouvrent des enjeux de gouvernance à explorer.

La diversité des données en éducation

¹ Au sens de ce qui définit les lignes d'interprétation et les vecteurs d'analyse du réel, chez Berthelot (2004) par exemple.

² Sur le rapport entre ces notions dans le contexte de l'éducation, voir par exemple : Lehmans, A. (2018). Limites de la norme, constructions de la connaissance et durabilité de l'information. *Revue COSSI*, (5). https://doi.org/10.34745/numerev_1610

³ Voir par exemple OCDE (2021). *Cadre d'action en matière de bonne gouvernance publique : Éléments fondamentaux pour le bon fonctionnement des administrations publiques*, Éditions OCDE, Paris. <https://doi.org/10.1787/0ad89895-fr>

⁴ L'étude de la question de la gouvernance des données de l'éducation s'inscrit ici dans le cadre d'un projet de recherche consacré à l'analyse des représentations et usages des données pour l'éducation, la formation et l'innovation, et en constitue l'un des axes. Il s'agit du projet porté par l'équipe RUDII de l'IMS GTNum DEFI, groupe thématique numérique sur la littératie des données, *Education, Numérique, Recherche*, 08/03/2022. <https://edunumrech.hypotheses.org/4185>

Les données de l'éducation peuvent être définies comme les données liées à la vie scolaire (depuis l'administration jusqu'à la pédagogie) et concernant les acteurs de la communauté éducative. Elles sont produites, stockées et analysées pour des objectifs spécifiques, exploitées à des fins de suivi pédagogique des élèves, d'organisation et de pilotage du service public éducatif, d'élaboration de ressources pédagogiques ainsi que de statistiques d'évaluation et de recherches (Atal et Froidevaux, 2020). Elles concernent, en premier lieu, les élèves avec des données administratives et de scolarité (inscriptions, emplois du temps, absences, retards, redoublements, notes...) et leurs familles, par exemple avec les données sociales. Il s'agit également des données pédagogiques produites par les enseignants dans les activités d'enseignement (contenus et supports de cours, évaluations...). Il existe également des « données d'interactions » considérées comme fondamentales par le Comité d'éthique pour les données d'éducation : celles-ci sont produites dans les interactions entre les élèves et leurs enseignants à travers des documents ou sur les dispositifs numériques utilisés au cours de travaux collaboratifs entre les élèves, par exemple lors d'échanges sur l'évaluation d'un travail, ou encore dans les messages rédigés par un enseignant à destination d'une famille. Toutes ces données préexistent au numérique, mais leur circulation dans des plateformes de type Environnement Numérique de Travail (ENT), sous forme de traces numériques, permet d'en faciliter l'accès ainsi que la "retrouvabilité". L'usage des plateformes produit également un ensemble divers de données dites « comportementales », qui témoignent de la façon dont l'utilisateur se comporte sur la plateforme, à travers, par exemple, ses temps de connexion, ses actions, telles que le nombre de clics, de *scrolls*, ou de messages envoyés.... Ces données comportementales permettent ainsi de savoir précisément ce que font les acteurs, de quelle façon, et à quel moment. Le traitement des données, c'est-à-dire "toute opération portant sur des données (ici, personnelles), comme stocker, modifier, analyser ou croiser avec d'autres données, et quel que soit le procédé utilisé"⁵ est encadré par des procédures qui relèvent de choix politiques et juridiques. Le Ministère de l'éducation nationale, de la jeunesse et des sports indique sur son site "élaborer une gouvernance des données pour en garantir un meilleur traitement, collecte, circulation et valorisation"⁶.

Les enjeux de la gouvernance des données

Le principe de la gouvernance de l'information dans les organisations, pour lequel de nombreux travaux de recherche existent, s'applique aux données. Les acteurs des organisations privées, comme des services publics, ont progressivement pris conscience de la valeur des données qu'ils produisent. La question de la gouvernance des données est devenue un enjeu essentiel (Lehmans, 2018), complexe et critique, appelant un cadre fixant des règles de gestion liées au contrôle et à l'accessibilité de l'information dans une perspective de valorisation mais aussi de sécurisation des systèmes d'information, dans un écosystème politique ouvert (Maurel, Zwarich, Verlaet, 2022). La gouvernance correspond en effet à la mise en place d'un cadre négocié, complexe et dynamique, qui transforme la production, la perception et les usages des normes en s'appuyant sur des acteurs (experts, comités, acteurs de la société civile dans une logique de participation...) et des moyens organisationnels et juridiques nouveaux (normes "à exigence de réflexivité", recommandations, standards) (Bernatchez, 2020). Elle repose, dans la tradition canadienne par exemple, sur "une vision globale pour l'organisation, qui permet d'utiliser les connaissances à plein escient afin de mener l'organisation vers la destination souhaitée. Elle met l'accent sur les processus et offre un cadre pour mieux les exploiter" (Caron et al. 2020). Elle ouvre une perspective de maîtrise, de rationalisation et d'efficacité de l'organisation, doublée d'une utopie de transparence (le gouvernement ouvert), voire de participation. Dans les discours, la maîtrise de la gouvernance est aussi source d'innovation, et par rebond, éventuellement, de nouveaux profits.

Le traitement des données personnelles est encadré par la loi du 6 janvier 1978 et le règlement général européen sur la protection des données (RGPD). L'application de ce dernier depuis 2018 se traduit par la mise à disposition d'une documentation et de ressources en ligne produites par la Commission nationale de l'informatique et des libertés (CNIL). La collecte, le stockage, l'archivage, l'accès et

5- CNIL (sans date). RGPD : de quoi parle-t-on ?, *cnil.fr*. URL : <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on>

6- MENJS (2023). Le comité d'éthique pour les données d'éducation. *Education.gouv.fr*, février 2023. URL : <https://www.education.gouv.fr/le-comite-d-ethique-pour-les-donnees-d-education-12146>

l'utilisation des données concernant les élèves et leurs apprentissages représentent des enjeux éthiques et juridiques spécifiques. Le Ministère doit prévoir des actions de supervision et de médiation, dans chaque académie, notamment par l'intermédiaire d'un Délégué à la Protection des Données (DPD), censé être l'interlocuteur privilégié sur ces questions.

Cependant, des contradictions majeures apparaissent, entre la nécessité de continuité pédagogique pour les enseignants, les choix d'outils efficaces, familiers et ergonomiques pour enseigner, les projets se revendiquant de l'innovation éducative en provenance de l'industrie de l'éducation (la *EdTech*), et les politiques d'accès, d'usages et de stockage des données visant la sécurité des données. Elles risquent d'accentuer les freins aux usages numériques pédagogiques. D'autre part, la collecte et la production de données font l'objet d'une obligation de mise à disposition en accès ouvert (*open data*) pour permettre leur libre réutilisation, en application de la loi pour une République numérique de 2016. C'est ce que rappelle la feuille de route 2021-2023 relative à la politique des données, des algorithmes et des codes sources, élaborée par le Ministère de l'éducation nationale, de la jeunesse et des sports⁷. Cette ouverture a des incidences fortes sur la gouvernance des données dans les services qui doivent adapter leurs modes d'organisation à l'obligation d'ouverture, dans un environnement chargé de données personnelles et sensibles⁸. Elle mobilise aussi des efforts de formation et de médiation pour qu'elle ait un effet réel. Pour l'Éducation nationale, la mise en conformité des pratiques avec les normes suppose une transformation des modes d'organisation par rapport au cycle de vie des données, et aux usages des systèmes d'information et de communication, du point de vue de l'administration. Elle peut aussi avoir des conséquences sur l'agenda politique dans la mise en relief, par exemple, des inégalités scolaires. L'obligation de publier l'Indice de position sociale (IPS) des écoles et collèges n'a ainsi pu être mise en application qu'à la suite de la décision du Tribunal administratif de Paris du 13 juillet 2022. Celui-ci avait été saisi par un journaliste de la Gazette des communes, Alexandre Léchenet, à qui la Commission d'accès aux documents administratifs avait donné raison dans un premier temps pour exiger cette publication, dans la perspective de comprendre, notamment, l'algorithme d'attribution des places dans les établissements par secteurs⁹. Le choix des données à publier reste donc stratégique et sensible, si l'on considère leurs usages possibles dans une perspective de mesure de la performance et de classement par exemple. Du point de vue des acteurs que sont les enseignants, la formation initiale et continue doit assurer la connaissance et la maîtrise des outils numériques employés et leur conformité avec les normes, le RGPD en tête. Les dimensions éthiques, politiques et juridiques font désormais partie de la culture numérique attendue chez les enseignants (Cordier et al., 2018), à travers le cadre de référence des compétences numériques pour l'éducation (CRCN-Édu), et indirectement chez les élèves (CRCN) via la plateforme PIX qui évalue leurs connaissances et compétences numériques. Cette évolution des contours de la culture numérique est liée tout autant à l'exigence de connaissance des règles juridiques qu'à celle de développement de l'autonomie et de la capacité à agir. Ces exigences sont au cœur de débats qui engagent des normes mais aussi des représentations.

Explorer les représentations des données et de leur gouvernance : une enquête qualitative

Notre enquête explore la façon dont les acteurs, dans leurs interactions contextualisées, définissent et comprennent les enjeux de la collecte, la production, du stockage, de la mise à disposition et la réutilisation des données qui relèvent du champ de l'Éducation nationale. Nous faisons l'hypothèse que la gouvernance engage des enjeux politiques en termes de transparence pour l'administration de l'éducation et de respect des libertés individuelles, mais aussi organisationnels dans le sens d'une

7- MENJS (2023). La feuille de route 2021-2023 relative à la politique des données, des algorithmes et des codes sources, *education.gouv.fr*. <https://www.education.gouv.fr/la-feuille-de-route-2021-2023-du-ministere-de-l-education-nationale-de-la-jeunesse-et-des-sports-sur-340319>

⁸ Voir CNIL, 2022, Rapport annuel : Agir pour un futur numérique responsable, https://www.cnil.fr/sites/cnil/files/2023-05/cnil_-_43e_rapport_annuel_-_2022.pdf

⁹ Voir <https://www.politico.eu/article/coulisses-ips-indicateur-secoue-mixite-ecole-pap-ndiaye/>

amélioration et d'une rationalisation de son fonctionnement. Une autre hypothèse est que les données, dont une grande partie est issue des activités des élèves, sont comprises comme pouvant être réutilisées dans un cadre éducatif pour la construction de connaissances par les enseignants et les élèves. Nous cherchons donc à identifier les politiques et les moyens de gouvernance mis en œuvre concernant les données de l'éducation :

- dans leur production, du point de vue des acteurs concernés et de leurs représentations,
- dans leur gestion, du point de vue du cycle de vie des données,
- dans leur utilisation, du point de vue des usages potentiels et des perspectives d'innovation perçues et projetées, dans le contexte éducatif, depuis les établissements scolaires jusqu'aux services centraux du Ministère de l'éducation nationale.

La représentation des données par les acteurs de l'éducation, évoquée ici, constitue une première entrée dans la question de leur gouvernance. Elle s'appuie sur la délimitation, la collecte et l'analyse de discours sur les données mobilisant des documents institutionnels et des ressources à destination des acteurs de l'Éducation nationale. La perspective sémio-pragmatique permet d'interroger les représentations telles qu'elles apparaissent dans les discours en faisant le lien entre la construction du sens en contexte et ses effets, dans la tradition portée par Roger Odin (2011). Le contexte d'énonciation des discours des acteurs constitue une base pour l'analyse, qui permet de dessiner le cadre de leur expérience et de le mettre en relation avec les normes et les valeurs mobilisées. Une série d'entretiens semi-directifs menés auprès d'acteurs concernés par la gouvernance des données à différents niveaux hiérarchiques des services centraux et déconcentrés du Ministère de l'éducation nationale, ainsi que des établissements scolaires et des collectivités territoriales, permet de saisir la façon dont ils comprennent et interprètent ce cadre. La méthodologie de recherche est qualitative et compréhensive. Sur la base des premiers résultats de ce travail, nous analysons les éléments de langage en circulation, dans une approche à la fois lexicale et thématique, les représentations sous-jacentes aux discours et les systèmes de valeurs qui les portent.

Pour cette étude en cours, nous avons interrogé 15 acteurs impliqués dans la gouvernance des données du point de vue de la gestion, de la surveillance et/ou de la médiation : l'administrateur des données au sein de la Direction du Numérique pour l'Éducation (DNE), des Délégués Régionaux Académiques au Numérique Éducatif (DRANE) de régions différentes, des membres du Comité d'éthique pour les données d'éducation et de la CNIL, des ingénieurs *Open Data* travaillant dans deux métropoles qui développent une politique active d'ouverture et de médiation des données dans une perspective d'éducation, des responsables des opérateurs de médiation que sont le Réseau Canopé et le Comité de liaison de l'éducation et des médias d'information (CLEMI). Nous avons également pu recueillir les propos de trois délégués à la protection des données, deux référents pour les ressources et usages pédagogiques numériques (RRUPN) et deux chefs d'établissements, dans trois académies différentes. Ces acteurs sont positionnés à des niveaux très différents de la gouvernance des données, et porteurs de cultures professionnelles hétérogènes. On peut relever trois cadres de la construction identitaire et sociale a priori : la culture de l'enseignement, la culture informatique, et la culture managériale. Nous nous appuyons sur une analyse de discours et des dispositifs de communication à partir d'un corpus de transcriptions d'entretiens et de documents collectés auprès des acteurs (documents ressources de référence, pages web institutionnelles, documents d'information ou de médiation...). Il s'agit d'« appréhender le discours comme articulation de textes et de lieux sociaux » (Maingueneau, 2021), en nous appuyant sur les registres mobilisés, ainsi que les dimensions spatio-temporelles, lexicales et thématiques. La manière dont les acteurs font référence à quatre éléments clés est explorée, en lien avec notre recherche : la gouvernance des données de l'éducation, les enjeux perçus de cette gouvernance, les usages des données de l'éducation (actuels ou futurs), l'éducation aux données. Les codifications thématiques établies permettent de mettre en évidence les représentations des acteurs, à partir de différentes postures, de l'engagement et de l'affect exprimé par les acteurs, et d'occurrences lexicales dessinant des champs lexicaux relevant de colorations techniques (outils), politiques (normes et missions assignées) ou éthiques (valeurs mobilisées).

Représentations et intelligibilité des données dans l'éducation nationale

La gouvernance des données de l'Éducation nationale s'organise autour d'une diversité d'acteurs des secteurs public, privé et associatif. Ces acteurs interviennent auprès des publics de l'Éducation nationale, ainsi que du grand public, dans une visée plus large de communication et d'affichage de transparence. Certains agissent pour la gouvernance des données, d'autres davantage en faveur d'une médiation des données, chacun étant en interaction au sein d'un écosystème.

Les acteurs que nous avons identifiés peuvent être regroupés en quatre pôles. Le premier pôle est centré sur la protection des données, à l'intersection entre l'administration de l'Éducation nationale et la CNIL. Dans ce pôle, des acteurs de l'éducation sont impliqués, comme la DNE, les DRANE, ainsi que les DPD. Des dispositifs sociotechniques (Coutant et Domenger, 2014), tels que le Gestionnaire d'Accès aux Ressources (GAR) ou l'*Education Data Hub*¹⁰, répondent à des exigences légales (notamment de protection des données personnelles), techniques et organisationnelles qui traduisent des stratégies et des représentations sur la manière dont les données sont gérées et mises à disposition des différents acteurs de l'éducation (élèves, familles, enseignants, éditeurs de ressources pédagogiques...). Un deuxième pôle est celui de l'ouverture des données, depuis la création de la mission Etalab en 2011 et de la plateforme Data.gouv.fr, avec l'obligation pour les collectivités territoriales et les organisations gérant des missions de service public de publier les données dans la loi pour une République numérique de 2016. Les données de l'éducation, longtemps restées dans une nébuleuse peu lisible, sont désormais accessibles et mises en visibilité sur une plateforme dédiée, data.education.gouv.fr. Le Ministère de l'enseignement supérieur et de la recherche s'inscrit également dans ce pôle, mettant en avant la démarche d'ouverture des données de la recherche dans ses discours et ses incitations politiques, dans une tradition antérieure à celle de l'Éducation nationale.

Un troisième pôle rassemble les acteurs qui œuvrent pour l'éducation aux données avec certains acteurs de l'Éducation nationale (des DRANE notamment et la DNE, le réseau Canopé, le CLEMI), quelques Instituts Nationaux Supérieurs du Professorat et de l'Éducation (INSPE), ainsi que des associations d'éducation populaire telles que les Petits débrouillards, la Ligue de l'enseignement, la Fédération des centres sociaux ou encore Fréquence École. Un quatrième pôle émerge, entre le monde de la recherche (notamment avec la valorisation de l'informatique appliquée et des disciplines rattachées aux neurosciences, mises en avant dans les discours scientifiques sur l'éducation, par exemple au sein du Conseil scientifique de l'éducation nationale) et celui des entreprises liées aux technologies pour l'éducation (*EdTech*), dont PIX, *start-up* d'État, fait partie. Il traite notamment de l'exploitation des données au service de la modélisation cognitive pour l'éducation, en partie autour des perspectives de l'intelligence artificielle pour les apprentissages. Le Ministère de l'éducation nationale collabore notamment avec l'INRIA (Institut national de recherche en informatique et en automatique, rebaptisé Institut national de recherche en sciences et technologie du numérique), particulièrement sollicité et chargé de la conception et de la valorisation de la plateforme de données de l'éducation.

Entre ces quatre pôles, différents acteurs interviennent ou sont consultés pour participer à la réflexion sur la gouvernance des données. C'est le cas de plusieurs collectifs comme le Conseil scientifique de l'éducation nationale, le Comité d'éthique des données de l'éducation, également mis en place par le Ministère de l'éducation nationale¹¹, le collectif Educnum qui regroupe un ensemble d'organisations publiques et privées, des entreprises de la *EdTech*, des associations comme l'An@é. Sur la médiation des données, d'autre part, les RRUPN, des Inspecteurs de l'Éducation Nationale (IEN, IA-IPR) ainsi que les Enseignants Référents aux Usages du Numérique (ERUN) dans l'enseignement primaire et les Interlocuteurs Académiques du Numérique (IAN) dans l'enseignement secondaire, interviennent dans la mise en œuvre des choix politiques en les traduisant diversement sur le terrain. L'écosystème dans

¹⁰ Plateforme de données de l'éducation et de l'enseignement, proposée à la suite des états généraux du numérique pour l'éducation réunis en novembre 2020, qui vise à coordonner et structurer les données en provenance de différents acteurs, dont les collectivités territoriales, pour les mettre à disposition. Le projet est mené en collaboration avec l'INRIA.

¹¹- Comité créé en 2019, à mission consultative, composé de 13 membres, chercheurs, parlementaires, acteurs du monde économique et social.

lequel évoluent ces acteurs dessine le contexte de l'analyse des discours qu'ils produisent. Celle-ci révèle en partie les tensions qui se jouent entre les exigences normatives et de protection d'un côté, d'ouverture et de transparence de l'autre. Entre les deux, la nécessité d'éduquer au et par le numérique suppose des choix dans le recours à des outils numériques qui ont nécessairement des incidences en matière de production, de gestion et de protection des données de l'éducation.

Des régimes en tension : approche par la valeur vs approche par le risque

Si la mise en relief des risques liés aux usages des données a longtemps été de mise dans les discours en et sur l'éducation (Capelle, Liquète, 2022), elle est contrebalancée aujourd'hui par la valorisation des potentialités des usages des données. Celles-ci deviennent des ressources, qui restent néanmoins soumises à un contrôle centralisé dans le cadre de systèmes de valeurs plastiques.

Les données comme ressources politiques

Les représentations des données en éducation à travers leur gouvernance cristallisent des enjeux politiques, non seulement sur le plan des principes, mais aussi sur le plan de l'action. Les acteurs interrogés évoquent une responsabilité collective, pluri-catégorielle et qui ne se joue pas seulement en France mais au niveau européen, concernant les données de l'éducation. Dans le cadrage politique national tel qu'il ressort des états généraux du numérique pour l'éducation menés en 2020 (et dont les traces ont disparu sur le site du Ministère), la politique des données vise surtout à "faire avancer la recherche en éducation, éclairer les décisions publiques et construire des services plus performants"¹². Alors que de plus en plus de logiciels sont utilisés pour collecter, stocker, traiter ces données, et que ces dernières se massifient, les acteurs évoquent une prise de conscience récente et soudaine, accélérée par la crise sanitaire du Covid-19, de la valeur des données : "*la donnée c'est le terreau qui fera pousser les futurs arbres* " pour l'un des acteurs interrogés. Elle est perçue comme une ressource au cœur des enjeux politiques en éducation, à laquelle il est nécessaire d'acculturer le public, élèves, enseignants, parents, dans une visée pragmatique d'amélioration du service public de l'éducation d'une part, mais aussi dans une perspective de construction de l'autonomie, de libération, d'inclusion ou d'employabilité et de compréhension des enjeux notamment autour des communs numériques.

Les données comme ressources contrôlées

Si la liberté est valorisée du point de vue des usages possibles, le contrôle centralisé de la gestion demeure un point central dans les discours des acteurs. Pour l'un d'eux "*il n'est pas question qu'un tiers autre que le ministère ait les informations sur les élèves*". La thématique de la souveraineté des données, apparue très nettement éclairée ces dernières années, renvoie à des valeurs politiques qui ne sont plus seulement liées aux questions de droits individuels et de libertés (la protection des données personnelles contre les risques d'utilisation par les GAFAM notamment), mais bien plus aux équilibres stratégiques internationaux (le poids de la France et de l'Europe dans le monde). Dans le discours politique interne, la thématique de la souveraineté résonne favorablement pour une partie de l'électorat et des partis politiques. On observe alors des variations dans les prises de position entre acteurs publics et privés. Le Ministère de l'éducation nationale, contrairement au Ministère de l'enseignement supérieur et de la recherche, qui contrôle systématiquement les données de la recherche et met en valeur la science ouverte depuis plusieurs années, reprend le contrôle sur des données qui lui échappaient : c'est le cas par exemple des données collectées dans le logiciel Pronote, détenu par le groupe Index Education. Alors que les établissements scolaires du secondaire se sont abonnés en masse à cette solution pratique de gestion de la vie scolaire, les acteurs de la gouvernance se sont aperçus que ces données ne pouvaient pas rester la propriété d'une entreprise privée. Dans sa stratégie visant à reprendre le contrôle sur les données, le Ministère investit également dans la conception et la réalisation de solutions numériques à destination des enseignants et des familles. C'est le cas du portail

12- Cité par le rapport de la mission Bothorel « Pour une politique publique de la donnée », décembre 2020, p. 118.

apps.education qui a vocation à remplacer les outils privés, hébergeant souvent les données en dehors de la France et même de l'Europe, que les enseignants utilisent encore massivement.

La plupart des acteurs interrogés s'inquiètent de l'information en circulation dans l'espace public et de l'image politique renvoyée par la problématique du traitement et de la gouvernance des données. Sur cette question, certains jouent un rôle de lanceurs d'alerte, comme l'association Human Rights Watch, citée par plusieurs interviewés, et dont le rapport 2022 sur les produits d'apprentissage en ligne utilisés pendant la pandémie de 2020 a mis l'accent sur les risques de voir les données personnelles des enfants et des jeunes exploitées par les entreprises en l'absence de vigilance de la part des gouvernants. Avant même cette alerte médiatisée, un rapport de la Cour des comptes de 2019¹³ soulignait la puissance des stratégies d'influence de la filière industrielle et la nécessité d'une vigilance sur la protection des données personnelles. L'exemple de Pronote est cité dans le rapport qui mentionne une situation de quasi-monopole d'une entreprise privée sur les données de vie scolaire et les risques encourus pour leur sécurité. À la suite de cette alerte, relayée par la presse¹⁴, le Ministère de l'éducation nationale a fait racheter la majorité des parts d'Index Education par Docaposte, filiale de La Poste. Sur le site d'Index Education, on observe une stratégie de communication à destination des parents d'élèves et visant à rassurer le public quant à la protection des données. Le groupe met en avant le fait qu'il dispose de son propre serveur sécurisé pour héberger les données depuis 2010. Il se présente aussi comme le « premier hébergeur des données de l'éducation ». Pronote reste l'emblème des difficultés à concilier des principes contradictoires : ergonomie et accessibilité du point de vue des usages ; sanctuarisation et garantie de la sécurité des données personnelles par l'État.

Dans la gouvernance, les dimensions normatives et contraignantes sont contrebalancées, du point de vue du l'image et de la communication, par la valorisation des données dans un discours d'accompagnement et de formation aux usages qui construisent le socle de la "citoyenneté numérique". C'est donc la valeur des données qui est mise en avant plutôt que les dangers de leurs usages. Cependant, sur le terrain, cette communication ne semble pas toujours atteindre ses objectifs. La représentation de la gouvernance des données est plutôt centrée sur la contrainte et les risques du point de vue des pratiques des enseignants. Elle devient un moyen de contrôle de leurs activités, un contrôle pourtant vécu comme défaillant par ses acteurs eux-mêmes. Les DPD se sentent investis d'une mission pour laquelle ils ne disposent pas des moyens suffisants ni de la confiance qu'ils appellent de leurs vœux pour assurer leur mission. Les chefs d'établissements sont parfois tiraillés entre des injonctions contradictoires et qu'ils peinent à comprendre. La coordination des activités et des discours est loin d'être claire, comme celle des services dont les systèmes de valeurs, les cultures professionnelles de référence et les outils sont isolés les uns des autres.

Une intelligibilité plastique

De nombreux acteurs, anciens (comme la CNIL) ou nouveaux (comme le Comité d'éthique pour les données d'éducation) interviennent sur la scène publique des données en éducation et dessinent un horizon idéologique fortement orienté vers les questions éthiques, composant la "bonne conscience" de la gouvernance. Dans la mission de ces acteurs, l'accent est mis sur les bonnes pratiques, les codes de conduite, l'accompagnement, le conseil et la formation, les outils de communication, beaucoup plus que sur les contraintes. Cet effacement de la contrainte (et donc de force obligatoire) se double d'une absence d'évaluation des actions, mentionnée par plusieurs acteurs chargés d'expertise, de conseil ou d'accompagnement, et parfois d'une impossibilité de situer leur action à un moment qui lui donnerait une certaine force (en amont des décisions politiques notamment). Un acteur souligne même que c'est l'ensemble de l'activité ou de la ressource construite sur des données, depuis le fait même d'y recourir, qui n'est pas vraiment évalué.

Ainsi, des concepts clés ressortent des discours dans leurs références au cadre institutionnel, mais ces concepts ne s'appuient pas sur une doxa clairement identifiée dans son régime de vérité. Ils sont plastiques et susceptibles d'adaptations interprétatives du point de vue de l'intelligibilité, selon les

13- Cour des Comptes (2019). Le service public numérique pour l'éducation. Un concept sans stratégie, un déploiement inachevé. *Ccomptes.fr*, Juillet 2009.

14- Tourette, L. (2022). La vie scolaire sous Pronote. *Le Monde diplomatique*, janvier 2022, p. 20-21

cultures de référence des acteurs. Les concepts les plus récurrents sont ceux de *communs du numérique* et de *souveraineté*, considérés comme des principes de base de la gouvernance des données en éducation. Du côté des communs, l'horizon est ouvert, il s'agit de mettre en évidence les valeurs de partage, d'égalité, d'accessibilité, mais aussi d'utilisabilité ouverte, y compris dans des perspectives d'innovation industrielle et d'amélioration des outils pour enseigner. Du côté de la souveraineté, l'horizon est fermé, il s'agit de mettre en relief le nécessaire contrôle de la puissance publique. Cela notamment vis-à-vis des puissances étrangères, pour préserver les usages commerciaux, et dans le cadre de la guerre de l'information que sont susceptibles de se livrer les États, mais aussi vis-à-vis des puissances économiques intérieures.

Du point de vue des méthodes de gouvernance, les collectivités territoriales qui ont travaillé sur la question de l'ouverture des données depuis longtemps, comme Rennes Métropole, s'appuient sur leur expérience pour situer leur activité dans une vision écosystémique qui n'est pas encore véritablement de mise dans l'Éducation nationale, et qui n'inclut pas totalement celle-ci au niveau local. Cette perspective s'ancre dans une organisation très ouverte des métiers et des compétences autour de la donnée, qui intègre les dimensions techniques liées à la gestion des systèmes d'information mais aussi les dimensions communicationnelles et managériales, en s'appuyant sur des méthodes qui valorisent la négociation : *“acculturation, médiation, accompagnement au changement”*, selon un acteur d'une collectivité. Dans l'Éducation nationale, ces méthodes sont également préconisées. Mais leur mise en œuvre reste freinée par une organisation hiérarchique et cloisonnée qui crée une distance entre les promoteurs d'une politique de la donnée et les acteurs de terrain. Certains acteurs engagés et militants du réseau Canopé, chargé de la médiation, ont construit une solide expérience dans le domaine, par exemple autour des « *datasprints* » pédagogiques. Ils peinent à mobiliser les enseignants de terrain sur un objet qui reste très lointain dans les imaginaires collectifs et marqué par la technicité, les compétences mathématiques ou l'image de l'interdit associé aux données personnelles. Les acteurs mettent en avant un système de contraintes organisationnelles et normatives, qui passe par les programmes, les obligations réglementaires, le souci de sécurité, le manque de moyens et de temps.

Ce constat de la difficulté d'installer un écosystème cohérent autour de la donnée, pour les services internes de l'administration, mais aussi dans les relations avec les collectivités territoriales et les partenaires, est fait par l'institution elle-même. Un rapport de l'Inspection générale de l'éducation, du sport et la jeunesse, daté de 2022¹⁵, par exemple, revient sur la feuille de route numérique ministérielle de septembre 2021, sur la spécificité de la “donnée scolaire”. Il souligne la difficulté pour l'administration centrale de mettre en place une doctrine cohérente ainsi que l'extrême hétérogénéité des situations locales. Les entretiens confirment cette hétérogénéité et le sentiment de “bricolage” des acteurs qui sont pourtant souvent fortement investis dans une forme de mission tout en mentionnant le manque de formation et l'absence de médiation. Entre les injonctions politiques nationales, et les réalités de la perception et des pratiques de terrain, certains territoires semblent se démarquer. C'est le cas de la Bretagne : les entretiens avec des acteurs de ce territoire montrent qu'ils sont très mobilisés dans les collectivités territoriales, les universités, l'Éducation nationale, le tissu associatif, et qu'ils partagent une représentation basée sur les valeurs de coopération et de participation. Ces valeurs se retrouvent dans leurs discours et dans les dispositifs cités (coopératives pédagogiques numériques, Territoire numérique éducatif...). Les discours valorisent le principe d'une approche de terrain, qui part des besoins des acteurs, et s'appuie sur une stratégie de communication visant à *“rendre visible sans imposer”*. Cette approche repose aussi sur la construction de représentations communes, car *“le levier qui est intéressant, c'est quand se rejoignent à la fois le public des élèves, des parents et des enseignants, sur un degré de connaissance qui est le même”*, nous dit un cadre de l'Éducation nationale.

Conclusion

15- Quéré, M., Costantini, N., Delécluse, Y. (2022). *Les conventions d'échange de données entre les académies et les collectivités territoriales : état des lieux, pratiques et préconisations*. Rapport de l'Inspection générale de l'éducation, du sport et de la recherche, Octobre 2022.

Finalement, il ressort de cette tentative de compréhension de la gouvernance et de l'intelligibilité des données, que n'existe pas encore une culture commune et partagée des données, malgré les efforts de communication politique et les évolutions de la perception des enjeux des données. Les acteurs accordent une attention certaine et grandissante aux données depuis quelques années, mais les considèrent dans le cadre de systèmes de contraintes et de valeurs liés à des représentations et à des imaginaires de la technique qui restent hétérogènes. Le souci d'éduquer aux données, sans réserver cette éducation aux élèves des filières destinées aux métiers de l'ingénierie informatique, ne se traduit pas toujours dans des pratiques effectives de littératie de l'information. La perspective de démocratisation de l'accès aux communs du numérique et de la connaissance, qui croise ouverture des données et éducation à leurs usages, semble même s'effacer au profit des promesses de l'intelligence artificielle (Minichiello, 2022). Celle-ci absorbe discours et projets sur les données en posant de nouveaux défis à la question de la gouvernance.

Bibliographie

ATAL Ignacio et FROIDEVAUX Christine, 2020, *Enjeux d'éthique des usages des données numériques d'éducation dans le contexte de la pandémie*, Avis n°2020-1. Comité d'éthique des données d'éducation, Ministère de l'Éducation Nationale, de la Jeunesse et des Sports.

BERNATCHEZ Stéphane, 2020, Le droit en transition : le droit de la gouvernance et le paradigme cybernétique, dans A. Bailleux, *Le droit en transition : les clés juridiques d'une prospérité sans croissance*, Bruxelles, Presses universitaires de Saint-Louis.

BERTHELOT Jean-Michel, 2004, *Les vertus de l'incertitude*. Paris, Presses Universitaires de France, p. 45-85.

CAPELLE Camille, LIQUETE Vincent, 2022, *Perceptions et analyses des risques numériques*, Londres, ISTE Group.

CARON Daniel J., BHERER Harold, et BERNARDI Sara, 2020, *La gouvernance informationnelle au sein de l'administration publique*, Chaire de recherche en exploitation des ressources informationnelles, ENAP.

CORDIER Anne, CAPELLE Camille, LEHMANS Anne, 2018, « L'éthique du numérique » au sein de l'éducation à l'information : regards et pratiques d'enseignants en début de carrière, dans L. Balicco, *L'éthique en contexte info-communicationnel numérique : Déontologie, régulation, algorithme, espace public*, De Boeck Supérieur, p. 141-151.

COUTANT Alexandre, DOMENGET Jean-Claude, 2014, Un cadre méthodologique pour enquêter sur les dispositifs socio-techniques d'information et de communication, dans H. Bourdeloie et D. Douyère, *Méthodes de recherche sur l'information et la communication : regards croisés*. Paris, Mare & Martin, p. 231-254.

GOFFMAN Erving, 1991, *Les cadres de l'expérience*, Paris, Ed. De Minuit.

HASSANI Nadia, 2019, Le paradoxe de la protection des données personnelles à l'heure de la libre circulation des informations. *Terminal*. 124, <https://doi.org/10.4000/terminal.4040>

ISAAC Henri, 2022, Quelle souveraineté numérique européenne ? *Revue française de gestion*, 48, 305, p. 63-78.

KRIEG-PLANQUE Alice, OGER Claire, 2010, Discours institutionnels. Perspectives pour les sciences de la communication. *Mots. Les langages du politique*, 94, p. 91-96. <https://doi.org/10.4000/mots.19870>

LECLERE Fabrice, DESFRICHES-DORIA Orélie, 2019, Intégrer le RGPD à la politique de l'établissement et travailler en réseau : le rôle du délégué à la protection des données à l'Université. *I2D - Information, données documents*. 1, 1, p. 51-54.

LEHMANS Anne, 2018, L'horizon d'une culture de la donnée ouverte : de l'utopie aux pratiques de gouvernance des données. *Revue COSSI : communication, organisation, société du savoir et information*.

MAINGUENEAU Dominique, 2021, *Discours et analyse du discours : Une introduction*. Paris, Armand Colin.

MAUREL Dominique, ZWARICH Natasha, VERLAET Lise, 2022, Introduction : Pouvoir et stratégie des acteurs de la gouvernance de l'information, *Communication & management*, 19, p. 5-10.

MINICHIELLO Federica, 2022, Les défis pour utiliser les données en éducation, *Revue internationale d'éducation de Sèvres*, 91, décembre 2022, <http://journals.openedition.org/ries/13121>

ODIN Roger, 2011, *Les espaces de communication. Introduction à la sémio-pragmatique*. Grenoble, Presses universitaires de Grenoble.

ROUVROY Antoinette, 2021, Les interstices entre le monde et les représentations, entre les sujets et les normes, nous permettent de mettre les représentations et les normes à distance, de les discuter de manière collective, *Parole de*, Conseil National du Numérique.

THEVENOT Laurent, 2006, *L'Action au pluriel. Sociologie des régimes d'engagement*, Paris, Éditions La Découverte.