

LA CULTURE DES DONNÉES OUVERTES ET L'ENSEIGNEMENT

Marthe-Aline Jutand¹, Camille Capelle²

¹ ISPED université de Bordeaux, 146 rue Léo Saignat, 33076 Bordeaux cedex

Laboratoire CeDS (Cultures et Diffusion des Savoirs) (EA 7440) marthe-aline.jutand@u-bordeaux.fr

²IMS UMR 5218 CNRS Université de Bordeaux, 351 Cours de la Libération, 33400 Talence, camille.capelle@u-bordeaux.fr

Résumé. Après que les industriels se soient emparés du Big Data, il en revient maintenant à la sphère sociale de s'emparer de l'Open Data selon Gurstein (2011). L'institution scolaire et les enseignants ont un rôle à jouer pour permettre aux élèves, en tant que futurs citoyens, d'appréhender la question des données, de comprendre leurs contenus et le processus ayant permis de les produire, et ainsi de développer les compétences nécessaires pour en faire usage. L'accès à la pluralité des données ouvertes, tant en termes de quantité que de qualité, nous conduit à nous interroger sur les usages pédagogiques qui peuvent en être faits. Il ne s'agit plus seulement d'apprendre à constituer des données pour valider des hypothèses, mais d'interroger la structure et le contenu de données préexistantes pour se les approprier et pour construire de nouveaux savoirs. L'analyse des scénarios pédagogiques construits par les enseignants et des savoirs mis en jeu pour les élèves, nous conduit à mettre en évidence les besoins en matière d'offre de données ouvertes pour permettre leur appropriation citoyenne.

Mots-clés. Données ouvertes, open data, scénario pédagogique, formation aux données, data culture

Abstract. After industrialists seized Big Data, it is now up to the social sphere to take over the Open Data (Gurstein, 2011). The teachers have a role to play in enabling students, future citizens, to understand the question of data, to understand their content and the process by which they have been produced, and thus to develop necessary skills to make use of. Access to the plurality of open data, both in terms of quantity and quality, leads us to question the pedagogical uses that can be made of it. It is no longer a question of learning how to build up data to validate hypotheses, but of questioning the structure and content of pre-existing data in order to appropriate them and to build new knowledge. The analysis of the pedagogical scenarios constructed by the teachers and of the knowledge brought into play for the pupils leads us to highlight the needs for the provision of open data in order to allow them to be appropriated by the citizens.

Keywords. Open data, pedagogical scenario, data training, data culture

1 Introduction

L'accès aux données publiques sous un format ouvert, réutilisable, communément appelé Open Data, invite les citoyens à s'en saisir et à les exploiter comme *communs de la connaissance* selon Hess et Ostrom (2007). Ces données numériques reflètent notre quotidien sous forme « éclatée, segmentée, distribuée, décontextualisée, déshistoricisée » comme le souligne Rouvroy (2014). Il s'agit de données, la plupart du temps, chiffrées, de relevés d'informations, de mesures d'objets, de réponses à des enquêtes, de statistiques, de comptages, tout un ensemble de données collectées par les organismes publics ou privés et mises à disposition en format numérique sur des plateformes nationales ou locales. La mise en relation, le regroupement et/ou la

confrontation de différentes sources de données permettent de répondre à certaines questions posées et d'explorer des pistes de recherche. Le mouvement de l'Open Data ouvre d'intéressantes possibilités ; en facilitant l'accès et en permettant la mise à disposition gratuite des données produites par les services publics, chacun peut accéder à des données, les mettre en regard avec d'autres données, ou produire des résumés d'informations à confronter à d'autres informations, tout ceci afin de construire du sens, de nouvelles connaissances, pour, par exemple, améliorer des prises de décisions ou pour produire de nouvelles connaissances.

L'intégration des données ouvertes, comme ressources pédagogiques, ou comme concept à diffuser, reste pour l'instant très marginale dans le cadre de l'enseignement ou de la formation. Certains enseignants construisent des séquences pédagogiques mettant en œuvre ce nouvel apprentissage d'usage et de manipulation de données ouvertes. La construction de ces séquences et la mise en œuvre en classe correspond à notre objet d'étude. Après avoir exposé l'enjeu d'une approche des données ouvertes pour former les futurs citoyens à une culture et aux usages des données, nous définirons ce que nous désignons par scénarios pédagogiques. La méthodologie choisie pour explorer les stratégies pédagogiques de formation aux données ouvertes sera ensuite exposée, ainsi que les premiers résultats de la recherche. Celle-ci nous conduit à proposer une typologie de scénarios pédagogiques possibles autour des données ouvertes et à mettre en évidence des besoins de mise en forme des données ouvertes en vue de permettre leur appropriation citoyenne et de favoriser la diffusion d'une culture des données.

2 Les données ouvertes et les usages pédagogiques

L'ouverture des données publiques pose avant tout des questions de choix et de stratégies qui reposent sur leur gouvernance, pour être à même de proposer une offre suffisamment claire, lisible, et documentée, permettant à chacun de s'en saisir. Il est alors important de considérer ces choix dans le processus d'appropriation des données par les usagers.

2.1 Données ouvertes et enjeux : l'importance des choix de formation et de méthodes

La loi Lemaire de 2016, a introduit la notion de "données d'intérêt général" qui implique que toute administration a désormais l'obligation de mettre à disposition les documents, informations et données collectés dans le cadre d'une mission de service public, dans un format électronique, en standard ouvert, permettant une exploitation entièrement libre par tous ceux qui le souhaitent (citoyens, entreprises, associations, administrations, etc.). Ces données dites "ouvertes" (*open data*) sont donc des données publiques, qui ont vocation à être librement accessibles pour produire des informations, de la connaissance et développer des compétences. L'absence d'indication de modalités exactes de mises en forme des données, induit nécessairement des choix en matière de formats, licences, métadonnées, notamment, qui reposent sur la gouvernance et la politique mise en œuvre au sein de chaque administration. Or, ces choix sont essentiels étant à la source du processus d'utilisation qui peut s'illustrer à travers le schéma suivant. On peut considérer trois niveaux d'usage, que l'on peut mettre en regard avec différents objectifs pédagogiques, permettant de passer de la notion de données, à la notion d'information, puis de connaissance et enfin de compétence.

Ces données concernent tous les domaines (la démographie, la santé, l'environnement, l'urbanisme, l'enseignement, la culture...) et sont ainsi susceptibles d'intéresser les enseignants et les élèves dans de nombreuses disciplines, et tout particulièrement en statistique. En effet, ces données principalement quantitatives peuvent être traitées, triées,

croisées, exploitées, et visualisées. Cela nécessite donc des compétences informationnelles, documentaires, statistiques mais aussi techniques (informatiques) visant dans le même temps à soulever des questions socio-économiques, politiques, historiques ou géographiques. Une véritable acculturation à l'environnement propre de ces données est donc nécessaire.

Dans ce processus d'acculturation aux données, une médiation, faisant référence, selon Rézeau (2002) à "l'ensemble des aides ou des supports qu'une personne peut offrir à une autre personne en vue de lui rendre plus accessible un savoir quelconque" apparaît alors indispensable. L'enseignant en est un acteur clé car il peut jouer son rôle de médiateur, en proposant différents scénarios pédagogiques pour former à l'aptitude d'accéder aux données à des fins de compréhension, d'usage et d'analyse, potentiellement pour des prises de décisions sociétales. C'est l'hypothèse que nous interrogeons afin d'identifier des moyens et méthodes favorisant une acculturation et une démocratisation des usages de données.

Pour les élèves, manipuler ce type de données fait partie de la découverte et du développement d'une culture numérique telle qu'elle est présentée dans le socle commun de connaissances et de compétences à travers le domaine 1 "Les langages pour penser et communiquer"^[1]. L'usage des données ouvertes dans la société civile comporte plusieurs enjeux éducatifs qui incitent à s'interroger sur la formation tels que :

- informationnels, car ces données livrées sans médiation nécessitent de maîtriser des techniques de recherche d'informations pour les étayer et les analyser ;
- cognitifs, car elles nécessitent des capacités d'abstraction pour les interpréter, ou pour les manipuler, sans modifier, voire fausser leur sens ;
- médiatiques, car les représentations graphiques et visuelles permettant d'illustrer les données ont un attrait qui peut séduire, mais elles peuvent cependant déformer ou orienter les informations en fonction des intentions de leurs auteurs ;
- techniques, car des compétences en informatique sont nécessaires pour travailler les données et donc permettre de donner à voir et d'en communiquer leur sens ;
- méthodologiques, car les approches quantitatives des big data tendent à imposer leur manière de voir le monde et à standardiser les analyses, au détriment d'une recherche plus compréhensive et plus fine qui pourrait s'appuyer sur les thick data.

2.2. Quel stratégie pédagogique pour développer les apprentissages d'usage des données ouvertes ?

Nous proposons d'analyser les stratégies professionnelles mises en œuvre par les enseignants autour des données ouvertes à travers le filtre des scénarios pédagogiques. Ces scénarios décrivent les dispositifs technologiques, l'utilisation des données, dans une mise en forme qui s'appuie sur des hypothèses concernant l'apprentissage, les compétences, ainsi que les postures et les stratégies professionnelles, dans le cadre imposé par la gouvernance de ces données. Les scénarios pédagogiques peuvent être définis, selon Paquette (2003) comme « ensembles ordonnés d'activités, régies par des acteurs qui utilisent et produisent des ressources » ; ils organisent, d'après Wenger (1998), le répertoire de ressources partagées par une communauté de pratiques. La scénarisation est associée aux gestes professionnels de l'enseignant et à une dimension de créativité dans la conception, autant qu'à un acte d'ingénierie réflexive. Le scénario pédagogique a pour objet de proposer un environnement permettant de mettre en œuvre le processus d'apprentissage en s'appuyant sur des

ressources pédagogiques issues d'open data fabriquées en amont (processus externe) imposant à l'enseignant un certain nombre de contraintes. Nous interrogeons la façon dont s'opère la transition "open data - ressources pédagogique". S'intéresser à la scénarisation pédagogique dans le cadre d'usages de données nécessite dans un premier temps de comprendre les différents niveaux d'usages qui peuvent être considérés comme les apprentissages devant être développés.

3. Cadre méthodologique pour l'analyse des usages émergents dans l'éducation

La recherche s'inscrit dans le cadre d'un projet Peps Idex CNRS coordonné par l'équipe RUDII (Représentations, Usages, Développements, Ingénieries de l'Information) du laboratoire IMS UMR 5218 CNRS. Le projet de recherche MVDE (Médiation et Valorisation des Données ouvertes pour l'Éducation) s'est constitué autour de la recherche en sciences de l'information et de la communication, en sciences de l'éducation, en didactique des statistiques et en didactique des sciences économiques et sociales mais aussi autour du design. Ce projet a pour ambition de diffuser la connaissance des usages des données ouvertes en milieu scolaire et d'amorcer une réflexion collective quant à ces pratiques. L'observation d'usages émergents dans l'éducation, la mise en relation des acteurs et la réflexion collective autour des besoins et ressources facilitant le développement des usages pédagogiques des données ouvertes sont principalement visés. Le travail d'enseignants a été suivi au long de l'année scolaire et un ensemble de documents relatifs à leur travail en classe a été collecté à savoir des fiches projets en réponse un appel de la Délégation pour l'Éducation aux Médias et à l'Information, leurs descriptifs de séquences pédagogiques, leurs supports de cours, des fiches d'aides pour les élèves, un wiki alimenté par des enseignants, des productions d'élèves.

L'analyse d'entretiens auprès d'enseignants et de leurs documents pédagogiques a été réalisée, ceci afin de caractériser les choix pédagogiques permettant d'identifier le processus de conception de l'activité d'apprentissage, ainsi que les stratégies et positions des enseignants visant à sensibiliser les élèves aux problématiques liées aux données, à leur traitement et à leurs usages. Ces éléments nous permettent de caractériser différents types de scénarios pédagogiques et de design de l'activité en classe.

Trois types de scénarios pédagogiques ont été identifiés. Nous les désignerons comme scénario transmissif, scénario participatif et scénario d'autonomisation. Cette typologie témoigne de la construction en cours d'une culture professionnelle autour des données ouvertes et sera développé au cours de la communication.

4. Discussion : les conditions de construction d'une culture autour des données ouvertes

L'utilisation pédagogique des données repose sur des scénarios établis par les enseignants en appui sur leur propre culture des données. Cette culture peut être plus fortement influencée par une culture disciplinaire (lorsqu'il s'agit en priorité de répondre à des objectifs curriculaires) ; numérique (liée aux enjeux du partage des ressources et de l'open source) ; citoyenne (liée à l'accompagnement des élèves dans des initiatives participatives au sein d'une communauté) ; technique (lorsqu'il s'agit de permettre aux élèves de s'autonomiser dans leurs usages des objets et outils numériques).

Les modèles de design et de scénarios de l'activité pédagogique construits par les enseignants mettent en évidence les étapes de travail nécessaires pour sélectionner, transformer, transposer les ressources sociétales (les jeux de données mis à disposition) en ressources pour apprendre. Nous identifions donc plusieurs obstacles dans l'usage des données ouvertes telles qu'elles sont aujourd'hui accessibles sur les plateformes. En effet, très peu d'enseignants s'engagent véritablement dans la mise en activité des élèves à partir de jeux de données directement issus des plateformes d'accès. Les enseignants choisissent plus souvent la construction de données avec les élèves plutôt que la réutilisation de données préexistantes. La dimension "thick data" qui permet d'interpréter et de comprendre reste difficilement accessible pour les élèves mais aussi pour les enseignants. Ainsi, pour développer l'acculturation aux données, il apparaît essentiel de faciliter l'accessibilité grâce à des ressources de médiation, voire de formation. Une possibilité serait de mettre en place un travail avec des documentalistes permettant d'indexer les données, notamment à l'aide de métadonnées, afin d'en faciliter la lecture et l'intelligibilité par des personnes extérieures à leur contexte de production d'origine.

BIBLIOGRAPHIE

[1] Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone?. *First Monday*, 16(2).

[2] Hess, C., Ostrom, E., (dir.) (2007). *Understanding Knowledge as a Commons : From Theory to Practice*. Cambridge: The MIT Press.

[3] Paquette, G., Bourdeau, J., Basque, J., Leonard, M., Henri, F., & Maina, M., (2003). Construction d'une base de connaissances et d'une banque de ressources pour le domaine du téléapprentissage. *Sciences et Techniques Éducatives*, 10.

[4] Rézeau, J. (2002). Médiation, médiatisation et instruments d'enseignement : du triangle au « carré pédagogique ». *ASp*, 35-36, 183-200.

[5] Rouvroy, A. (2014). Des données sans personne : le fétichisme de la donnée à caractère personnel à l'épreuve de l'idéologie des Big Data. In J. Richard & L. Cytermann (dir.) *Etude annuelle du Conseil d'Etat "Le numérique et les droits fondamentaux"*. La documentation française.

[6] Wenger, E. (1998). *Communities of practice : learning, meaning and identity*. New York: Cambridge University Press.

[1] Voir sur le décret n° 2015-372 : http://www.education.gouv.fr/pid25535/bulletin_officiel.html?cid_bo=87834