
N° 4 | 2018

Méthodes et stratégies de gestion de l'information par les organisations : des big data aux thick data

Stratégies pédagogiques pour diffuser la culture des données ouvertes

Camille CAPELLE *Maître de conférences*

IMS UMR5218 CNRS

University of Bordeaux

Marthe-Aline JUTAND *Maître de conférences*

Faculté sciences de l'éducation

CeDS

University of Bordeaux

Franc MORANDI *Professeur des universités*

IMS UMR5218 CNRS

University of Bordeaux

Édition électronique :

URL :

<https://revue-cossi.numerev.com/articles/revue-4/66-strategies-pedagogiques-pour-diffuser-la-culture-des-donnees-ouvertes>

DOI : numerev_1601

Date de publication : 04/09/2018

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : CAPELLE, C., JUTAND, M.-A., MORANDI, F. (2018) Stratégies pédagogiques pour diffuser la culture des données ouvertes. *Revue COSSI*, (4). https://doi.org/10.34745/numerev_1601

Après que les industriels se soient emparés du Big Data, il en revient à la sphère sociale de s'emparer de l'Open Data. Les enseignants ont un rôle à jouer pour permettre aux élèves, en tant que futurs citoyens, d'appréhender les données, de comprendre leurs contenus et le processus ayant permis de les produire, et ainsi de développer les compétences nécessaires pour en faire usage. Il ne s'agit plus seulement de constituer des données pour valider des hypothèses, mais d'interroger la structure et le contenu de données préexistantes pour se les approprier et construire de nouveaux savoirs. Selon leurs modalités de collecte et de mise à disposition, la construction de sens et la pertinence des informations produites sont loin d'être évidentes. 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 :

Open data, Données ouvertes, Scénario pédagogique, Design 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. 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. 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 build new knowledge. Depending on how they are collected and made available, the meaning construction and relevance of the information produced are not obvious. 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, pedagogical design, data training, data culture

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* (Hess et Ostrom, 2007). Ces données numériques reflètent notre quotidien sous forme « éclatée, segmentée, distribuée, décontextualisée, déshistoricisée » (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. Cette mise à disposition permet leur accès pour une manipulation par les citoyens ou les entreprises, à des fins personnelles, de recherche, de diagnostic ou de mise en évidence de sens. 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. C'est ce que font, avec des données de masse, les grandes entreprises du numérique ou les services de renseignement gouvernementaux à l'aide de moyens dont ne peuvent disposer les plus petites entreprises et encore moins les citoyens. Ces « intelligences algorithmiques » influencent nos façons d'être et d'agir, en organisant l'information et ainsi en permettant de prendre des décisions (Cardon, 2015). Le premier enjeu de l'usage des données se situe alors dans la capacité à s'en saisir et à les « faire parler », à travers les inégalités de moyens pour agir avec, face aux puissances économiques et politiques, qui s'appuient sur la puissance algorithmique (approche « *big data* »). Le second enjeu, qui est un enjeu de citoyenneté, concerne la formation à la compréhension et à l'exploitation de données. Il concerne la place de l'intelligence cognitive et de l'approche compréhensive pour utiliser les données, à travers leurs dimensions culturelles et sociales notamment avec des informations plus qualitatives (approche « *thick data* »). 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, car la mise en analyse de ces données bouleverse « les catégories d'interprétation traditionnelles (qui) ne tiennent plus d'elles-mêmes et (...) il faut commencer par regarder avant d'interpréter » (Cardon, 2012 : 142).

Après avoir exposé l'enjeu d'une approche des données ouvertes par les *thick data* pour former les futurs citoyens à une culture et aux usages des données, nous définissons ce que nous désignons par scénarios pédagogiques et par design pédagogique. La méthodologie choisie pour explorer les stratégies pédagogiques de formation aux données ouvertes est 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.

LES USAGES PÉDAGOGIQUES DES DONNÉES OUVERTES

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.

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 (cf. figure 1). 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.



Figure 1 : Adaptation de la pyramide de Russel Akoff (1989) aux usages pédagogiques des données ouvertes

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. Pour bien comprendre comment ce processus peut s'opérer, rappelons tout d'abord la particularité de ce type de données et les enjeux qu'elles soulèvent en matière d'appropriation à des fins d'enseignement.

Ces données, parfois collectées en masse, notamment par la mise en place systématique de capteurs, 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, ainsi que les chercheurs, les entrepreneurs, etc... En effet, ces données, principalement quantitatives, peuvent être traitées, triées, croisées, exploitées, et visualisées à des fins de recherche, de développement, d'innovation ou d'enseignement. Leur traitement nécessite une intelligence humaine permettant leur intégration aux activités des professionnels ou du public (analyse, évaluation, visualisation, cartographie, etc.). 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 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 » (Rézeau, 2002) 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], dans le programme d'Éducation aux Médias et à l'Information (EMI) en cycle 3 ou dans certains programmes d'enseignement^[2].

L'usage des données ouvertes dans la société civile comporte aussi plusieurs enjeux éducatifs qui incitent à s'interroger sur la formation. Ces enjeux sont de plusieurs ordres :

- 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.

L'approche par les « *thick data* » (Hine, 2015), développée dans le champ de l'ethnographie et plus particulièrement des *internet studies*, permet d'appréhender les données d'une autre façon que par l'approche quantitative qui consiste à appliquer des algorithmes pour traiter les données massives. Cette autre approche repose sur une approche qualitative qui implique de s'intéresser aux données dans leur épaisseur (environnement et processus de construction) pouvant se traduire à travers différentes couches d'informations relatives à une approche socio-culturelle des données, mais aussi documentaires et techniques. Cette épaisseur peut permettre de comprendre la chaîne de production des données ouvertes, par la connaissance de leur origine, des choix de leur mise en forme, leur description. Les usages des données ouvertes dans l'éducation impliquent de s'intéresser à ce type d'approche pour permettre de traduire en informations les données et construire des connaissances. Il convient en effet d'en comprendre les fondements dans toutes leurs dimensions là où l'approche par le *big data* ferait l'impasse de cette démarche compréhensive pouvant ainsi conduire aux risques évoqués précédemment. Antoinette Rouvroy dénonce l'instauration d'un « régime de vérité » numérique, une nouvelle manière de rendre le monde signifiant : « la réalité (...) y serait saisie (...), non plus au niveau de ses représentations et transcriptions ou de ses interprétations individuelles ou collectives, mais au niveau quasiment atomique ou génétique de la donnée, considérée comme un fait ultime, parlant d'elle-même, sans médiation » (Rouvroy 2014 : 413). La connaissance et la prise en compte des *thick data*, semble être une piste pour construire une démarche pédagogique. Cette démarche relève donc d'un enjeu démocratique en ce qu'elle peut permettre d'amener les citoyens à comprendre et à agir intelligemment, à choisir, à imaginer et à innover, en prévenant les risques d'une approche purement algorithmique.

Quel design 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 comme « ensembles ordonnés d'activités, régies par des acteurs qui utilisent et produisent des ressources » (Paquette, 2005) ; ils organisent le répertoire de ressources partagées par une communauté de pratiques (Wenger, 1998). 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. A la différence des schémas trop rigides ou de la description des *lesson plans* chez les anglo-saxons, elle organise le faire des acteurs autour d'un évènement, d'une situation qui mobilise les acteurs et induit un déroulement d'actions. Autour du projet avec des données, les scénarios peuvent associer différentes dimensions, depuis la logique pédagogique et jusqu'à la gouvernance des données, elle-même mise à l'épreuve de l'exploitation et de

la réutilisation des données ouvertes par des usagers.

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 (cf. figure 2).



Figure 2 : Processus de scénarisation pour l'usage pédagogique des données ouvertes

Nous interrogeons la façon dont s'opère la transition « *open data* - ressources pédagogiques ». Cela implique une appropriation par l'enseignant puis par les élèves des ressources proposées pour pouvoir les transformer en informations et développer des compétences.

La métaphore du scénario, de la mise en scène et de son script est utilisée pour marquer les caractères d'une planification anticipatrice et d'une logique suivie, qui organise le déroulement de l'activité à travers les différentes séquences pédagogiques qui s'y rapportent. Plus que l'objet à enseigner et à apprendre, le scénario représente la trame explicitée et communicable de l'activité. Le scénario est le résultat du processus de conception pédagogique qui prend sa source dans les différents modèles pédagogiques. Du point de vue didactique, Emmanuelle Villot-Leclerc (2007 : 18) souligne que « l'émergence du concept de scénario est révélatrice du déplacement du paradigme épistémologique de la connaissance d'une logique de la transmission à une logique d'apprentissage (Jonnaert, 2002) ». Un scénario pédagogique est constitué par deux aspects, l'un descriptif, inscrit dans les pratiques proposées par l'enseignant décrivant « ce qu'il prévoit de faire ou fait » (Villot-Leclerc ; 2007). Cet aspect se caractérise par des objets de descriptions écrits, ce qui implique que les scénarios constituent un répertoire et une typologie de pratiques partageables. Le second aspect, prospectif, est celui des intentions pédagogiques, des stratégies, un angle de prise de position sur le mode de construction de savoirs, qui modélise une intention. C'est le niveau de conception d'une activité d'enseignement et d'apprentissage donnant sa forme à l'activité pédagogique, son design.

Gilbert Paquette définit le concept de « *design* pédagogique » comme l'« ensemble des théories et des modèles permettant de comprendre, d'améliorer, et d'appliquer des méthodes d'enseignement favorisant l'apprentissage » (Paquette, 2002, p.111). Il ajoute que « par le *design* des scénarios pédagogiques, le concepteur établit les liens entre les sources d'information et les différents acteurs. [...] Le concepteur prévoit les types de communication, les stratégies pédagogiques, les modes de collaboration entre les acteurs » (Paquette et al., 2003). Un exemple est celui de la pédagogie inversée (*flipped learning*, ou apprentissage inversé), souvent plébiscitée. Ce scénario décline un *design*, celui de l'inversion, supportant une série d'hypothèses pédagogiques. La « pédagogie inversée » propose d'adapter l'ordre pédagogique aux activités d'apprentissage sous forme d'alternance et d'inversion : le contenu des cours (partie

théorique) est accessible aux élèves en amont, à l'opposé de l'enseignement magistral, à l'aide de plateformes numériques, sous forme de ressources (proposant par exemple des capsules vidéo élaborées par l'enseignant). Les élèves s'en saisissent avant la classe, et peuvent ainsi prendre des notes, en les complétant par leurs propres recherches. Puis, le présentiel de la classe est consacré au travail d'équipe, à la mise en projet, aux activités permettant de mettre en œuvre les notions présentées. L'enseignant fournit des aides, des explications complémentaires aux élèves, selon leurs besoins spécifiques. Dans sa posture, il n'est plus un transmetteur (même s'il reste maître d'œuvre du dispositif), mais accompagne le travail de l'élève.

Dans le cadre d'apprentissage par usage des données, trois approches peuvent caractériser les intentions du *design* pédagogique. Une première serait centrée sur les données ressources, essentiellement pour les enseignants et leurs apports, dans la verticalité du processus « enseigner ». Un second type d'approche se centre sur les activités, dans l'horizontalité du « faire apprendre », associée aux pédagogies actives. Une dernière direction organisatrice met l'accent sur des approches centrées sur les interactions, pédagogique à travers des projets organisationnels, des parcours de compétences (transversalité). Elle propose une dimension collaborative, associée à un projet de formation à l'information et aux données, assortie de démarches d'accompagnement.

Les scénarios peuvent alors être examinés à différents niveaux :

- un niveau *macro*, le scénario général de structuration pédagogique, celui des intentions ;
- un niveau intermédiaire ou *meso*, le scénario d'activités qui organise une séquence d'activités,
- un niveau *micro*, le scénario de déroulement d'activités.

L'entrée par la scénarisation et le design pédagogique de l'enseignant nous permet d'interroger la façon dont sont conceptualisés les usages des données en situations d'enseignement et d'accéder aux intentions pédagogiques des enseignants. Nous proposons donc d'analyser les scénarios et les stratégies professionnelles mises en œuvre pour former les élèves à ces nouveaux objets de savoirs que sont les données ouvertes. 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.

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. Nous présentons dans un premier temps les principaux

objectifs de ce projet avant de décrire la méthode d'observation et de collecte du corpus ayant permis l'analyse exposée.

Le projet de recherche sur l'usage des données ouvertes pour l'éducation

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. Au début de cette recherche, la question de l'apprentissage de l'usage des données ouvertes n'apparaît qu'en toile de fond de certains programmes disciplinaires (sciences de gestion ou sciences économiques et sociales) dans l'enseignement secondaire. La notion de données est abordée au sens large (données numériques, données personnelles, données de référence,...), notamment en mettant à l'épreuve l'apprentissage permettant de développer des compétences numériques chez les élèves. En revanche, celle de « données ouvertes » ne figure dans aucune directive ministérielle ni curricula. L'intégration des données ouvertes dans l'enseignement n'est donc pas une injonction institutionnelle, et très peu, voire aucune ressource pédagogique n'est proposée aux enseignants.

Un collectif de chercheurs, d'enseignants du secondaire et de l'enseignement supérieur, mais aussi de professionnels et d'étudiants s'est réuni dans l'objectif d'amorcer une réflexion collective sur l'intégration de l'usage de ces données dans la formation, et de réfléchir à des propositions concrètes permettant de faciliter leur manipulation et leur appropriation dans l'enseignement. 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.

Les acteurs qui s'intéressent à cette question ont été identifiés en Aquitaine, autour d'enseignants de l'ESPE, d'Inspecteurs pédagogiques Régionaux notamment en sciences économiques et sociales, du CLEMI (Centre de Liaison pour l'Education aux Médias et à l'Information) au rectorat, d'enseignants de sciences de gestion, d'associations comme la FING (Fédération Internet Nouvelle Génération), et de professionnels^[3]. Ce travail a été facilité par l'intégration dans un groupe de Travaux Académiques Mutualisés (TraAM)^[4] sur le thème « Datavisualisation et cybercitoyenneté ». Ce groupe de travail a permis d'identifier des enseignants en sciences de gestion, en sciences économiques et sociales, en technologie ou en informatique, s'intéressant à la question des données ouvertes, certains s'en étant déjà emparés pour leur enseignement. Ils participent à la réflexion autour des données ouvertes et construisent des séquences pédagogiques dans des disciplines et selon des approches variées avec des élèves de niveaux collège, lycée et BTS.

Observations, entretiens et analyse du design pédagogique

La méthodologie adoptée pour ce projet repose sur l'approche ethnographique issue des travaux de l'École de Chicago qui consiste à s'immerger au cœur des pratiques afin d'acquérir une meilleure compréhension de l'activité du groupe observé. Elle a dans un premier temps consisté à participer aux réunions de groupe de travail TraAM et à interviewer des acteurs de différents profils : inspecteurs et cadres de l'éducation, enseignants, élèves, responsables de la mise en accès de données publiques des collectivités. Nous avons mené une vingtaine d'entretiens semi-directifs. Les axes sur lesquels nous les avons interrogés concernent leur avis sur :

- La représentation de l'offre de données ouvertes : diversité, dimensions sociales et politiques, en lien avec une discipline ou non ;
- Les portails de données, les types de données utilisés, et les outils et tutoriels éventuels utilisés pour la visualisation de données, et les formes de représentations graphiques de l'information choisies ;
- Les besoins de formations aux données, aux outils, aux développements de compétences techniques ou disciplinaires ;
- Les usages pédagogiques : fréquence, contenu, adossement à des dispositifs et des programmes, ainsi que les modalités d'organisation (projet, accompagnement, travail collaboratif, pluridisciplinaire ou non) ;
- Les notions à aborder : big data, communs de la connaissance, algorithmes, compétences professionnelles de traitement des données, critique de l'information, etc...
- Les dimensions économique et politique des données (*communs de la connaissance*) ;
- L'émergence d'une communauté épistémique sur ces questions : réseaux interne à l'Education Nationale, liens avec l'extérieur, participation en ligne et présence sur les réseaux sociaux.

Puis, le travail des enseignants a été suivi au long de l'année scolaire et nous avons collecté un ensemble de documents relatifs à leur travail en classe, à savoir des fiches projets en réponse un appel de la Délégation pour l'Education 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.

Nous avons comparé leurs perceptions de l'usage pédagogique des données ouvertes à travers les discours recueillis, et avons analysé leurs documents pédagogiques. Nous nous sommes intéressés aux 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.

CARACTÉRISATION DES SCÉNARIOS PÉDAGOGIQUES D'USAGE DES DONNÉES OUVERTES EN CLASSE

Trois types de scénarios pédagogiques sont identifiés. Nous les désignons 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.

Le scénario transmissif : préparation des données par l'enseignant

Dans le scénario dit « transmissif » (Figure 3), l'enseignant prépare un jeu de données adapté à ses enseignements disciplinaires, à partir duquel les élèves répondent à une question de cours. Ce scénario repose sur plusieurs contraintes identifiées par les enseignants pour réutiliser des jeux de données ouvertes dans un but pédagogique. D'une part, les jeux de données disponibles répondent, la plupart du temps, difficilement aux objectifs pédagogiques que l'enseignant souhaite aborder avec sa classe, objectifs le plus souvent liés à une approche disciplinaire. L'offre de jeux de données, encore très limitée et lacunaire, conduit les enseignants à sélectionner des fichiers et à les retravailler pour les rendre pertinents et exploitables. Pour certains, l'usage de données ouvertes ne doit pas s'imposer comme une contrainte dans le scénario pédagogique. Il s'agit d'abord pour eux de former les élèves à des connaissances disciplinaires dans lesquelles les enseignements techniques peinent à trouver leur place, alors qu'ils sont souvent associés aux enseignements disciplinaires à l'ère du numérique.

D'autre part, l'usage de ces données « *nécessite des connaissances techniques, ne serait-ce que pour ouvrir les données dans le bon logiciel* » (enseignant de technologie). Les libellés des variables, rédigés par des experts du domaine et parfois abrégés, rendent certains jeux de données difficiles à comprendre pour des personnes extérieures. Les enseignants effectuent alors souvent un premier travail de reformulation et simplification du fichier de données visant à le rendre plus lisible. Ce travail va parfois plus loin. Il peut nécessiter de convertir le format de fichier ou même de le modifier ou le compléter pour éviter son incomplétude. Ce travail est souvent présenté comme nécessaire de la part de l'enseignant pour permettre de rendre les données accessibles aux élèves dans des délais souvent restreints. Les savoirs travaillés avec les élèves ne reposent donc pas ici sur le traitement des données, jugé trop technique ou chronophage, mais plutôt sur l'analyse des données en relation avec une question de cours et sur leur représentation graphique informationnelle. Des supports pédagogiques, fiches ou ressources d'aides, peuvent être préparés par l'enseignant qui anticipe ainsi sur les difficultés des élèves.



Figure 3 : Le scénario transmissif

L'approche pédagogique choisie s'inscrit ici plutôt dans la verticalité visant des apprentissages déductifs puisque l'enseignant définit les savoirs à enseigner en appui de ressources précises. Ces ressources ont fait l'objet d'une transformation par l'enseignant pour en faire des ressources directement exploitables en classe par les élèves, à des fins d'apprentissages disciplinaires plutôt que d'usages des données ouvertes telles qu'elles sont mises à disposition sur les plateformes d'open data. Néanmoins, dans cette configuration le scénario proposé par l'enseignant peut contenir une intention de mise en activité des élèves suivant une approche horizontale, visant à faire apprendre aux élèves par eux-mêmes en travaillant à l'aide des ressources à disposition. Pour l'enseignant ici, les jeux de données produits et mis à disposition par les organismes publics, ne permettent pas directement une utilisation pédagogique, pour des raisons de contraintes de temps, mais aussi compte tenu des objectifs pédagogiques qui se rapportent avant tout à un programme disciplinaire. Dans ce cas, la problématisation vient davantage de l'enseignant metteur en scène de l'activité, que des élèves.

Le scénario participatif : production de données et contribution de la part des élèves

Un scénario participatif (Figure 4) peut viser différents objectifs. Nous en avons observé deux. Le premier inscrit la mise en activité des élèves dans la culture du libre sur internet avec pour intention pédagogique de faire découvrir et participer les élèves à une démarche collaborative open source. Le second inscrit la mise en activité des élèves sur un plan de participation à la politique locale à travers la collecte de données dans leur environnement proche pour produire des informations directement utiles à la communauté.

Pour le premier objectif nous décrivons l'exemple d'une enseignante du primaire ayant développé un projet de type cartographique, de contribution de données sur une carte collaborative en ligne (ex : Umap d'Openstreetmap, Framacarte, etc.). Elèves et enseignants peuvent ainsi travailler sur un territoire, autour de l'établissement scolaire ou proche de chez eux, pour collecter les données nécessaires et pour pouvoir les mettre à disposition dans le format attendu sur le dispositif. Pour cette enseignante, il s'agit de permettre aux élèves de « *construire leur propre représentation de ce que peut être Internet, la culture contributive du réseau, la connaissance comme bien commun élaboré collectivement* ». L'intention pédagogique est ici clairement liée à des choix et à un engagement de l'enseignante pour la culture du partage, la culture du libre, qui privilégie l'ouverture des données en tant que communs de la connaissance. La création et la mise en forme des données par les élèves permet ici d'éviter les difficultés liées aux formats et aux contenus des jeux de données existants sur les plateformes. Cette approche permet ainsi de comprendre comment se collectent et se produisent les données et facilite le processus d'interprétation puisque les données sont collectées dans leur contexte pour répondre à un besoin d'information. La recontextualisation des données n'est donc pas ici une problématique dans l'activité pédagogique. Celle-ci est par ailleurs modélisée par le dispositif en ligne qui implique de collecter des données de types et de formats précis, par exemple sur la géolocalisation, le type d'objets relevés, etc.

Pour le second objectif, nous prenons l'exemple d'un enseignant de sciences économiques et sociales qui définit un cadre pédagogique en laissant le choix aux élèves de déterminer la nature des données qui les intéressent pour participer à la vie de leur établissement et être force de proposition (proposition d'aménagements des temps scolaires pour réduire les temps d'attente à la cantine ou pour fluidifier les entrées et sorties dans l'établissement). Pour lui, le choix repose sur des apprentissages liés à la compréhension du processus par la production des données. L'enseignant guide les élèves qui identifient les données dont ils ont besoin pour répondre à une ou plusieurs questions et les recherchent sur des plateformes ou les collectent eux-mêmes. Il s'agit d'aller recueillir des données au sein de leur établissement scolaire et de les organiser pour les analyser. Dans ce contexte, les enseignants choisissent de travailler les compétences d'investigation et de collecte des données avec les élèves. Comme l'indique cet enseignant, «*produire de l'information chiffrée locale qui n'existe pas pour montrer l'impact que cela peut avoir sur les décisions locales*» est un objectif pédagogique. Produire des données qui n'existent pas suppose de faire un travail de relevés et de mesures, qui peut parfois être long et fastidieux tout particulièrement dans le temps pédagogique imparti. La séquence pédagogique est donc plutôt envisagée sur le mode de projet. Elle aboutira à la création d'une information visuelle et à sa communication auprès de l'établissement qui viendra appuyer les propositions concrètes des élèves. La pédagogie active visant l'apprentissage par projet amène les élèves à fixer leur objectif : celui d'apporter des réponses et propositions concrètes par rapport à ce qui a été observé.



Figure 4 : Le scénario participatif

On est ici pleinement dans une dimension horizontale du « faire apprendre » par l'activité, ainsi qu'une dimension organisatrice centrée sur les interactions soutenues par l'enseignant qui attend un résultat précis et accompagne les élèves pour atteindre cet objectif. Les élèves doivent rechercher des données existantes ou les produire pour répondre aux questions qu'ils posent. Une fois les données recueillies, ils devront retravailler les données initiales pour créer une information et mettre en scène des connaissances afin de répondre précisément à la question posée par l'enseignant.

Le scénario d'autonomisation : usage critique et construction collective du sens sur le mode de projet

Dans le scénario d'autonomisation (Figure 5), les élèves choisissent un jeu de données ouvertes disponible sur une plateforme d'open data et apprennent avec l'enseignant à « faire parler » les données. L'objectif est alors d'apprendre aux élèves à manipuler les données, pour en comprendre le sens et à construire une problématique autour. Comme l'explique un enseignant de technologie, il s'agit de « *traiter des données de la vie réelle, d'être producteur d'informations et de pouvoir communiquer avec ces informations* ». Ce travail mené par des enseignants notamment en technologie ou en mathématiques en collège dans une perspective d'Éducation aux Médias et à l'Information, consiste à comprendre l'usage qui peut être fait des données réelles

découvertes dans les jeux de données existants sur les plateformes d'open data. La démarche proposée aux élèves est de problématiser et de construire des hypothèses pour les infirmer ou les confirmer à partir d'un jeu de données sélectionné. Il peut également arriver que les élèves formulent des questions qui appellent d'autres données que celles qui sont disponibles. Ainsi, lorsque le jeu de données qui les intéresse n'est pas accessible, les enseignants accompagnent les élèves pour en faire la demande auprès des organismes concernés. Cette démarche associe connaissances et compétences aux questions de citoyenneté pour identifier le bon interlocuteur et formuler sa demande. Quand le jeu de données est obtenu, l'utilisation d'un tableur puis d'outils de visualisation de données permettent aux élèves de leur donner une forme et de communiquer le message souhaité. Les élèves produisent ainsi des infographies ou peuvent aller jusqu'au développement web d'applications au niveau du BTS.



Figure 5 : Le scénario d'autonomisation

Ce scénario s'inscrit ici clairement dans un modèle horizontal et transversal à travers un projet impliquant le développement de compétences, la collaboration et visant l'autonomisation des élèves par la mise en œuvre de leur propre projet. La dimension collaborative, associée au projet et assortie de démarches d'accompagnement est aussi présente. Les enseignants interviennent surtout pour soulever des questions qui se posent pour pouvoir interpréter les données et s'intéressent pour l'évaluation aussi bien au processus mis en œuvre qu'à la production des élèves.

Dans tous les scénarios pédagogiques, la production finale d'informations se fait sous forme de visualisation. L'analyse fait apparaître également que les méthodes de production des données sont parfois abordées, mais que, dans l'ensemble, l'accent est mis davantage sur l'exploitation des données (ce qu'on en fait), plutôt que sur la recherche d'information et de précision assurant la fiabilité de l'interprétation (ce qu'elles permettent de savoir réellement), par exemple à l'aide de métadonnées ou de recherches sur le contexte de production des données. Développer la capacité des élèves à avoir une lecture critique de la production des données est pourtant un objectif pédagogique transversal visé dans l'enseignement, et l'Éducation aux Médias et à l'Information en est le fil conducteur. L'engouement et l'attrait de la datavisualisation des données chez les élèves, peuvent parfois prendre le dessus sur cet objectif. Cependant, on voit bien qu'« il s'agit moins de s'attacher à la mise en données (datafication) du monde par la raison informatique qu'à la mise en expériences du monde par le design numérique. » (Vial, 2015). Ainsi, l'approche « *thick data* » est clairement plébiscitée pour apprendre à utiliser les données ouvertes.

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. La documentarisation, entendue comme « le travail consistant à équiper un support pérenne des attributs qui faciliteront sa circulation dans l'espace, le temps et les communautés d'interprétation » (Zacklad, 2007 : 23), ainsi qu'une redocumentarisation (Salaün, 2007), nous paraissent être des étapes clés dans le processus de production des jeux de données ouvertes. La mise en forme des données ouvertes (leur *design*) gagnerait donc à être pensée avec les acteurs de l'éducation et les documentalistes, acteurs incontournables des médiations liées aux nouveaux objets numériques.

BIBLIOGRAPHIE

Ackoff, R.L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis*, 16, 3-9.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5). 662-679.

Cardon, D. (2012). Regarder les données. *Multitudes*, 2(49), 138-142.

Repéré à :
[http://www.cairn.info.docelec.u-bordeaux.fr/revue-multitudes-2012-2-page-138.htm](http://www.cairn.info/docelec.u-bordeaux.fr/revue-multitudes-2012-2-page-138.htm)

Cardon, D. (2012). *A quoi rêvent les algorithmes : Nos vies à l'heure des big data*. Paris : Seuil.

- Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone?. *First Monday*, 16(2).
- Hess, C., Ostrom, E., (dir.) (2007). *Understanding Knowledge as a Commons : From Theory to Practice*. Cambridge: The MIT Press.
- Hine, C. (2015). *Ethnography for the Internet. Embedded, Embodied and Everyday*, London: Bloomsbury.
- Morandi, F. (2006). *Introduction à la pédagogie*. Paris : Armand Colin.
- Morandi, F. (2006a). *Modèles et méthodes en pédagogie*. Paris : Nathan.
- Paquette, G. (2002). Apprentissage sur Internet : des plateformes aux portails d'objets à base de connaissance. In S. Pierre (Eds.), *Innovations et tendances en technologies de formation et d'apprentissage*. Presses de l'école polytechnique de Montréal, 1-30.
- 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.
- Rézeau, J. (2002). Médiation, médiatisation et instruments d'enseignement : du triangle au « carré pédagogique ». *ASp*, 35-36, 183-200.
- 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.
- Salaün, J.M. (2007). La redocumentarisation, un défi pour les sciences de l'information. *Études de communication*, 30, Repéré à : <http://edc.revues.org/428>
- Vial, S. (2014). De la spécificité du projet en design : une démonstration. *Communication et organisation*, n° 46, 17-32.
- Wenger, E. (1998). *Communities of practice: learning, meaning and identity*. New York: Cambridge University Press.
- Zacklad, M. (2007). Une théorisation communicationnelle et documentaire des TIC, In Brossaud, C. et Rebert, B. (dirs) *Humanités numérique 2, Socio-informatique et démocratie cognitive*. Paris : Hermès Science Publications, 20-35.
-

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

^[2] En 2016, voir les options “Informatique et Création Numérique” en Seconde, ou “Information et Sciences du Numérique” en terminale scientifique.

^[3] Armelle Gilliard, experte travaillant sur la question des données ouvertes, voir <http://www.lareinemerlin.org/>

^[4] Pour plus d’informations, voir : <http://eduscol.education.fr/cid98083/les-travaux-academiques-mutualises-traams-des-laboratoires-des-pratiques-numeriques.html>