

Les *datasprints*, un dispositif d'éducation aux données ? Une étude exploratoire via le cas de « Traces de Soldats »

Béatrice Drot-Delange¹ et Françoise Tort²

¹ ACTé, Université Clermont Auvergne

² ENS Paris Saclay

beatrice.drot-delange@uca.fr francoise.tort@ens-paris-saclay.fr

Résumé. L'avènement de la société de la donnée représente de véritables enjeux pour l'éducation, tels que l'identification des compétences nécessaires. Nous analysons, à l'aide de la modélisation des concepts et pratiques de l'éducation aux données de Grillenberger et Romeike, le cas du *datasprint* « Traces de soldats » créé par l'Atelier Canopé 94. L'analyse porte sur un corpus hétérogène de documents rassemblant un document d'intention, un document d'accompagnement et des productions de collégiens. Les résultats obtenus mettent en évidence que la quasi-totalité des compétences proposées par le modèle sont mobilisables dans le *datasprint*. A l'inverse, des compétences susceptibles d'être mises en œuvre dans le *datasprint* ne sont pas prises en charge par le modèle. Des pistes de recherche sont proposées pour contribuer à une didactique de la donnée.

Mots clés : donnée, didactique, enseignement secondaire, situation authentique, compétences, littératie

1 Introduction

Les données, et leurs traitements, s'invitent régulièrement dans le débat public. L'exigence de publication des algorithmes, tels ceux utilisés pour l'affectation des étudiants après le baccalauréat, la mise en œuvre du règlement général sur la protection des données, les amendes infligées aux grands groupes internationaux par la Commission Nationale de l'Informatique et des Libertés en sont quelques exemples récents. Le développement de l'Internet des objets accroît de manière vertigineuse la production, le stockage et l'exploitation de données. Chaque individu ou collectif produit, plus ou moins consciemment, des données qui seront ensuite exploitées par des industries ou des services. Si la valeur économique des données n'est plus à démontrer, leur valeur patrimoniale est probablement moins connue. Dans ce contexte, on peut qualifier notre société contemporaine de société de la donnée.

Ces évolutions sociétales et techniques représentent de véritables enjeux pour l'éducation et la formation. L'une des questions qui se posent est celle des compétences nécessaires dans cette société de la donnée. Un cadre général pour cette réflexion est proposé par ce que des chercheurs nomment la littératie des données et l'éducation aux données.

Crusoe [2] en propose une définition qu'il élabore en synthétisant les différentes facettes repérées dans la littérature et qu'il juge essentielles : « La littératie des données est la connaissance de ce que sont les données, de la manière dont elles sont collectées, analysées, visualisées et partagées, et est la compréhension de la manière dont les données sont utilisées à des fins bénéfiques ou défavorables, dans le contexte culturel de la sécurité et de la vie privée » (p. 38). Cette définition intègre notamment la connaissance de l'étendue des phénomènes ou objets que l'on peut décrire par des données, des différentes méthodes de collecte, des analyses statistiques et des représentations visuelles, au-delà des graphiques canoniques, des différents modèles de partage des données, en prenant en compte les contextes réglementaires s'appliquant à la protection des données, leur sécurité et leur confidentialité.

L'éducation aux données n'existe pas en tant que telle dans les enseignements scolaires en France. Toutefois, l'enseignement d'informatique introduit depuis 2015 et reconfiguré en 2017 constitue une introduction à quelques dimensions d'une éducation aux données, dans une approche disciplinaire. Nous avons montré [3] que le programme de l'enseignement des Sciences Numériques et Technologie (SNT) et ses manuels scolaires abordent certaines phases du cycle de vie des données, telles que l'acquisition, l'implémentation, l'analyse et le partage des données. Ils n'abordent pas ce qui pourrait relever de l'éducation aux médias et à l'information. Le programme ne propose pas de développer un esprit critique vis-à-vis des données, au sens de Womack [7], à savoir les compétences d'évaluation critique de l'information et la capacité à l'utiliser dans un objectif précis. C'est un enseignement disciplinaire qui vise avant tout les concepts et éventuellement certaines pratiques associées aux données.

D'autres approches de l'éducation aux données s'inscrivent davantage dans cette dimension analytique et critique. C'est le cas des *datasprints* pédagogiques. Ils sont définis comme « un dispositif de médiation numérique des savoirs contributif limité dans le temps où les participants unissent leurs compétences pour explorer, augmenter un jeu de données et proposer des visualisations éclairant une question déterminée » [1]. Ces dispositifs sont réputés ne pas nécessiter d'expertise technique ou scientifique. Leurs promoteurs estiment qu'ils favoriseraient la créativité, la collaboration et le questionnement. Les participants développeraient des habiletés numériques et critiques tout en construisant collectivement un savoir. Il s'agit pour leurs promoteurs de faire travailler les élèves dans le champ des humanités numériques [1]. Il ne s'agit pas d'un enseignement disciplinaire, mais d'une approche qui considère les données comme un matériau.

Dans la perspective d'une contribution à une didactique de la donnée, cet article présente une analyse des concepts et des pratiques relevant d'une éducation aux données mobilisés par les activités du *datasprint* « Traces de soldat ». Nous mobilisons la méthode que nous avons élaborée pour analyser les programmes de SNT [3] et qui s'appuie sur le modèle proposé par Grillenberger et Romeike [5] pour concevoir une éducation aux données. Il s'agit d'identifier les compétences nécessaires aux élèves mais aussi aux enseignants pour se lancer dans ce type de projets. Cette analyse vise également à accroître la robustesse du modèle, pour un niveau d'enseignement secondaire, voire primaire, dans le contexte scolaire français.

2 Cadre d'analyse : un modèle d'éducation aux données

L'identification des compétences nécessaires dans le champ de la littératie des données est un travail mené par de nombreux chercheurs. Parmi eux, les travaux de Grillenberger et Romeike portent sur l'identification des idées fondamentales de ce champ. Ils élaborent un modèle d'éducation aux données [5]. Selon eux, ce modèle serait plus adapté à l'enseignement primaire et secondaire que les travaux déjà existants dans le champ, s'adressant plutôt à l'enseignement supérieur.

Ils proposent un modèle qui distingue d'une part les concepts scientifiques sous-jacents à l'éducation aux données – intitulés « *content area* » dans le modèle – et d'autre part les pratiques avec et sur les données – intitulés « *process* » – basées sur le cycle de vie des données.

Les concepts relèvent principalement de l'informatique. Ils sont regroupés en quatre domaines. Le premier (C1) concerne les concepts de base tels que la distinction entre donnée et information ou la représentation numérique de l'information, par exemple. Le deuxième (C2), recouvre les concepts relevant du stockage et de l'accès aux données, et aborde la duplication ou la synchronisation des données. Le troisième (C3) se focalise sur les méthodes, algorithmes et principes nécessaires à l'analyse des données. Enfin le quatrième (C4) inclut les questions d'éthique, de sécurité et de confidentialité des données.

Quant aux pratiques, elles sont définies en référence au cycle de vie des données. Elles couvrent, selon les auteurs, l'ensemble des pratiques généralement mentionnées dans les définitions de l'éducation aux données : acquisition, nettoyage, modélisation, implémentation, optimisation, analyse, visualisation, évaluation, partage, suppression ou archivage. Le cycle de vie sous cette forme n'étant pas facilement utilisable en contexte scolaire, les auteurs l'ont adapté en identifiant, avec des enseignants et des chercheurs, les phases utiles lors de la mise en œuvre d'une leçon concernant l'éducation aux données. Dans leur modèle, les pratiques sont regroupés en quatre phases : « Collecter, modéliser et nettoyer » (P1), « Implémenter et optimiser » (P2), « Analyser, visualiser et interpréter » (P3) et « Partager, archiver et effacer » (P4).

Le modèle proposé combine ces deux dimensions (concepts et pratiques) sous forme d'une matrice permettant d'identifier précisément les compétences à l'intersection de chacune de ces deux dimensions. Ainsi, le domaine P1 recouvre les capacités de répondre aux questions suivantes : quelles caractéristiques du système à modéliser doit-on collecter en tant que données (P1C1) ? Comment peut-on les collecter (P1C2) ? Comment stocker ces données de manière à pouvoir les utiliser plus tard ? Les données collectées sont-elles utiles à l'atteinte des objectifs visés (P1C3) ? Comment peut-on pratiquement assurer la sécurité des données collectées (P1C4) ? Ces questions sont détaillées pour chaque pratique dans Drot-Delange et Tort [3]. Nous utilisons cette matrice pour identifier les concepts et les compétences mobilisées dans les activités proposées par un *datasprint*.

3 Méthodologie et corpus : le *datasprint* « Traces de soldat »

Le *datasprint* « Traces de soldats », a été créé par l'Atelier Canopé 94. Il proposait aux participants, élèves et enseignants de travailler à partir des traces inscrites sur les monuments commémoratifs et des données numériques pour comprendre les parcours de « Poilus » de la Grande Guerre. Il s'est tenu pendant l'année 2018.

Des ressources à destination des enseignants ont été produites, des productions d'élèves publiées. L'activité a également généré de nombreux échanges sur les réseaux sociaux. Dans cette étude exploratoire, nous avons constitué et analysé quatre corpus constitués de documents hétérogènes. Les corpus 1 et 2 comportent des documents d'accompagnement des enseignants qui nous renseignent sur les intentions pédagogiques du projet, les corpus 3 et 4 des productions d'élèves

Le corpus 1 est un article de Franck Bodin, directeur de l'atelier Canopé 94, publié sur medium.com, qui présente la démarche et l'intérêt pédagogique du travail avec les données¹. Il comporte plusieurs vidéos et des liens vers d'autres articles. Le corpus 2 est un kit pédagogique produit par Philippe Chadeaux, enseignant de Sciences numériques et technologie (SNT), et publié sur la plateforme Canoprof². Il comporte plusieurs pages et des tutoriels vidéo réalisés par l'auteur ou disponibles en ligne.

Ce projet a été réalisé dans de nombreux établissements, écoles, collèges, lycées, de toutes les régions de France. Il a abouti à la publication, par le réseau Canopé d'un jeu de données sur les poilus et de 62 « réutilisations » de ce jeu, correspondant à des productions d'élèves³. Pour chaque réutilisation, un court texte descriptif présente le travail sur les données réalisé par les élèves. De plus, les enseignants ont témoigné de l'avancée des travaux de leurs élèves sur twitter ([#tracesDeSoldats](https://twitter.com/#tracesDeSoldats)), postant parfois des photographies de productions intermédiaires sur papiers ou sur ordinateur.

Pour mener un premier travail exploratoire d'analyse des productions des élèves, nous avons sélectionné les productions de deux établissements scolaires, pour lesquelles nous avons accès à du matériel varié : copies d'écran, vidéos interactives, interview des élèves, affiches, etc. Ainsi, nous avons construit un troisième corpus 3 avec la production publiée du collège Simone de Beauvoir de Créteil (v. figure 1) publiée et 4 photographies d'écrans présentant un travail intermédiaire sur tableur (v. figure 2), tweetées par l'enseignant⁴. Un quatrième corpus 4 comporte les onze productions publiées du collège Achille Mauzan et du lycée Aristide Briand de Gap, ainsi qu'une vidéo

¹ Apprendre avec les données numériques, les *datasprints* pédagogiques : le cas de Traces de Soldats, Bodin F., 29 mai 2018, medium.com, consulté le 10/01/22.

² Traces de soldats - le *datasprint* historique, Chadeaux P, consulté le 10/01/22 : <https://philippe-chadeaux.canoprof.fr/eleve/TdS/>

³ Jeu de données "Première Guerre mondiale - Les Poilus morts pour la France (à compléter)", Réseau Canopé, data.gouv.fr

⁴ Tweet du 12/10/2018 : <https://twitter.com/GirardinYG/status/1050676563602071552> et du 16/10/2018 : <https://twitter.com/GirardinYG/status/1052223734860394496i>

Le travail d'analyse a consisté en une analyse thématique descendante du discours des corpus 1 et 2 (texte et description du contenu des vidéos avec verbatim). Pour chaque portion de texte, nous en avons inféré les concepts et les pratiques en matière d'éducation aux données. Ensuite nous les avons classés selon la matrice de Grillenberger et Romeike (op. cit.). Les corpus 3 et 4 ont été utilisés, lors de la synthèse de l'interprétation pour émettre des hypothèses sur le travail effectivement réalisé par les élèves.

Tableau. 1. Extrait des témoignages des élèves, retranscrits de la bande son de la vidéo interactive produite par le collège Achille Mauzan et le lycée Aristide Briand de Gap

"On est parti d'une base de données, la fiche synthèse des matricules. C'était un tableau Excel pas complété. On a dû faire des recherches pour le compléter. On avait leur matricule, date de mort, de naissance, endroit d'où ils venaient, leur lieu de mort, des fois c'était pas en France."
"Le prof nous a donné des thèmes pour qu'on soit pas perdus. Par exemple : le premier mort et le dernier mort, le nombre de soldat par l'année de naissance, le déplacement dans les tranchées."
"On a commencé par certains sites : mémoire haute alpes, mémoire des hommes. On a trouvé toutes les infos. On pouvait accéder à une photo de leur pièce d'identité et chercher ce qui n'était pas forcément visible"
"On a fait des moyennes, et avec pictochar on a réussi à faire comme une présentation pour des affiches, et c'est plus agréable à lire. On a été voir des anciens combattants qui ont pu faire des témoignages. Il y avait des poèmes. Des lycéens ont vu notre travail."

4 Résultats

L'interprétation proposée porte sur la couverture par les activités du *datasprint* du domaine de l'éducation aux données, tel que modélisé par Grillenberger et Romeike. Une synthèse est proposée dans la figure 3.

La pratique « Collecter, modéliser et nettoyer » (P1) est très largement couverte par les activités proposées. Le corpus 1 précise que « le fichier de données fourni initialement doit être enrichi par un travail de *crowdsourcing* effectué en classe », qui met en œuvre la compétence de sélection et d'exploitation de sources de données (P1C1). Dans une section intitulée « récolter des données », le corpus 2 montre comment l'enseignant peut réaliser le jeu de données initial depuis le site du ministère des armées « Mémoires des hommes »⁶ et précise « Les élèves auront pour travail de compléter la base de données déjà constituées avec d'autres sources et de l'uniformiser pour pouvoir l'exploiter ». Il s'agit ici de sélectionner et formater les données pour respecter le modèle proposé (P1C2). Plusieurs vidéos montrent comment repérer les données manquantes dans un tableau de données et comment les compléter en allant les chercher sur des sites spécialisés. Il est également question d'aller collecter des données sur le terrain (auprès des mairies, des familles, etc.). Dans le corpus 3, le texte descriptif de la réutilisation précise que les élèves ont complété le jeu de données des causes de la mort, présentes sur la fiche numérisée accessible via « Mémoire des Hommes », et des métiers des

⁶ <https://www.memoiredeshommes.sga.defense.gouv.fr/>

Poilus, trouvés dans les registres matricules. Dans le corpus 4, une élève témoigne d'une pratique effective de collecte de données pour compléter un fichier au moyen d'un tableur (v. tableau 1, ligne 1). Enfin, nous associons à la compétence d'analyse lors de la collecte des données (P1C3) la vérification de la fiabilité des données. Celle-ci est abordée dans les deux corpus. Le corpus 1 précise que « Repérer la donnée fiable, connaître les sources primaires, vérifier, croiser les informations secondaires sont des compétences développées à travers le *datasprint* ». Le corpus 2 reprend la notion de sources primaires pour insister sur leur fiabilité, et les processus de vérification.

Pratiques Concepts	P1 Collecter, modéliser et nettoyer	P2 Implémenter et optimiser	P3 Analyser, visualiser et interpréter	P4 Partager, archiver et effacer
C1 Données, informations				
C2 Accès et stockage des données				
C3 Analyse des données				
C4 Éthique et sécurité des données				

Compétences couvertes
 Compétences non couvertes

Fig. 3. Compétences relevant de l'éducation aux données dans le *datasprint* Traces de soldats

La pratique « Implémenter et optimiser » (P2) est également couverte par les activités proposées. Dans une section intitulée « Traitement des données », le corpus 2 propose des tutoriels vidéo détaillant comment réaliser pratiquement les recherches d'informations complémentaires, par *web scraping* ou par recherche d'informations et leur ajout dans le jeu de données ; comment mettre en forme le jeu de données (scinder des données en plusieurs colonnes) ; comment réaliser certains calculs statistiques simples avec un tableur ; comment filtrer et trier des données. Ces activités relèvent de l'implémentation et l'optimisation de la mise en forme des données pour la collecte (P2C1) et pour l'analyse (P2C3). Il est précisé : « D'autres outils [que le tableur] existent mais ils sont plus difficiles à prendre en main pour des élèves et le tableur est un outil au programme dès le collège ». Les copies d'écrans du corpus 3 (v. figure 2) montrent des travaux d'élèves en cours de réalisation avec un tableur, laissant supposer qu'ils ont réalisé ce type de traitement sur les données. On peut supposer que le travail collaboratif a amené les élèves et les enseignants à choisir les formats de fichiers à partager (P2C2).

La pratique « Analyser, visualiser et interpréter » (P3) est centrale dans les intentions énoncées par le corpus 1. Les élèves doivent « être en mesure d'analyser les données disponibles, de leur donner un sens, de formuler des hypothèses, puis de les vérifier en

opérant des choix : quelles données garder, quelles données compléter, quelles données vérifier, croiser. » (P3C1) Cette analyse et ces choix sont guidés par l'objectif de produire une ou plusieurs visualisations de données. Dans une section intitulée « faire parler les données : datavisualisation », le corpus 2 détaille les différentes représentations graphiques et leur utilisation. Pour chacune, sont énoncées des règles permettant d'en améliorer la lisibilité et une vidéo présente un kit pédagogique sur le design graphique qui introduit les notions de typographie, couleur, mise en page. Ceci doit permettre aux élèves de choisir les visualisations adaptées à leur objectif (P3C3).

La pratique « Partager, archiver et effacer » (P4) est absente des corpus 1 et 2. Pourtant le projet a comporté une activité de partage des données et des visualisations produites sur le site data.gouv.fr. Il semble que la préparation du jeu de données, les choix afférents au partage, et le dépôt ont été réalisés par les membres des ateliers Canopé. Nous n'avons pas d'information sur la façon dont les élèves et les enseignants y ont été associés.

Les concepts relevant de l'éthique et de la sécurité des données (C4) ne sont pas abordés dans le corpus. Ceci peut s'expliquer par le sujet lui-même : les données portent sur des individus décédés pour lesquels les questions de protection de données personnelles et d'anonymisation ne se posent pas⁷.

5 Discussion et perspectives

L'ensemble des supports analysés dans notre étude exploratoire est constitué d'un document que l'on pourrait qualifier d'intention (corpus 1), d'une trame qui explore de manière quasi-exhaustive les activités pouvant être réalisées avec les élèves et qui fournit des ressources pour les mener à bien (corpus 2), de différentes traces des activités des élèves de deux établissements (corpus 3 et 4). Il s'agit des productions des élèves, publiées sur le site des données ouvertes data.gouv.fr par le réseau Canopé, des photographies d'écran d'ordinateur prises par l'enseignant pendant une séance de travail avec ses élèves, et de témoignages des élèves enregistrés pour la bande son d'une production interactive. Nous n'avons pas réalisé d'observation directe de l'activité.

Le corpus 1 comporte beaucoup d'expressions pour qualifier l'approche pédagogique et ses objectifs, que l'on ne retrouve pas dans l'enseignement disciplinaire de SNT : la rupture avec l'approche disciplinaire des savoirs (même si le programme de SNT incite les enseignants de cette discipline à collaborer avec des collègues d'autres disciplines), la co-réalisation en équipe, la participation à un commun numérique, le développement de l'esprit critique, l'expérience d'éducation par la recherche, etc. Nous avons relevé ce qui nous apparaît comme une contradiction entre les corpus 1 et 2. Le premier précise que « Paradoxalement les compétences les plus largement travaillées relèvent moins des habiletés numériques que des compétences critiques », alors que le corpus 2 comporte des tutoriels présentant des habiletés techniques assez poussées de

⁷ Voir sur le site de l'INSEE: <https://www.data.gouv.fr/fr/datasets/fichier-des-personnes-decedees/>

recherche d'information sur le Web, d'utilisation d'outils dédiées ou de l'outil tableur. Cette contradiction est en fait révélatrice de l'adaptabilité offerte par le *datasprint* en termes d'objectifs pédagogiques et d'appropriation par les enseignants et les élèves. L'analyse des déclinaisons d'un même *datasprint* serait intéressante à mener du point de vue de la discipline enseignée par l'enseignant qui le propose. Les corpus 3 et 4 n'ont pas été analysés systématiquement, mais il ressort des similitudes dans les réalisations. Si la ville concernée change, ce sont les mêmes questions qui sont posées, les mêmes informations complémentaires qui sont recherchées. Dès lors, on peut s'interroger sur l'objectif de créativité énoncé dans le corpus. On pourrait parler d'autonomie guidée, mais aussi contrainte par ce que permet ou non le jeu de données.

En tant que dispositif d'éducation aux données, l'analyse des corpus et des productions réalisées montre que le *datasprint* « Traces de soldats » s'apparente à la conception d'une situation authentique d'apprentissage. Duval et Pagé [4] synthétisent, à partir d'une revue de la littérature, les cinq caractéristiques de ces situations par les traits suivants que nous déclinons dans le cas de « Traces de soldats ». (1) Le contexte d'apprentissage a un caractère réaliste. Travailler avec des données historiques et locales assure le caractère réaliste de la situation dans le cas de « Traces de soldats ». (2) La situation demande l'accomplissement d'une réalisation plutôt qu'une simple reprise d'information. Toutes les actions des élèves sont ici sous-tendues par l'objectif de réalisation d'une production partageable, diffusable et contribuant à la production de connaissances. (3) La situation authentique propose des tâches complexes qui favorisent le jugement et l'innovation. L'analyse du corpus montre que les élèves sont sollicités tout au long d'un processus assez complet d'analyse de données : depuis le choix des sources de collecte de nouvelles données, jusqu'à l'interprétation permise par une représentation graphique. Les différentes pratiques de l'éducation aux données, à l'exception du partage, sont mobilisées dans « Traces de soldats », nécessitant des habiletés informatiques développées. Les compétences, dans le *datasprint*, deviennent effectives, celles-ci résultant « de la mobilisation, de la sélection, de la coordination, de la mise en œuvre et des nombreux ajustements des ressources utiles pour le traitement des tâches dans une situation donnée ou une classe de situations » [6, p. 40]. Néanmoins, l'analyse du corpus ne permet pas toujours de savoir quelle est la part d'autonomie dans les choix réalisés par les élèves. (4) La situation authentique nécessite une consultation entre les élèves et une rétroaction en vue de l'amélioration de la réalisation. Cette caractéristique n'est pas vraiment identifiable dans notre corpus. En effet, nous n'avons pas d'information sur les modalités d'accompagnement des apprentissages par les enseignants ou les évaluations menées. (5) La situation authentique génère une forte motivation, qui dépasse le désir d'obtenir une bonne note. Ce point, comme le précédent, n'est pas directement identifiable dans l'analyse de nos corpus, puisqu'il nécessiterait d'observer l'activité effectivement menée.

Notre analyse exploratoire montre la richesse de ces situations d'apprentissage en matière d'éducation aux données que fournissent les *datasprints*. Cependant, nous nous interrogeons pour savoir si le cas emblématique que nous avons étudié est représentatif

des *datasprints*. Une étude serait à mener pour savoir quels rôles jouent la mise à disposition de ressources et la mise en place d'ateliers dans l'adoption par les enseignants de cette modalité pédagogique. Nous avons également souligné les limites de notre étude qui ne se base que sur une partie des ressources et des productions d'élèves disponibles. Outre l'activité des élèves, l'activité conjointe enseignants-élèves lors du déroulement de ces *datasprints* constitue une piste de recherche.

Enfin si le *datasprint* étudié couvre la quasi-totalité des compétences proposées par le modèle de Grillenberger et Romeike, celui-ci ne prend pas en compte les compétences liées à la réutilisation de données collectées par d'autres et la question de la vérification de la fiabilité des sources. Les pistes évoquées pourront contribuer à l'élaboration d'une didactique de la donnée.

Contribution

Cette recherche participe des travaux menés par le GIS 2IF (Innovation, interdisciplinarité et formation) dans le cadre du Groupe Thématique Numérique 2020-2022 « Humanités numériques : entre recherche et formation », bénéficiant du soutien du Ministère de l'Éducation nationale, de la Jeunesse et des Sports.

Références

1. Bodin, F. (2018). « Apprendre avec les données numériques, les *datasprints* pédagogiques : le cas de Traces de Soldats ». *Medium*. 17 octobre 2018.
2. Crusoe, D. (2016). Data Literacy defined pro populo : To read this article, please provide a little information. *The Journal of Community Informatics*, 12(3).
3. Drot-Delange, B., Tort, F. (à paraître). Éducation aux données vs. Enseignement des données : une contribution aux humanités numériques à l'école ? *Humanités numériques*, vol. 5, « Enseigner et apprendre les humanités numériques ».
4. Duval, A.-M., Pagé, M. (2013). *La situation authentique : de la conception à l'évaluation. Une formule pédagogique pour toutes les disciplines*. Montréal. AQPC
5. Grillenberger, A. et Romeike, R. (2018). « Developing a theoretically founded data literacy competency model ». Dans *Proceedings of the 13th Workshop in Primary and Secondary Computing Education*, 1-10.
6. Jonnaert, P., & Vander Borght, C. (2003). Créer des conditions d'apprentissage : Un cadre de référence socioconstructiviste pour une formation didactique des enseignants. De Boeck.
7. Womack, Ryan. 2014. « Data Visualization and Information Literacy ». *IASSIST Quarterly* 38 (1) : 12-17.