



**HAL**  
open science

## Former à la littératie de données : les stratégies de réduction de la complexité

Camille Capelle, Marie Chagnoux

### ► To cite this version:

Camille Capelle, Marie Chagnoux. Former à la littératie de données : les stratégies de réduction de la complexité. La numérisation des sociétés, SFSIC, Jun 2023, Bordeaux, France. hal-04598702

**HAL Id: hal-04598702**

**<https://hal.science/hal-04598702>**

Submitted on 3 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

# **Former à la littératie de données : les stratégies de réduction de la complexité**

## **Training in Data Literacy: strategies for reducing complexity**

Camille CAPELLE, Maître de conférences en Sciences de l'Information et de la Communication, IMS UMR CNRS 5218, Université de Bordeaux  
[camille.capelle@u-bordeaux.fr](mailto:camille.capelle@u-bordeaux.fr)

Marie CHAGNOUX, Maître de conférences en Sciences de l'Information et de la Communication, CEMTI, Université Paris 8  
[marie.chagnoux@univ-paris8.fr](mailto:marie.chagnoux@univ-paris8.fr)

### **Mots clés**

Littératie des données, transdisciplinarité, formation, complexité, culture numérique

### **Keywords**

Data literacy, transdisciplinarity, training, complexity, digital literacy

### **Résumé**

A partir de premières explorations menées dans le cadre du GTNum DEFI (Données pour l'Education, la Formation et l'Innovation), nous proposons une analyse d'un corpus de formation à la littératie des données et d'une enquête menée auprès des enseignants. Nous montrons comment la complexité, intrinsèque à cette littératie, implique des stratégies de simplification pour les formateurs.

### **Summary**

Based on initial explorations carried out within the framework of the GTNum DEFI (Data for Education, Training and Innovation), we propose an analysis of a data literacy training corpus and a survey addressed to teachers. We show how the complexity, intrinsic to this literacy, implies simplification strategies for trainers.

# **Littératie des données : une nécessaire réduction de complexité en contexte pédagogique ?**

Camille CAPELLE

Marie CHAGNOUX

La littératie des données, ou *data literacy*, désigne la capacité à comprendre les enjeux de la production, de l'organisation et de l'exploitation des données, et à les utiliser efficacement, de manière critique et créative (Schild, 2004). Elle fait désormais partie intégrante des référentiels de compétences sur le numérique, de politiques de formation, de pratiques pédagogiques dans l'enseignement, du primaire au supérieur, ainsi que de travaux de recherches (Lehmans et Capelle 2016, Daunay et Fluckiger, 2018, Arruabarrena et al, 2019, Verdi et Le Deuff, 2020, entre autres) mais sa mise en œuvre dans les établissements scolaires et dans la formation des enseignants reste encore insuffisamment documentée. En réponse à cela, le groupe thématique numérique sur la littératie des données, GTnum DEFI (Données pour l'Éducation, la Formation, l'Innovation)<sup>1</sup>, initie depuis 2022 une recherche-intervention pour analyser et suggérer des pistes visant à développer une culture des données des élèves. Il se déploie sur trois axes qui constituent la base de groupes de travail spécifiques. Nous nous proposons ici de présenter des premières pistes d'exploration de l'axe 2 portant sur la médiation et la médiatisation des savoirs avec les données que nous corroborons à plusieurs terrains de recherche sur la formation aux données. Nous exposerons les principes méthodologiques adoptés pour ces terrains et, à partir de premières observations des acteurs de l'éducation appréhendant la question de la formation aux traitements des données, ainsi que des résultats d'une enquête menée auprès de 736 enseignants, nous montrerons comment la complexité, intrinsèque à cette littératie, implique des stratégies de simplification pour les formateurs.

## **1. Appréhender les logiques de formation**

Les analyses proposées dans ce travail mobilisent plusieurs terrains complémentaires résumés dans le tableau 1 ci-dessous. Il s'agit principalement de formations destinées à faire appréhender à de futurs formateurs, ou professionnels, néophytes, des contextes concrets de

---

<sup>1</sup> Voir sur le carnet de recherche Edunumrech à l'adresse <https://edunumrech.hypotheses.org/4185#more-4185>

mise en œuvre de littératie de données. Chaque situation a donné lieu à une production finalisée de la part des participants à la formation (article, datavisualisation ou ressource multimédia).

Terrain	Projets	Contexte	Nombre de participants	Format
1	Projet MVDE IDEX CNRS (2016-2017)	Formation d'élèves à l'école primaire, en collèges et lycées dans différentes disciplines	25-30 élèves, 1 enseignant	Séance ou séquence pédagogique
2	Madona (2018-2022)	Formation au datajournalisme	9 étudiants, 1 formateur	Une semaine de formation intensive au datajournalisme assurée en Master par un professionnel
3	Madona (2018-2022)	Hackathon "Raconter Rennes par la donnée"	11 participants, 6 facilitateurs	2 jours de hackathon autour des données Open Data de Rennes, ouvert à tous (professionnels et amateurs)
4	DEFI (2021-2024)	Initiation au traitement de données	16 participants, 2 formateurs	2 jours de formation au traitement des données dans un Master Documentation
5	DEFI (2021-2024)	DataSpring "Femmes et médias"	Paris (5 participants, 1 formateur) INSPE Bordeaux (13 étudiants, 2 formateurs, 2 facilitateurs)	Une demi-journée de dataspring "Femmes et Média" proposée au national dans le cadre du Prontemps de l'EMI par Canopé

Tableau 1 – Présentation des terrains mobilisés

Pour chacun de ces terrains, plusieurs méthodologies ont été appliquées. Pour les terrains 2, 4 et 5, les participants ont rempli un journal de bord sous forme de « notebook » avec des catégories prédéfinies, destiné à formaliser les étapes de réalisation. Pour les terrains 3, 4 et 5, les formations ont fait l'objet d'observations détaillées et des entretiens ont été menés avec certains participants. Pour ces mêmes formations, un questionnaire en ligne portant sur les représentations des données dans l'enseignement a été soumis en amont des formations. Pour la présente étude qui porte sur les scénarios pédagogiques proposés par les formateurs, nous mobiliserons les documents et observations en étayant nos propositions par les données issues du questionnaire d'enquête.

Afin de mieux cerner ce à quoi renvoient les données pour les enseignants, selon leur niveau d'enseignement et leur discipline, nous avons élaboré un questionnaire d'enquête à destination de tous. Ce questionnaire comprend une première partie sur la définition des données et de leurs enjeux, ainsi que les pratiques d'enseignement et compétences du CRCN (Cadre de Référence des Compétences Numériques) liées aux données enseignées auprès des élèves. Une deuxième partie porte sur la connaissance des données ouvertes, leur utilité et facilités ou difficultés d'usage. Une troisième partie concerne la visualisation de données, ses usages, intérêts, facilités ou difficultés. Puis une dernière partie porte sur la formation aux données reçues et les besoins ressentis. Cette enquête a été diffusée en ligne et par des mails ciblés, entre mai 2022 et mars

2023, par différentes stratégies ayant pour but de toucher un maximum d'enseignants, qu'ils soient ou non aguerris et proches de par leur discipline, avec la thématique et le numérique au sens large. Dans un premier temps, la stratégie a été de diffuser l'enquête par messagerie aux participants aux différentes actions de médiation et de formation aux données, en amont de celles-ci. La plupart étaient professeurs documentalistes étudiants ou stagiaires. Dans un second temps, nous avons sollicité les DANE<sup>2</sup> et DRANE<sup>3</sup> pour une diffusion à un public plus étendu d'enseignants et dans plusieurs académies, mais ces démarches n'ont pas abouti. Le questionnaire a aussi fait l'objet d'une communication via le carnet de recherche de la Direction du Numérique pour l'Éducation et plus largement sur le réseau social Twitter. Nous avons pris le parti de diffuser plusieurs messages différents : certains mentionnaient le sujet de l'enquête, en l'occurrence la littératie des données, et d'autres pas, car nous pensions que l'intitulé aurait pu rebuter bon nombre d'enseignants qui n'étaient pas intéressés ou qui se sentaient trop éloignés du sujet. Aux vues du faible nombre de répondants, et surtout des disparités dans la représentation des différentes disciplines enseignées, nous sommes ensuite passés à une stratégie de communication plus ciblée, en contactant par messagerie des Inspecteurs Pédagogiques Régionaux de certaines disciplines absentes de l'enquête ainsi que des directions d'écoles primaires pour le premier degré d'enseignement. Au final, nous avons obtenu 736 réponses à l'enquête. Parmi ces enseignants, 19% sont professeurs des écoles alors que les autres enseignent en collège et lycées et 84% ont plus de 5 ans d'expérience dans le métier. Dans le second degré, les profils de répondants sont assez variés, malgré la surreprésentation des professeurs documentalistes, compte tenu de nos actions de formation auprès de ce public, et la faible participation de certains enseignants de disciplines telles que les sciences économiques et sociales, les arts plastiques, l'éducation musicale.

Par ailleurs, 71% d'entre eux se déclarent intéressés pour bénéficier d'une formation à l'usage pédagogique des données ouvertes, ce qui témoigne de leur intérêt pour le sujet et de la perception de sa complexité.

## **2. La complexité de la littératie de données et les stratégies pédagogiques de réduction en contexte de formation**

---

<sup>2</sup> Délégations académique au numérique éducatif

<sup>3</sup> Directions de Région Académique du Numérique pour l'Éducation

Du point de vue épistémologique, la formation à la littératie des données, à la suite de la formation à la culture numérique qui s'est construite notamment à travers l'EMI, l'éducation aux médias et à l'information, et autour des référentiels de compétences successifs (B2i, C2i, puis le CRCN avec la plateforme PIX<sup>4</sup>), constituerait une étape marquante dans une histoire de la pédagogie qui s'est toujours construite autour des domaines disciplinaires classiques. La particularité de la littératie des données réside en effet dans la complexité des objets et des compétences qu'elle appréhende, qui combinent, par-delà ses frontières académiques, des connaissances à la fois mathématiques (statistiques), technologiques, informatiques, informationnelles, sémiotiques et narratives. Elle relève ainsi de ces domaines émergents qui impliquent un changement de paradigme, par le recours à une épistémologie pluraliste qui sollicite des concepts et des méthodes transdisciplinaires que l'on trouve chez différents pédagogues et penseurs (Piaget, 1967 ; Morin, 1994 ; Nicolescu, 1996). Cette particularité constitue une gageure pour les formateurs qui ne peuvent y répondre que par une réduction de cette complexité. Cette réduction passe par différents types de stratégies que nous nous proposons de décrire et d'analyser dans cette proposition.

**2.1. Des formats « applicatifs » et collaboratifs sous forme de challenge ou de cas pratiques**  
« Datasprint », « hackathon », « datachallenge », « carto » ou « cryptoparty »... ces néologismes désignent les formats pédagogiques couramment adoptés dans les formations, décrits pour certains comme le datasprint pédagogique (Bodin, 2018 ; Drot-Delange & Tort, 2022) et qui ont été adoptés pour 4 terrains sur 5 de notre corpus. Ils désignent des temps courts de collaboration qui mettent à profit les compétences et les connaissances des participants avec les données dans un esprit d'émulation. Plusieurs travaux ont néanmoins montré les limites théoriques de ces formats, qui participent à l'invisibilisation de la « fabrique des données » et se focalisent de façon quasi-exclusive sur l'exploration et la visualisation (Gruson-Daniel & de Quatrebarbes, 2017).

## **2.2. Une familiarisation via la datavisualisation**

Parmi les enseignants interrogés dans notre enquête, 57% d'entre eux affirment réaliser des datavisualisations avec leurs élèves. Très souvent, en effet, même dans le cadre de formations plus longues et plus institutionnelles comme plusieurs des nôtres [Terrains 1, 2], l'initiation aux données s'inscrit dans une perspective et une démarche simplifiée de design de l'information

---

<sup>4</sup> <https://pix.fr/>

(Capelle *et al.*, 2018 ; Chagnoux, 2020). Il s'agit de sélectionner et communiquer des données dans une visée essentiellement fonctionnelle, sémiologique et esthétique. Pour une grande majorité des enseignants, la datavisualisation présente plusieurs intérêts auprès des élèves car elle aide à comprendre une information pour 87%, elle paraît plus facile à lire qu'un texte pour 77% et elle peut susciter l'envie de lire un texte pour 74% des enseignants. Ce format s'inscrit donc dans une volonté pédagogique de rendre plus accessibles et plus attractives des données auprès des élèves.

### **2.3. Une focalisation sur les statistiques**

Selon les disciplines et les approches, c'est moins l'intelligibilité de la représentation des données que la représentation ou les chiffres eux-mêmes qui sont au centre de la réflexion et des compétences visées. C'est notamment le cas chez les enseignants de Sciences Économiques et Sociales observés dans le terrain 1. En cela, on peut se demander ce qui différencie l'enseignement des statistiques de l'éducation aux données, qui semblent parfois confondues (Erickson, 2017). On remarque en effet que dans l'imaginaire des enseignants, le mot « données » renvoie pour beaucoup au champ sémantique des statistiques à travers l'association à des termes tels que « chiffres », « chiffré », « mathématique », « statistique », etc., à côté des champs sémantiques de l'information et de l'informatique.

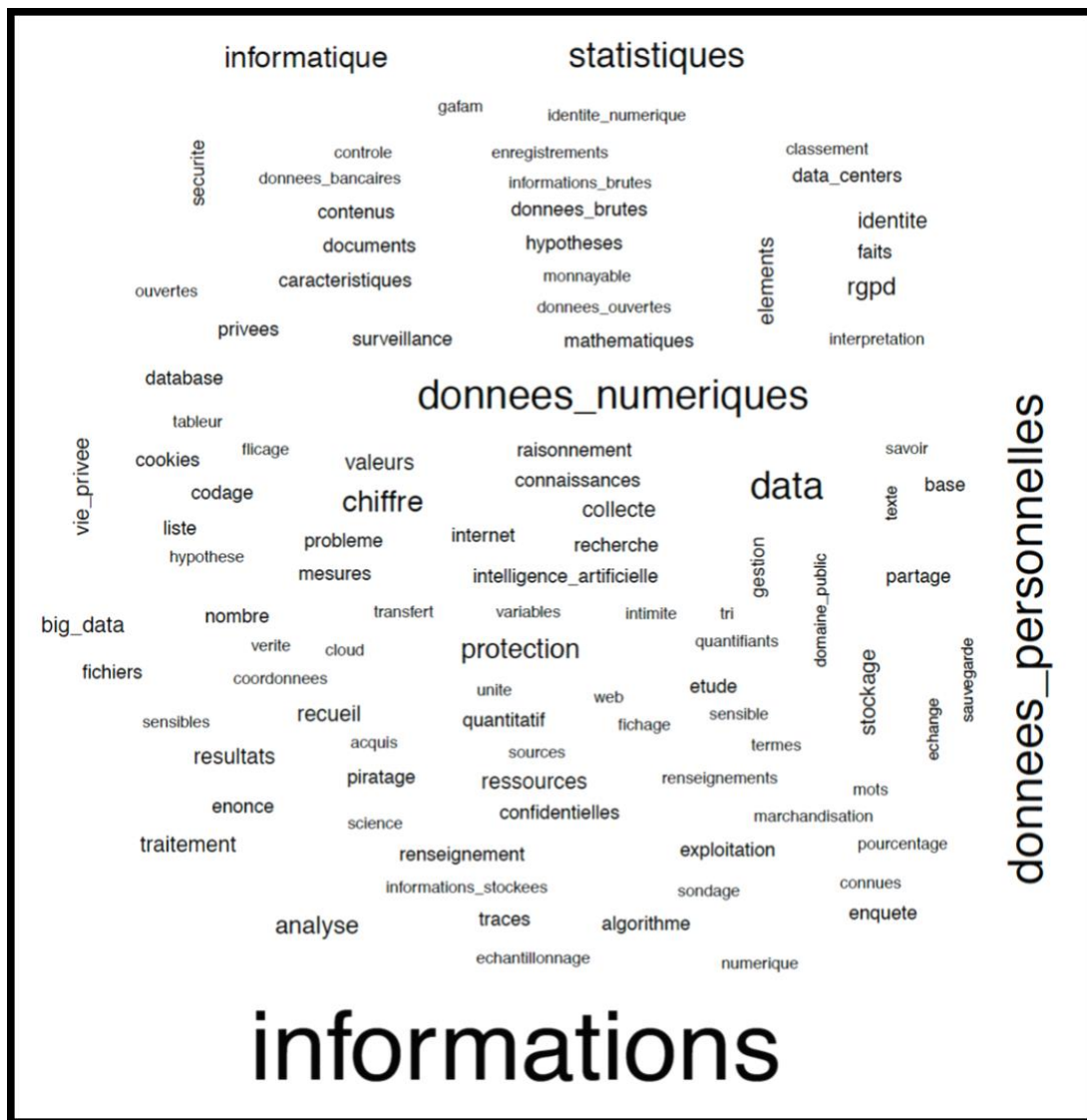


Figure 1 – Nuage de mots associés aux données par les enseignants

## 2.4. Une réinscription dans des cadres disciplinaires

Le terrain 4 confirme aussi l’hypothèse des contraintes de temps et du cadrage disciplinaire qui limitent la portée des initiations (Capelle et Lehmans, 2018 ; Capelle *et al.*, 2018) : ”[l]e processus d’utilisation de données ouvertes par des élèves dans un projet pédagogique apparaît comme trop long et trop complexe, alors que les compétences autour des données ne sont pas au cœur des programmes scolaires”. Non seulement les jeux de données vont souvent être adaptés par les formateurs, mais ils sont mis au service de scénarios fortement empreints de la culture disciplinaire native de l’enseignant. Bien que les données soient abordées en classe d’un point de vue disciplinaire, on remarque dans l’enquête qu’il s’agit bien d’un objet transversal. En effet, les enseignants les considèrent pouvoir être abordées dans de nombreuses disciplines telles que les SVT (pour 85%), l’histoire géographique (pour 81%) ou le français (pour 60%),



même si les mathématiques (pour 91%) et l'informatique (pour 89%) dominent sur le reste des disciplines. Il apparaît également que parmi les enseignants qui ont répondu et qui s'intéressent donc pour la plupart d'entre eux, au numérique et aux données dans l'enseignement, tous ne viennent pas de parcours scientifiques (40% d'entre eux seulement). 30% d'entre eux ont un parcours littéraire, 14% dans les sciences économiques et sociale et 15% mixte ou autre. Cela confirme en partie l'hypothèse selon laquelle l'éducation aux données représente un domaine transdisciplinaire, qui peut être partagé par tous les enseignants.

### **2.5. Une fragmentation de traitement**

Enfin, la contrainte temporelle ou la spécialisation de l'enseignant est segmentante. L'approche de l'éducation aux données est généralement parcellaire puisque le travail va porter soit sur la recherche de données, soit sur la vérification et la validation, soit sur la production d'un livrable à partir de données déjà traitées. Mis à part le terrain 3 qui incitait à mobiliser toute la chaîne de traitement, toutes les observations illustrent des focalisations sur certaines étapes, avec la préparation de jeux de données pré-définis. D'ailleurs, parmi les difficultés évoquées par les enseignants interrogés dans l'enquête, la première est d'avoir à corriger et nettoyer des jeux de données ouvertes. En effet, cette étape très chronophage n'est jamais réalisée avec les élèves en formation alors qu'elle fait pourtant partie du travail nécessaire lorsqu'il s'agit de réutiliser un jeu de données ouvertes comme ceux disponibles sur les plateformes d'open data. Ainsi, le processus global, qui consiste à passer d'un jeu de données ouvertes, brut, à la production d'informations par le traitement et la mise en forme des données, en passant par la formulation d'hypothèses n'est que très rarement mis en œuvre (Capelle *et al*, 2018). L'approche qui consiste à segmenter le processus avec les élèves ne permet donc pas de les mettre en situation réelle face aux données, ni d'appréhender l'ensemble des compétences de la littératie des données nécessaires à leur autonomie.

## **3. Discussion**

Ces premiers résultats montrent qu'une réflexion reste à mener pour accompagner les enseignants et formateurs dans "leur initiation à la complexité" (Arduino, 2000:§16) afin qu'ils puissent aborder la littératie des données non seulement en termes de compétences techniques, appréhendables dans des cadres de référence, mais surtout en terme de savoirs et savoir-faire (Verdi & Le Deuff, 2020) dans une approche transdisciplinaire, « entre, à travers et au-delà de toute discipline » (Nicolescu, 1996). Si la culture des données ambitionne d'être effectivement

ce creuset où l'esprit critique rencontre le pouvoir d'agir, l'*empowerment* citoyen, il est nécessaire d'imaginer et de mettre à la disposition des acteurs de l'éducation des leviers de formation aux enjeux éthiques, politiques, socio-économiques, juridiques, culturels et rhétoriques. Comme l'y invite Alhadeff-Jones (2008), la notion de complexité de l'objet d'études est aussi l'occasion de se soustraire aux habitus d'apprentissage pour élaborer de nouveaux modèles pédagogiques, pour que les transformations étudiées participent aussi des transformations des pratiques. L'appropriation de l'offre de données par les enseignants se heurte donc à d'importants enjeux en termes de médiation qu'il s'agira d'explorer davantage dans notre recherche.

### **Bibliographie**

Alhadeff-Jones, M. (2008). Trois générations de théories de la complexité : Nuances et ambiguïtés. Traduit et adapté par l'auteur de « Three generations of complexity theories: Nuances and ambiguities », *Educational Philosophy and Theory of Education*, 40(1), 66-82.

Ardoino, J. (2000). Éducation et politique aux regards de la pensée complexe. Dans : J. Ardoino, *Les Avatars de l'éducation : Problématiques et notions en devenir* (pp. 109-134). Paris cedex 14: Presses Universitaires de France.

Arruabarrena, B., Kembellec, G., & Chartron, G. (2019). Data littératie & SHS : développer des compétences pour l'analyse des données. In *CODATA*. Mars 2019, Marne la Vallée, France.

Bodin, F. (2018). « Apprendre avec les données numériques, les datasprints pédagogiques : le cas de Traces de Soldats ». *Medium*. 17 octobre 2018

Capelle, C., Jutand, M. A., & Morandi, F. (2018). Stratégies pédagogiques pour diffuser la culture des données ouvertes. Numéro coordonné par Christian Marcon, *COSSI*, 19. URL : <https://revue-cossi.numerev.com/articles/revue-4/1829-strategies-pedagogiques-pour-diffuser-la-culture-des-donnees-ouvertes>

Capelle, C. & Lehmans, A. (2018). Open data et émergence de communs de la donnée dans les territoires : les nécessaires médiations. *Communication & management*, 15, 87-99. <https://doi-org.docelec.u-bordeaux.fr/10.3917/comma.151.0087>

Chagnoux, M. (2020). La datavisualisation, double point d'entrée du datajournalisme dans la PQR. *Interfaces numériques*, 9(3).

Daunay, B., & Fluckiger, C. (2018). Écriture et numérique : pourquoi et comment parler de littéracie numérique. *Recherches*, 69, 71-86.

Drot-Delange, B. & Tort, F. (2022). Les datasprints, un dispositif d'éducation aux données ? Une étude exploratoire via le cas de " Traces de Soldats ". *Didapro 9 - DidaSTIC. L'informatique, objets d'enseignement et d'apprentissage. Quelles nouvelles perspectives pour la recherche ?* Mai 2022, Le Mans, France.

Erickson, T. (2017). Smelling Like Data Science – A Best-Case Scenario. <https://bestcase.wordpress.com/2017/02/21/smelling-like-data-science/>

Gruson-Daniel, C., & de Quatrebarbes, C. (2017). Les préparatifs d'un hackathon recherche : au coeur de la fabrique des données. *Sociologie et sociétés*, 49(2), 201-221. <https://doi.org/10.7202/1054279a>

Lehmans, A., & Capelle, C. (2016). Gouvernance et horizon d'attente des données ouvertes pour l'éducation. In *XIIe Colloque International EUTIC 2016 Réseau de recherche européen et pluridisciplinaire sur les enjeux et usages des TIC Logiques de réseaux et nouvelles gouvernances*, Université Nationale et Kapodistrienne d'Athènes.

Morin, E. (1994). « Interdisciplinarité et transdisciplinarité ». *Transversales, Science, Culture*, n° 29, p. 4-8.

Nicolescu, B. (1996). *La transdisciplinarité*. Paris : Éditions du Rocher.

Piaget, J. (1967). « Le système et la classification des sciences ». In J. Piaget (Ed.) *Logique et connaissance scientifique*, Paris : Gallimard. p. 1151-1224.

Schild M. (2005). Information Literacy, Statistical Literacy and Data Literacy. *IASSIST Quarterly, International Association for Social Science Information Service and Technology* 28 (2–3), 7–14.

Verdi, U., & Le Deuff, O. (2020). La data literacy distribuée. *Les Cahiers du numérique*, 16(2), 137-173.