



**HAL**  
open science

**GTnum IMS DEFI #DEFI – Formation aux données en  
Master MEEF Documentation - Groupes thématiques  
numériques de la Direction du numérique pour  
l'éducation (Ministère de l'Éducation nationale et de la  
Jeunesse) 2021-2024**

Camille Capelle

► **To cite this version:**

Camille Capelle. GTnum IMS DEFI #DEFI – Formation aux données en Master MEEF Documentation - Groupes thématiques numériques de la Direction du numérique pour l'éducation (Ministère de l'Éducation nationale et de la Jeunesse) 2021-2024. 2024. hal-04851643

**HAL Id: hal-04851643**

**<https://hal.science/hal-04851643v1>**

Submitted on 20 Dec 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.



Open licence - etalab



MINISTÈRE  
DE L'ÉDUCATION  
NATIONALE

*Liberté  
Égalité  
Fraternité*

Direction du numérique  
pour l'éducation



THÉMATIQUE

Littératie des données : évolution des  
compétences, enjeux éducatifs,  
pédagogiques et éthiques,  
perspectives pour la formation

# Formation aux données en Master MEEF Documentation

#GTnum DEFI  
2021-2024



# FORMATION AUX DONNEES EN MASTER MEEF DOCUMENTATION CAHIER D'EXPERIENCE

## Les auteurs / autrices

Camille Capelle, Université de Bordeaux-INSPE, IMS-RUDII

## Projet piloté par

Camille Capelle, Université de Bordeaux-INSPE, IMS-RUDII

Ces travaux sont publiés dans le cadre des groupes thématiques numériques soutenus par la Direction du numérique pour l'éducation.

• Eduscol <https://eduscol.education.fr/2174/enseigner-et-apprendre-avec-la-recherche-les-groupes-thematiques-numeriques-gtnum>

• Carnet Hypothèses « Éducation, numérique et recherche » <https://edunumrech.hypotheses.org/>

Novembre 2024

Conditions d'utilisation :  sauf indication contraire, tout le contenu de ce document est disponible sous [Licence Ouverte 2.0](https://creativecommons.org/licenses/by-nc-sa/2.0/)

# SOMMAIRE

---

Sommaire .....	3
Introduction.....	4
1 /// Présentation de l'expérience.....	5
Contexte et objectifs de l'action de formation.....	5
Caractérisation des publics participants.....	6
Consignes et modalités de formation .....	6
Les supports et outils à disposition des participants .....	7
Les productions des participants.....	8

---

2 /// Collecte de données .....	10
Dispositifs d'observation et de collecte.....	10
Itinéraires d'apprentissage sur et avec les données .....	12

---

3 /// Analyse critique du développement d'une littératie des données en formation .....	14
Techno-imaginaires liés aux données et manières de les déconstruire en formation.....	14
Développement d'une réflexion critique en Éducation aux médias et à l'information (EMI) .....	15

---

Références .....	16
------------------	----

# INTRODUCTION

L'axe 2 du [GTnum DEFI](#) (Données pour l'Education, la Formation, l'Innovation) s'intéresse aux médiations et médiatisations des savoirs autour des données. Dans ce cadre, nous avons participé à la mise en œuvre et au suivi durant 3 ans (2021-2024) de plusieurs expérimentations de formation aux données chez des publics variés. Chaque expérimentation a des objectifs singuliers et chacun des cahiers d'expérience rend compte du déroulement situé de l'expérimentation, ainsi que des manières singulières dont les formateurs et les participants ont développées des pratiques et savoirs info-communicationnelles à partir et autour de données (bases de données, données ouvertes (open data), statistiques sur l'orientation, données géolocalisées sur des cartes). Dans chacun des cahiers, un vocabulaire spécifique à la littératie des données est employé. Si vous souhaitez approfondir ce vocabulaire, nous vous invitons à vous référer au glossaire de la formation aux données, produit dans le cadre du GTnum DEFI.

# 1 /// PRESENTATION DE L'EXPERIENCE

## Contexte et objectifs de l'action de formation

Notre expérience se situe dans le cadre de la formation des futurs professeurs documentalistes en INSPE. Elle a été menée de 2021-2022 à 2023-2024, en collaboration avec deux médiatrices de l'atelier Canopé de Bordeaux puis avec un médiateur, membre de la fondation Open Street Map France. Les étudiants de master MEEF se destinent au métier de professeur documentaliste, métier qui repose à la fois sur des tâches de gestion et de valorisation de fonds documentaires, mais également à la sensibilisation des élèves aux enjeux propres à l'Éducation aux Médias et à l'Information (cf. circulaire de mission n° 2017-051 du 28-3-2017). Parmi ces enjeux, la compréhension du cycle de vie de la donnée, et de la manière dont des données sont produites et exploitées pour permettre de produire de l'information est essentielle. La dimension communicationnelle est également au cœur de cette formation qui implique de mener une réflexion sur les ressources libres de droit, les communs de la connaissance, l'interopérabilité des données et la fiabilité de l'information dans une perspective d'Éducation aux médias et à l'information. La formation s'inscrit dans le cadre d'une UE intitulée "Pédagogie, didactique, culture de l'information et des médias".

### Compétences du CRCN et CRCN-É

Les compétences visées sont celles du Cadre de référence des compétences numériques (CRCN) qui correspondent au niveau 4 de la compétence 1.3 « Traiter des données » :

- « Traiter des données pour analyser une problématique »
- « Appliquer une fonction statistique simple ».

Elles concernent aussi le Cadre de référence des compétences numériques pour l'Éducation (CRCN-É) relatives au domaine 2 « Ressources Numériques » :

Compétence 2.2. « Modifier et produire des ressources numériques

- Adapter des ressources numériques existantes en tenant compte des licences et des droits associés
- Créer ou co créer de nouvelles ressources éducatives numériques en fonction d'objectifs pédagogiques, des apprenants et du contexte d'apprentissage

## Caractérisation des publics participants

Les groupes se composent en moyenne de 30 étudiants par an, M1 et M2 réunis. Leurs formations de base sont assez diversifiées (parcours en lettres, histoire, sociologie, droit, psychologie, information-communication...). Aucun ne provient de parcours en sciences ou mathématiques. Leur appétence présumée avec les données est donc assez faible. Du côté des formateurs participants à l'expérimentation, il s'agit :

Pour les années 1 et 2 : de deux enseignants chercheurs en sciences de l'information de la communication, deux médiatrices du réseau Canopé,

Pour l'année 3 : de deux enseignants chercheurs en sciences de l'information de la communication, et d'un entrepreneur, médiateur, membre du Conseil d'Administration de la Fondation Open Street Map France et chercheur en géographie.

## Consignes et modalités de formation

Les objectifs de la formation pour les étudiants étaient, de savoir définir les données et ce que sont les données ouvertes, de comprendre le cycle de vie de la donnée à l'information et au document, d'être capable de rechercher des données ouvertes sur des plateformes d'open data, de les interroger, puis de sélectionner des données pour produire une visualisation de données. La formation aux données se composait de 2 à 3 séances ayant pour but de préparer les étudiants à la réalisation d'un challenge pédagogique final :

- un datasprint (années 1 et 2) : 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 (Bodin, 2018)
- une cartoparty (année 3) : balade urbaine qui vise à arpenter une zone géographique à plusieurs, afin de produire des données géo-référencées (...) on mesure, on capte, on photographie, on saisit les données récoltées pour les mettre en commun (l'Her et al. 2018) sur une base de données collaborative.

Voici ci-dessous le descriptif des séquences pour les 3 années d'expérimentation.

### Année 1 :

Séance 1. Introduction à la culture des données, données ouvertes et élaboration d'un scénario pédagogique à partir d'un jeu de données en Open Data – 4h

Séance 2. Data-viz : théorie et pratique – 6h

Séance 3. Datasprint pédagogique sur « La place des femmes dans les médias » proposé par le réseau Canopé dans de nombreux ateliers en France, à distance, et en présentiel à l'INSPE de Bordeaux sur une demi-journée - 4h

Consigne : explorer les données et proposer des visualisations éclairant une question à partir du jeu de données concernant le temps de parole des hommes et des femmes à la télévision et à la radio mis en ligne par l'Institut National de l'Audiovisuel (INA) :

<https://www.data.gouv.fr/fr/datasets/temps-de-parole-des-hommes-et-des-femmes-a-la-television-et-a-la-radio/#/community-reuses>

### **Année 2 :**

Séance 1. Introduction à la culture des données, données ouvertes, présentation de l'OSINT (Open Source Intelligence) - 4h

Séance 2. Traitement de données sur tableur (Excel) - 6h

Séance 3. Datavisualisations et design de l'information - 6h

Séance 4. Datasprint « Traces 2 roues » proposé par le réseau Canopé - 4h

Consigne : explorer les données et proposer des visualisations éclairant une question à partir du jeu de données des accidents corporels de la circulation routière incluant des véhicules à deux roues (2005 à 2022) mis en ligne par le Ministère de l'intérieur : <https://www.data.gouv.fr/fr/datasets/bases-de-donnees-annuelles-des-accidents-corporels-de-la-circulation-routiere-annees-de-2005-a-2022/>

### **Année 3 :**

Séance 1. Données ouvertes et construction d'un scénario pédagogique à partir d'un jeu de données ouvertes - 4h

Séance 2. Traitement de données sur tableur (Excel) - 6h

Séance 3. Dataviz (design de l'information) - 3h

Séance 4. Cartoparty animé par un médiateur Open Street Map - 5h

Consigne : aller explorer le site de formation avec les plans pour y relever tout ce qui n'apparaît pas sur le plan fourni. Ajouter ensuite les données recueillies dans la [base de données Open Street Map](#).

## **Les supports et outils à disposition des participants**

Les supports de cours se composaient de diaporamas permettant de définir les notions de base et d'explicitier les enjeux ainsi que la démarche pour exploiter les données. Des exemples de datavisualisations ou de cartes ont été présentés.

### **Documentation des activités**

Différents supports ont également été fournis les trois années dans le but d'aider les étudiants à porter un regard réflexif sur leurs pratiques et de partager leurs difficultés et stratégies adoptées pour y faire face.

**Année 1** : un [tableur](#) proposant des étapes à documenter a été proposé. Un étudiant par groupe avait pour rôle de compléter ce tableur en parallèle du travail du groupe

**Année 2** : [un livret du participant](#), carnet de bord reprenant à la fois des éléments de cours et des questions structurant la démarche a été proposé. Un étudiant par groupe avait à nouveau pour tâche de compléter ce carnet.

**Année 3** : un carnet multimédia de documentation de l'activité ([DoDoc](#), application développée par l'atelier des chercheurs) a été proposé aux étudiants pour rendre compte de leur travail avec les données tout au long de l'année.

Ces carnets de bord ont également été collectés pour le projet de recherche dans le but d'analyser les trajectoires et stratégies des étudiants et d'observer la manière dont chacun des groupes a vécu cette activité, en explicitant ou non, les difficultés rencontrées.

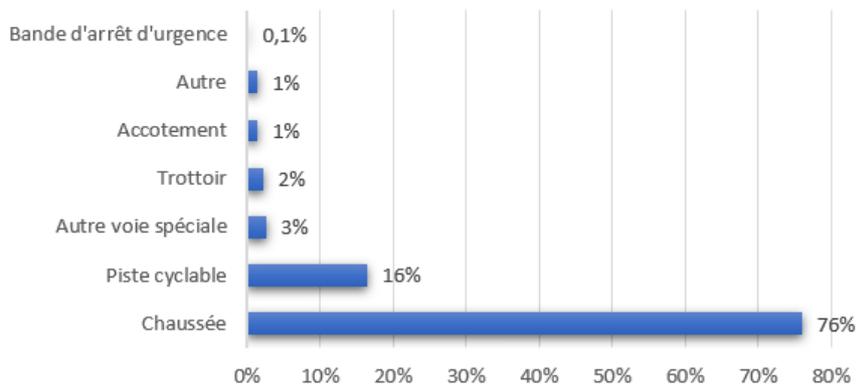
## Les productions des participants

Les productions présentées ici permettent d'illustrer le travail réalisé dans certains groupes d'étudiants. Ils donnent à voir des compétences plus ou moins développées, en matière de visualisation de données, et de cartographie dans une base de données collaborative telle qu'Open Street Map.

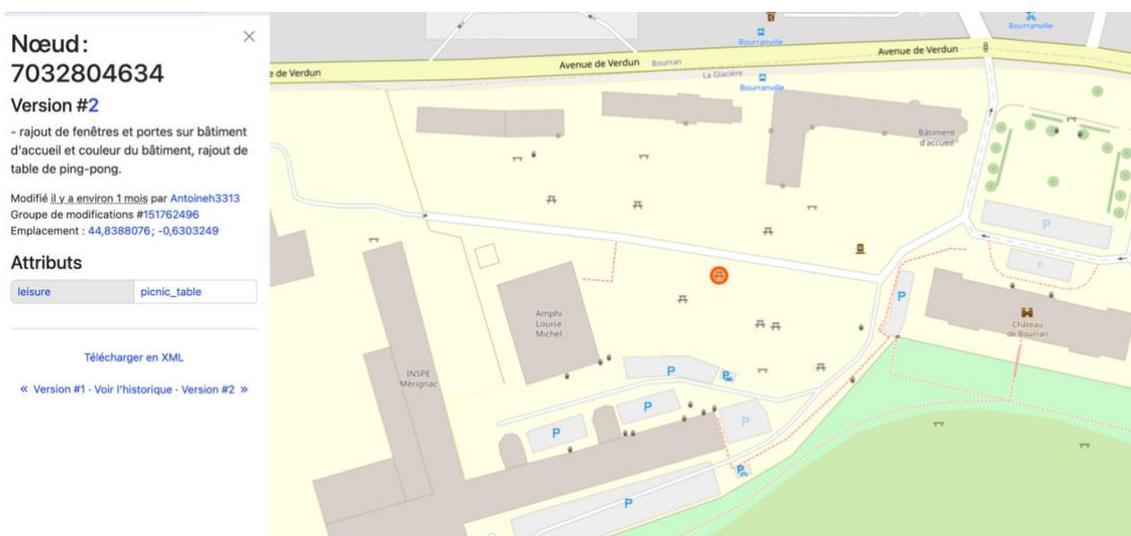


**Année 1** : Exemple de production d'étudiants réalisée lors du datasprint « La place des femmes dans les médias »

## Lieu d'accident des cyclistes en France en 2021.



Année 2 : Exemple de production d'étudiants réalisée lors du datasprint « Traces 2 roues »



Année 3 : Exemple de productions d'étudiants réalisées dans Open Street Map lors de la cartoparty : ajouts sur la carte du site de l'INSPE de Mérignac





**Illustration : travail en groupe lors du Datasprint année 2**



**Illustration : travail en groupe lors de la cartoparty année 3**

Certains dispositifs de collecte se sont révélés particulièrement contraignants et peu fructueux :

Formation aux données en Master MEEF Documentation

- La grille d'observation : trop complexe à renseigner en temps réel de manière exhaustive
- La captation audiovisuelle : produit des fichiers lourds et conséquents à analyser, mais peut permettre de revenir sur un moment marquant du travail de groupe
- Le carnet de bord papier, numérique et multimédia : difficile voire impossible pour les participants engagés dans l'activité, d'en sortir pour rendre compte de chaque étape du travail de groupe.

En revanche, l'enregistrement d'audios par les étudiants eux-mêmes à partir d'une fiche type « persona » les invitant à produire un autoportrait sur leur rapport aux données et à l'activité de formation avec les données, a été extrêmement riche. Ces autoportraits permettent de mettre en évidence des représentations et rapports personnels aux données et donnent à voir, pour l'étudiant lui-même, ainsi que pour les chercheurs, la manière dont des compétences ont pu être développées dans le cadre de la formation.

## Itinéraires d'apprentissage sur et avec les données

### Les obstacles et difficultés rencontrées

Les principales difficultés rencontrées par les étudiants se situent à plusieurs niveaux :

- **Difficultés sociotechniques** : liées aux habitudes des étudiants avec les équipements et logiciels propriétaires qui diffèrent des usages attendus avec les formats ouverts et libres de droits.
- **Difficultés dans l'interprétation des jeux de données ouverts** : liées à la masse de données et aux codifications particulières utilisées dans certains tableurs, donnant le sentiment que les données n'ont pas été produites par des humains mais par des machines comme l'illustre le propos de cette étudiante : « *Les données seraient peut-être plus propres si elles repassaient par une manipulation humaine* ».
- **Difficultés dans le passage des données à l'information** : croisant compétences en design de l'information et compétences statistiques, le processus de construction de l'information à partir des données reste fragile, incertain, « bricolé ». Les étudiants passent par de nombreuses phases d'essais-erreurs jusqu'à parvenir à une représentation qui leur paraisse intelligible et représentative des données, comme l'illustre le propos de cette étudiante : « *Nous n'avons pas arrêté de batailler et de trafiquer les chiffres pour produire une datavisualisation* ».

### Les leviers et stratégies

Le principal levier observé est l'appui sur un « expert » du traitement de données dans chaque groupe. La collaboration, ayant « pour objectif de mutualiser les connaissances de chaque membre du groupe en vue d'une réalisation commune » (Reverdy, 2016) permet ici le

croisement des compétences requises dans les différents domaines (statistiques, informatique, design de l'information...).

La documentation de l'activité, contrairement à ce que nous pensions, n'est pas perçue par les étudiants comme une ressource. Celle-ci reposant sur un processus intellectuel individuel, elle semble en effet peu compatible avec le travail collectif, réalisé sur un temps limité.

Les supports de cours sont également peu mobilisés dans l'activité. Les étudiants procèdent davantage par essais-erreurs, et se réfèrent peu à la documentation fournie. Le tâtonnement et le bricolage sont très souvent observés permettant aux étudiants d'élaborer leurs propres stratégies pour parvenir à l'objectif final.

# 3 /// ANALYSE CRITIQUE DU DEVELOPPEMENT D'UNE LITTERATIE DES DONNEES EN FORMATION

La littératie des données peut être définie comme « la capacité d'accéder, d'interpréter, d'évaluer, de gérer, de manipuler et d'utiliser de manière critique et éthique les données (Calzada-Prado et Marzal, 2013) grâce à une connaissance des caractéristiques de la donnée, de son cycle de vie et des différents impacts engendrés par son usage, notamment en termes de sécurité et de protection de la vie privée (Crusoe, 2016). » (Verdi, 2023). Cette notion fait cependant abstraction de l'« ensemble de représentations, (d'un univers de sens, une capacité à se projeter, un sentiment d'appartenance » qui caractérise la culture des données (Lehmans, 2017). La prise en compte de ces représentations et imaginaires apparaissent essentiels en formation pour développer un sentiment d'agir compétent (Bihl, 2020) avec les données, c'est-à-dire pour comprendre les conditions, les moyens et les effets de la production et de la circulation de l'information, et pouvoir agir en citoyen autonome. Cette définition renvoie à la culture de l'information allant de l'utilisation de l'information à l'agir informationnel : compréhension des traitements de l'information et des modes d'accès proposés, utilisation raisonnée de l'information, intégration dans une culture générale et éthique et dans les usages sociaux (Simmonot, 2014). C'est dans cette perspective du développement d'une culture des données allant de l'utilisation à l'agir que nous interrogeons ces pistes de formation des futurs enseignants aux données.

## Techno-imaginaires liés aux données et manières de les déconstruire en formation

Les techno-imaginaires renvoient aux « représentations positives ou négatives que l'on se fait des technologies, pouvant être issues de mythes, croyances, connaissances préalables qu'ont les acteurs. Ces représentations sont perceptibles en creux dans la pratique car elles orientent et guident l'action » (Plantard, 2016). Nos observations ont permis de montrer que les étudiants se représentent les données comme des relevés produits par des machines et non par des humains. Ainsi, le processus de collecte qui est pourtant issu d'intentions d'acteurs humains intervenants sur un domaine précis et opérant des choix pour indexer et organiser les données qu'ils collectent sont invisibles dans les représentations et manières d'agir des étudiants. Il apparaît que ces conceptions, qui prédominent dans le processus d'exploitation d'un jeu de données ouvertes (années 1 et 2 de notre expérimentation), sont moins présents dès lors que les étudiants sont engagés dans le processus de collecte des données et qu'ils ont également pour mission d'indexer ces données, en se référant à des propositions établies, comme cela a été le cas pour l'année 3 de notre expérimentation avec la base de données contributive Open Street Map. Un moyen de déconstruire les imaginaires liés aux données est alors de proposer dans la formation un travail consistant à partir d'une collecte propre de

données et de participer aux tâches d'indexation de ces données afin de les mettre en commun dans une base de données ouvertes.

## Développement d'une réflexion critique en Éducation aux médias et à l'information (EMI)

Le travail à partir de données permet de nourrir une réflexion critique qui se situe dans le champ de l'EMI. Celle-ci repose sur la prise de conscience que la production de statistiques et d'infographies sont en réalité construites à partir d'une sélection réductrice de données intégrant elles-mêmes de potentiels biais dans la manière dont elles sont construites. Ces biais sont susceptibles d'échapper aux auteurs de ces informations, et donc d'avoir des effets sur la fiabilité de l'information partagée. Ce processus de construction de l'information en formation permet de prendre du recul sur les discours considérant les données comme élément de preuve et de scientificité. Ce travail permet aussi de mettre en exergue les enjeux des discours promoteurs du pilotage par la donnée (*Data driven strategy*), dans lesquels l'organisation et la finalité de l'activité sont redessinées à partir de la donnée, notion qui devient un descripteur universel des objets (Pène, 2020). Or, il faut rappeler qu'une donnée n'est pas donnée, mais en réalité captée (*capta*, Becker, 1952). Les données sont donc par nature subjectives de même qu'elles suggèrent des interprétations susceptibles d'occulter la diversité et la complexité des pratiques d'acteurs et de situations singulières. En cela, une éducation critique aux médias et à l'information, consistant à analyser la manière dont sont collectées, interprétées et exploitées des données en société, nous paraît essentielle.

# RÉFÉRENCES

Becker, H. (1952). Science, culture, and society. *Philosophy of Science*, 19(4), 273-287.

Bihl, J. (2020). La littératie médiatique par et pour l'empowerment: La place du pouvoir d'agir dans un dispositif évaluatif de compétences. *Spirale-Revue de recherches en éducation*, 66(3), 65-75.

Bodin, F. (2018). Apprendre avec les données numériques, les datasprints pédagogiques : Le cas de Traces de Soldats. *Medium*. 17 octobre 2018. Consulté sur : <https://medium.com/digital-praxis/apprendre-avec-les-donn%C3%A9es-num%C3%A9riques-avec-les-datasprints-p%C3%A9dagogiques-le-cas-de-traces-de-50475fbc6b8e>

L'Her, G., Servières, M. et Siret, D. (2018). "La Cartopartie, une nouvelle forme de balade urbaine déployée par les villes", *Les Cahiers de la recherche architecturale urbaine et paysagère*, 3. URL: <http://journals.openedition.org/craup/1003>; DOI: <https://doi.org/10.4000/craup.1003>

Pène, S. (2020). Métiers de la fonction publique : motifs et modèles de transition numérique. *Approches Théoriques en Information-Communication (ATIC)*, 1, 58-80. <https://doi-org.docelec.u-bordeaux.fr/10.3917/atic.001.0058>

Reverdy, C. (2016). La coopération entre élèves : des recherches aux pratiques. *Dossier de veille de l'IFÉ*, 2016, 114, pp.1-32.

Simonnot, B (2014). Médiations et agir informationnels à l'ère des technologies numériques. *Les Cahiers d'Esquisse, Vers de nouvelles formes de médiation documentaire et bibliothéconomique*, 4, pp.21-33. (sic\_01124651)

Verdi, U. (2023). "Quelle(s) réponse(s) à l'enjeu d'acculturation aux données ? Un état de l'art des caractéristiques de la data literacy", *Revue française des sciences de l'information et de la communication*, 26. URL: [http://journals.openedition.org.docelec.u-bordeaux.fr/rfsic/14589](http://journals.openedition.org/docelec.u-bordeaux.fr/rfsic/14589)

**Ce document est rédigé par les équipes de recherche  
dans le cadre des GTnum du ministère de l'Éducation nationale.**

**La responsabilité des contenus publiés leur appartient.**