

# La Pertinence du Catalogue du Compte Personnel de Formation, vue au Prisme du Comportement des Utilisateurs

**Résumé.** En France, le Compte Personnel de Formation (CPF) sert de véhicule aux politiques publiques visant à financer la formation d'adultes. La plateforme Mon Compte Formation (MCF) permet de sélectionner la formation de son choix au sein d'un catalogue particulièrement fourni; elle met à disposition un certain nombre de données sur son offre, données qui à ce jour restent sous-exploitées. Nous présentons ici une réflexion méthodologique sur les opportunités offertes par ces traces. Elle vise à apporter un regard sociologique et géographique sur des analyses de traces réalisées par des méthodes issues de l'informatique, pour mieux contraster l'offre des catalogues de formation et les thématiques qui constituent les principaux centres d'intérêt du public de la plateforme, ceux-ci étant inférés sur la base des traces laissées via le moteur de recherche intégré au sein de MCF.

**Mots-clés :** CPF, fouille de règles d'association, formation d'adultes, traces, organisme de formation

**Abstract.** In France, the Personal Training Account (CPF in French) serves as the main tool for public policies aimed at financing adult training. The Mon Compte Formation (MCF) platform allows you to select the training of your choice from a particularly comprehensive catalog; it makes available a certain amount of data on its offer, data which to date remains under-exploited. In particular, we present analyzes mobilizing geographical approaches, that aim to highlight the interest in methods from computer science to better understand catalog construction strategies, as well as the possible inadequacy of the catalogs proposed in relation to the expectations of the public, expectations that can be analyzed thanks to the logs left in research engines used within MCF.

**Keywords:** CPF, association rule mining, lifelong learning, analytics, training establishment

## Introduction

### Contexte de l'étude et questions de recherche

En matière de formation d'adultes, le Compte Personnel de Formation (CPF) représente en France un instrument central des politiques publiques [18]. Rattaché à un individu plutôt qu'à son contrat de travail, il vise à offrir des droits continus à la formation

tout au long de la vie. Il assure ainsi un accès ininterrompu à des droits [7], et permet aux salariés une grande liberté dans le choix de leurs formations [16,19]. Ainsi, il n'est pas nécessaire de choisir une formation en fonction de la profession exercée, ni de demander l'approbation de l'employeur. La plateforme qui héberge le catalogue de formations se nomme Mon Compte Formation (MCF) depuis 2019 ; elle est gérée par la Caisse des Dépôts et Consignations (CDC).

Le lancement de la plateforme MCF constitue l'un des nombreux avatars de ce qu'un auteur comme Bullich [9, 10] nomme la « plateformisation » de la formation. Dimension centrale de cette plateformisation, des milliers d'organismes de formation (OF), en concurrence les uns avec les autres, proposent sur MCF une offre hétéroclite, allant de l'informatique au permis de conduire en passant par toutes sortes de titres et certifications. Les chercheurs d'emploi constituent un contingent important de la plateforme, mais des publics variés sont représentés [5]. La CDC s'intéresse de près aux retours sur investissement associés à la mise en place du CPF, en contribuant à des études sur les choix des formations achetées selon les profils d'apprenants, mais aussi en analysant les traces issues de la navigation sur MCF. Le premier type d'analyse fait l'objet de publications ouvertes au grand public [6,13,15,17]. Le second relève en revanche de réflexions menées uniquement en interne, conséquence des contraintes, politiques ou légales, associées à la diffusion de travaux mobilisant des traces [4].

Pourtant, ces dernières présentent un intérêt certain tant du point de vue des visées de la recherche que de celles, plus pragmatiques, des organismes de formation souhaitant disposer d'informations pertinentes pour développer leur catalogue de cours. Comprendre les déterminants de ce phénomène est crucial en termes d'efficacité de l'action politique, qui peut orienter les financements. Des entreprises comme Wedof<sup>1</sup> proposent toutes sortes de services aux OF, mais faute d'accès aux traces, détenues par la CDC, ils ne peuvent se situer au niveau de granularité que nous proposons ici. Notre démarche pourrait donc aussi bien informer des politiques publiques que servir de base à la conception de services à destination des organismes de formation.

Parmi les nombreuses opportunités qu'offre l'exploitation des traces, nous nous proposons dans cet article de discuter de la question de l'adéquation entre offre proposée par les OF et centres d'intérêt des utilisateurs de MCF, ceux-ci étant inférés sur la base de leur comportement de navigation sur la plateforme. Le choix du terme à employer a été l'objet de discussions : si la visée initiale du projet était de contraster offre et attentes, nous ne pouvons pas inférer quant aux attentes sur la base du seul comportement de navigation. Le fait est que les attentes d'un individu peuvent difficilement être capturées par de simples comportements ; une enquête s'impose. Le terme « centre d'intérêt » est plus neutre qu'attente ou demande, puisqu'il est plus légitime de supposer qu'un apprenant porte un certain intérêt à une thématique ou une formation dont il lit le descriptif ou qu'il entre dans un moteur de recherche.

En matière d'évaluation des politiques publiques, les enjeux d'études sur la pertinence des catalogues de formation sont considérables. Chaque année, l'Etat, ses opérateurs, les OPCO (opérateurs de compétences), et les entreprises investissent plus de trois milliards d'euros dans le CPF. Par ailleurs, cette somme ne prend pas en compte

---

<sup>1</sup> <https://www.wedof.fr/>

le fait qu'une partie significative des apprenants financent par eux-mêmes une part des frais d'inscriptions, à plus forte raison depuis le vote de l'amendement n°698 de l'article 49 du Code du Travail, en 2023 [11]. La bonne utilisation de ces sommes dépend largement de la pertinence du catalogue pour l'audience de la plateforme. Cette considération nous a amené à structurer le présent article autour de questions de recherche qui fait de cette contribution une réflexion de nature avant tout méthodologique structurée autour de deux questions :

Comment les données disponibles tant sur le catalogue de formation de MCF que sur le comportement de navigation au sein de la plateforme, et du moteur de recherche afférent, peuvent-elles être combinées pour évaluer l'adéquation de l'offre avec les thématiques de formation suscitant l'intérêt des apprenants ? Dans quelle mesure les données à disposition de la CDC permettent-elles d'analyser sous l'angle géographique et sociologique une telle adéquation ?

Aborder ces questions implique en premier lieu de construire des indicateurs sur la base des traces résultant de recherches, ou de la navigation au sein de MCF. Cette contribution mêle des résultats obtenus sur la version de juin 2023 du catalogue de MCF, pour ce qui est de l'offre, et de données simulées, pour ce qui est des traces. En effet, si des analyses de traces réelles ont été produites par nos étudiants pour le compte la CDC dans le cadre de leurs stages, les résultats correspondants ne peuvent être publiés dans cette contribution, du fait des contraintes politiques et légales [4]. Néanmoins, dans la mesure où il s'agit d'une preuve de concept, la substitution de dernière minute par des données artificielles ne change rien à la visée du propos.

### **De l'approche géographique des traces à la notion de visite**

Nous avons mené la présente étude en premier lieu sous l'angle des thématiques recherchées par l'utilisateur, que l'on peut ensuite contraster avec les thématiques proposées dans les catalogues, mais aussi sous l'angle des autres caractéristiques des formations en question, comme le niveau d'études correspondant. Le croisement entre d'une part des données sociologiques et économiques récoltées à l'échelon national et à une granularité fine – celle de la commune, et d'autre part des traces issues de plateformes, a été réalisé dans plusieurs études américaines [12, 20], mais n'est pas une pratique commune dans les recherches en France.

S'agissant plus précisément de l'interaction de cohortes d'apprenants avec une offre de cours, nous avons déjà illustré sur la base d'analyses de traces d'une plateforme de MOOC comment des centaines de milliers d'apprenants voyaient leur comportement de sélection de formation influencé par les caractéristiques d'un catalogue [2,3].

Cette recherche a employé plusieurs concepts, notamment celui de "clé d'entrée", qui désigne le processus ayant mené à la sélection d'un produit, correspondant ici à l'inscription à un cours. Ce concept se divise en deux catégories : la "clé d'entrée plateforme" et la "clé d'entrée produit". D'une part, un internaute peut visiter un site en ayant une idée générale de ce qu'il veut acheter, et d'autre part, il peut vouloir acquérir un ou plusieurs produits sans connaître précisément les sites où les trouver. Dans le premier

cas, la clé d'entrée est le site marchand à consulter, tandis que dans le second, il s'agit du produit spécifique à acheter [13].

Dans le contexte de la « visite expérientielle », les motivations sont principalement hédonistes, incluant la découverte d'offres originales, le plaisir de la surprise, et la promenade à travers un vaste assortiment de produits, évoquant la logique du lèche-vitrine. Des visites régulières sur un site de e-commerce sont des exemples de ce type d'achat expérientiel [13, p. 213]. Cela s'applique également aux participants qui, après avoir découvert MCF, y retournent fréquemment pour voir si de nouveaux cours susceptibles de les intéresser ont été ajoutés, même s'ils n'ont pas d'idée précise sur le thème de formation qu'ils recherchent. Dans le cas de la « visite expéditive », « l'internaute se rend sur un site pour réaliser un achat assez précis. Soit il connaît l'offre de ce site pour ce produit (il a vu une offre sur une publicité ou un ami lui en a parlé, par exemple) ; soit il sait [...] que ce site est le meilleur pour telle catégorie de produit. [...] À la différence du précédent, ce processus de décision est guidé par un objectif précis : les motivations de visite sont plutôt utilitaires » [13, p. 214].

## **Méthodologie**

La réflexion sur le contraste entre offre et demande a été menée en parallèle d'une part au sein du laboratoire AGORA, et d'autre part par des élèves-ingénieurs de CY Tech effectuant dans le cadre de leur stage de fin d'études des missions d'analyse de traces pour le compte de la CDC. Les informations sur les traces rapportées ici sont issues de leur mémoire de fin d'étude. Ne pouvant rapporter les graphiques produits dans ce cadre, contraintes légales obligent, nous avons présenté les données collectées quotidiennement via MCF, puis généré des données artificielles pour illustrer notre propos. Nous revenons ensuite sur l'approche suivie en parallèle pour étudier les catalogues de formation, dont les données sont mises à disposition du grand public.

### **Des traces de navigation collectées notamment par Matomo**

La CDC dispose d'une table contenant spécifiquement les logs sur les requêtes effectuées dans le moteur de recherche interne à MCF. En plus de l'horodatage, ils incluent la thématique recherchée (enregistrée via son formacode), la zone géographique (pour les formations non dispensées à distance), d'éventuels filtres mobilisés : certains sont obligatoires à chaque recherche, comme la modalité (face-à-face/présentiel), d'autres non. L'on peut par exemple effectuer des recherches par organisme de formation.

Les indicateurs relatifs aux recherches sont utiles pour inférer quant aux centres d'intérêt du public de la plateforme, mais ils ne représentent qu'un aspect de la question. L'on pourrait par exemple afficher en complément, par zone géographique, le nombre de dossiers refusés faute de place dans les formations. L'on pourrait ainsi mieux identifier les thématiques pour lesquelles la demande dépasse l'offre, mais l'on sortirait ce faisant du champ des learning analytics, et donc des EIAH. Or l'objectif de cette

contribution est justement de fournir une illustration supplémentaire de l'intérêt qu'il y a à se placer à l'intersection entre sciences humaines et sociales et learning analytics.

En plus des traces produites par le moteur de recherche, la CDC dispose de tables comprenant des données détaillées sur les événements produits par les interactions avec le site : visiter une page, cliquer sur un lien externe ou sur un téléchargement, recherche sur le site. Si toutes les actions sont stockées dans les serveurs de la CDC, le trafic est en partie suivi grâce à *Matomo*, outil grand public régulièrement mobilisé pour analyser les utilisations d'un site Internet, mais que nous ne mobilisons pas ici. Il permet une sélection des événements d'intérêt fournit une interface d'interaction ainsi qu'une API dont la fonction est d'extraire des traces précises par requêtage. Dans certaines conditions, il est possible de relier le *user\_id* stocké par Matomo avec les données dites du *Dossier*, qui sont des informations autodéclarées fournies par l'apprenant : dernier diplôme obtenu, âge, genre, statut socio-professionnelle. Le numéro de sécurité sociale, identifiant utilisé dans MCF, est exclu de ces données.

Nous avons produit, pour mimer les requêtes dans le moteur de recherche, des traces artificielles ayant strictement la même composition que les données réelles permet de contourner les obstacles légaux liés à la diffusion par la CDC de traces réelles, sans pour autant devoir renoncer à illustrer les concepts sous-tendant les indicateurs. Des étudiants formés par nos soins ont fourni dans leurs rapports, suite à un stage portant sur l'analyse des données brutes de la CDC, des illustrations de la structure des bases de données en jeu, avec des exemples permettant de montrer une trace typique, pour une lecture de descriptif de formation, ou pour une requête. Ces éléments sont suffisants pour produire des bases identiques, sur le plan des variables en jeu, à celles manipulées par les analystes de la Caisse des Dépôts.

Les cartes que nous produisons dans la section dédiée aux résultats ne semblent pas constituer un risque majeur d'exposition de données à caractère privé, et nous souhaitons dès lors nous attarder ici sur les raisons qui nous ont poussé à produire des données artificielles. En effet, il faut pour les produire passer par des étapes où l'on est capable d'identifier les termes entrés par un individu donné dans le moteur de recherche de MCF ; se posent donc des questions de données à caractère personnel, et de conventions lourdes à mettre en place.

La CDC est soumise au régime du RGPD, comme, au demeurant, tout opérateur de l'Etat. Si elle est légitimement en droit de mener des analyses sur les traces que MCF produit, elle ne peut pas transmettre ces données à des laboratoires de recherche sans la signature de conventions appropriées. De fait, ce n'est à notre connaissance qu'en 2023 que les conditions générales d'utilisation (CGU) de MCF ont inclus la possibilité d'utiliser les données selon une logique de recherche, et la signature des CGU mises à jour a donc constitué un préalable pour qu'un individu puisse être inclus dans une étude. Le travail de développement d'une relation professionnelle de confiance entre acteurs de la recherche et opérateurs de l'Etat, le processus de conventionnement, généralement long, l'exclusion, hors des bases manipulées, des individus n'ayant pas signé les CGU mises à jour : toutes ces étapes, souvent laborieuses, n'ont de sens que si l'étude porte sur des phénomènes capturés dans les traces. Elles sont impératives si la focale, par exemple, porte sur l'évolution de la nature des thématiques les plus plébiscitées dans tel ou tel département français. Dans la mesure où nous ne cherchons pas à nous inscrire

dans la mise en évidence d'un phénomène, mais plutôt à apporter un regard méthodologique sur la production d'indicateurs d'une part, et à mettre en lumière les opportunités offertes par les bases de MCF d'autre part, il ne nous a pas semblé pertinent d'investir à ce stade dans la mise en place d'une collaboration formelle avec l'opérateur de l'Etat.

### **Analyse de l'offre de formation**

La CDC met à disposition un jeu de données nommé *offre de formation*, mis à jour régulièrement sur son site<sup>2</sup>. Les analyses présentées ici correspondent à la version du 1<sup>er</sup> juin 2023. Nous avons analysé en premier lieu la base sous l'angle de la géographie, l'échelle retenue étant celle du département. En effet, à moins que l'utilisateur ne cherche à se former à distance, il recherche a priori le cours, en priorité dans son environnement immédiat, c'est-à-dire sans son département, ou éventuellement dans un département limitrophe à son lieu de résidence. Précisons que l'on ne peut connaître l'adresse de résidence d'un utilisateur connecté à la plateforme MCF que s'il s'est connecté – il est en effet possible de naviguer sans se connecter. C'est donc à la richesse de l'offre dans l'environnement géographique proche qu'il est intéressant de contraster le résultat des recherches effectuées – à moins que l'on se penche sur les formations à distance.

S'agissant de la variable retenue pour représenter les centres d'intérêt, nous avons priorisé la thématique et le niveau de sortie, équivalent à un niveau d'études obtenu à la suite de la formation. La thématique peut être vue à plusieurs échelles, celle du *Formacode* (plusieurs milliers), celle du *Champ sémantique* (aussi appelé *NSF*, au nombre de 65, comme *Pêche, aquaculture*), et celle du *Grand domaine* (au nombre de 14, comme *Agriculture, Environnement*). Nous avons par exemple travaillé avec la version du printemps 2023 de la taxonomie, disponible via le Centre Inffo<sup>3</sup>. Elle comprend un identifiant numérique, le libellé (Boulangerie, etc.), le grand domaine correspondant, sa place dans l'arborescence des formacodes.

---

<sup>2</sup> <https://opendata.caissedesdepots.fr/>

<sup>3</sup> <https://formacode.centre-inffo.fr/>

### Synthèse des indicateurs présentés

Sur la base des données relatives au moteur de recherche interne, de consultations de descriptifs et d'analyses de catalogue, de nombreux indicateurs peuvent être construits. Nous en proposons un récapitulatif dans le Tableau 1.

Indicateur	Mode de construction et limites
% de recherches positives	Analyse de la nature des requêtes issues de l'utilisation du moteur de recherche de MCF : mots-clés sélectionnés ou entrés dans le champ de texte libre. Si le moteur présente dans sa réponse au moins un élément correspondant effectivement à ces mots-clés, la recherche est alors considérée comme « positive ».
% d'une thématique donnée dans un catalogue de formation	100 * le nombre des formations correspondant à une thématique donnée au sein d'une aire géographique divisé par le nombre de formations proposées dans cette aire. Mobilisation de la base en open data (données réelles) nommée <i>offre de formation</i> .
Ratio demande / offre	% de la thématique dans les requêtes, divisé par le % de la thématique, en nombre de formation, au sein de l'offre locale

### Résultats

Nous présentons en premier lieu des analyses mêlant traces artificielles et analyse du catalogue réel pour contraster offre et demande, pour discuter une série de métriques jugées pertinentes pour mesurer l'adéquation entre les deux.

#### Analyses des données du moteur de recherche

Comme les recherches réalisées dans MCF requièrent systématiquement d'entrer un code postal lorsque l'on recherche une formation en présentiel – en précisant une distance par rapport à la commune recherchée, il est possible de réaliser des cartes à l'échelle de la commune. Ainsi, plusieurs indicateurs peuvent être construits et représentés sous la forme de cartes. Nous représentons ci-dessous des données simulées analogues aux données réelles, dans lesquels nous représentons un indicateur d'adéquation, en réalisant l'analyse avec d'une part avec tous les utilisateurs de MCF (Figure 1, Haut), et d'autre part ceux ayant déclaré n'avoir que le Bac (Figure 1, Bas). Dans ce dernier cas, l'objectif est d'introduire une dimension sociologique en plus de l'aspect géographique.

En Fig. 1, nous avons représenté le taux de recherches positives, c'est-à-dire le pourcentage des requêtes ayant retourné au moins un résultat pour la thématique recherchée, pour le niveau de distance à la commune sélectionné (qui peut être choisi par

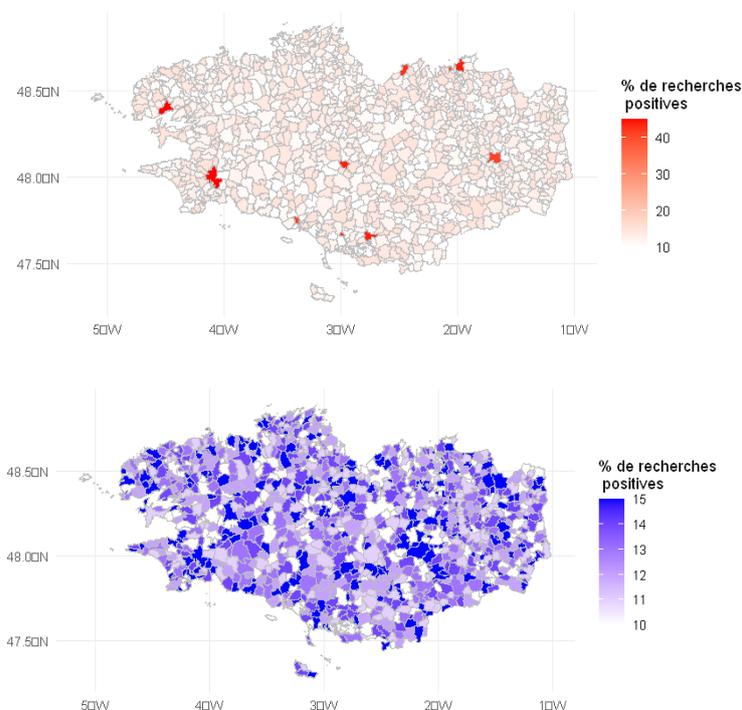
l'utilisateur entre 5, 10, 15, 50 ou 100 km par exemple), par contraste avec les recherches infructueuses. Il nous faut faire le distinguo ici entre une recherche, qui nous donne des indications sur une intention, et une consultation, qui implique qu'un résultat a été trouvé, et que le descriptif de la formation a été visionné.

Nous avons retenu ici un rayon de 5 km autour de la commune sélectionnée, tous grands domaines confondus ; ce seuil peut être modifié de manière arbitraire, selon que l'on s'intéresse aux formations situées dans l'environnement immédiat de l'apprenant, ou s'il lui est possible de se déplacer jusqu'à une certaine distance. On remarque pour ce jeu artificiel, dans le premier cas (Gauche), des valeurs plus élevées pour des centres urbains tels que Quimper, ou Rennes, signe que les habitants de ces communes trouvent davantage ce qu'ils cherchent via le moteur de MCF. Il suggère alors une meilleure adéquation de l'offre aux apprenants vivant en communes urbaines par rapport aux communes rurales, peut alors être utilisé dans la conception de politiques publiques visant à atténuer les inégalités d'accès à la formation entre la ville et la campagne, notamment via l'encouragement à la production de formations à distance. L'on peut également aborder la question sous un angle sociologique, en contrastant par exemple le pourcentage de recherches positives selon que le plus haut diplôme obtenu soit le baccalauréat, le master ou le doctorat.

En filtrant par le niveau de formation des apprenants (Figure 1, Bas), accessible par le dossier, l'on constate que les personnes peu diplômées sont logées à la même enseigne que les autres dans leurs recherches, sans distinction majeure entre les grandes villes et les communes rurales. L'interprétation de tels résultats nécessite ensuite un travail statistique, fondé sur des régressions, visant à identifier les variables sociodémographiques et socioprofessionnelles associées susceptibles d'expliquer des inégalités géographiques.

Comme nous l'avons souligné dans la section consacré à la méthodologie, il convient de compléter les analyses de traces du moteur de MCF par des analyses de consultations de descriptifs. En effet, un certain nombre d'apprenants découvrent ces descriptifs par des recherches sur des moteurs classiques comme Google, et non par le moteur interne à MCF. Par contraste avec le cas précédent, l'on ne peut ici parler de recherche infructueuse, puisque par définition une formation a été trouvée et son descriptif consulté.

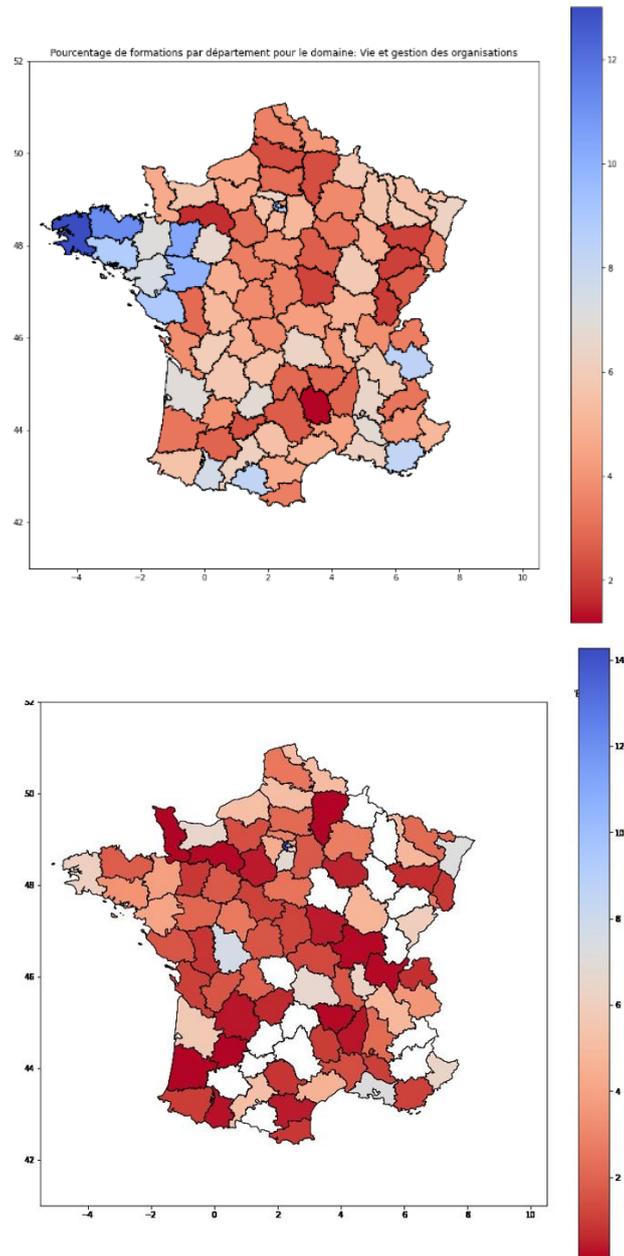
Aborder la question sous un angle géographique est moins intuitif, puisque la formation est certes associée à une adresse, mais il n'est pas obligatoire pour l'apprenant de spécifier sa propre adresse pour atteindre et visionner le descriptif. Une recherche sur Google ne permettant pas de prendre en considération des critères de distance au domicile, elle conduit a priori à découvrir fréquemment des formations organisées trop loin du lieu de résidence de l'apprenant, ce qui vient interférer avec une analyse sur la bonne couverture géographique des besoins de formation.



**Fig. 1.** Carte représentant, pour les communes de Bretagne, le pourcentage de recherches positives, tous utilisateurs confondus (Haut), et utilisateurs dont le diplôme maximal obtenu est le bac (Bac). Données simulées.

### Contraster offre et demande : un regard géographique et sociologique

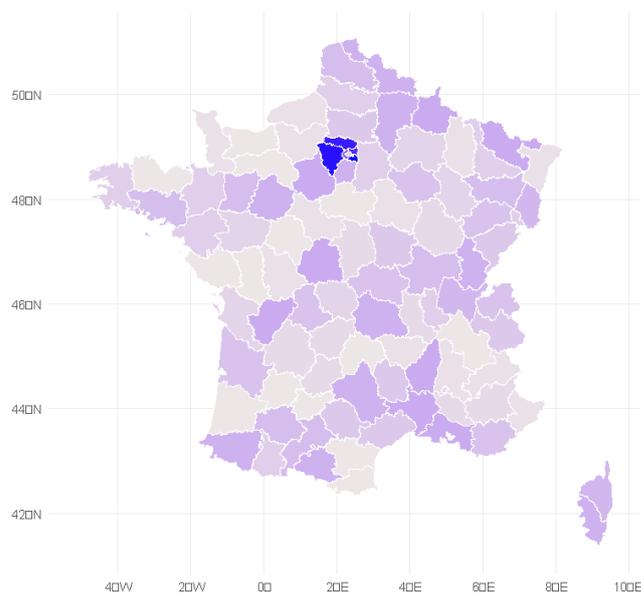
Nous revenons maintenant sur les indicateurs qui ne peuvent être construits seulement sur la base de traces, mais qui impliquent une mise en regard avec l'offre. Sur le plan géographique, il n'est pas pertinent de représenter à l'échelle de la commune, car il faut un nombre suffisamment important d'OF par unité géographique pour produire des statistiques pertinentes, ce qui invite à ne pas descendre en dessous du niveau départemental. Il s'agit, en se concentrant sur les formations dispensées présentiel, de calculer le ratio entre le pourcentage des apprenants effectuant une recherche dans une thématique donnée (niveau formacode, NSF ou Grand domaine), et le pourcentage des formations de l'unité géographique (du département par exemple) correspondant à cette thématique. S'il faudrait proposer des indicateurs plus fins prenant en compte le nombre de places de formation ouvertes, et non simplement le nombre de formations disponibles, l'on peut néanmoins voir que plus cet indicateur élevé, plus la demande est importante par rapport à l'offre, et inversement.



**Fig. 2.** Carte du pourcentage des formations du catalogue de MCF daté de juin 2023 relevant du domaine : Vie et gestion des Organisations (Haut) et pour les formations de niveau 7 (équivalent Master), les valeurs les plus élevées étant Paris (Bas). Données réelles.

En prenant en compte le catalogue de formation, et non plus seulement les traces laissées par les apprenants putatifs, nous gagnons en profondeur d'analyse. Sur la base des données fournies par la CDC, nous avons observé des disparités géographiques conséquentes eu égard à la part d'une thématique dans l'offre de formation. A des fins d'illustration, la Figure 2 (Haut) présente la place qu'occupe le grand domaine « Vie et gestion des organisations » dans le pourcentage de l'offre de formation d'un département ; cette thématique est par exemple surreprésentée en Bretagne. Les OF se positionnent largement sur ce sujet, sans que l'on ne sache précisément si la démarche a son équivalent dans les demandes.

Dans la Figure 2 (Bas), nous abordons la question de l'offre sous l'angle du niveau de sortie, représentant l'un des niveaux les plus élevés : le niveau 7, qui est un équivalent Master. De nombreux départements ne proposent qu'une offre limitée dans ce domaine, voire aucune offre – ils sont alors représentés en blanc. Seul Paris, avec 14% de l'offre, fait figure d'exception, sans doute car l'on y trouve un nombre important d'individus fortement diplômés, dont le taux de recours au CPF est également largement supérieur à la moyenne nationale [1], ce qui, a priori, stimule la création d'une offre distincte de celle du reste du territoire.



**Fig. 3.** Carte du ratio demande / offre, mesuré sur la base des requêtes dans le moteur de recherche interne, illustration pour la thématique Sports, Loisir, Tourisme. Données simulées.

Pour établir le lien entre offre et demande, nous proposons en Figure 3 une carte présentant un ratio entre le % de la thématique dans les requêtes, divisé par le % de la thématique, en nombre de formation, au sein de l'offre locale. En Île de-France, on constate pour le Grand domaine *Sports, Loisir, Tourisme* le rapport entre la demande, telle que perçue dans les requêtes du moteur de MCF, est nettement plus prononcée par rapport à l'offre existante que dans les autres régions de France. Une telle information pourrait être utile pour des OF qui envisageraient de développer ou d'enrichir leur offre dans ce secteur, en leur fournissant des indications quant aux dynamiques l'œuvre dans cette zone géographique.

## Discussion

Alors que les travaux de la Direction de l'Animation de la Recherche, des Études et des Statistiques (DARES) et de la CDC montrent que les analystes en charge de rendre compte des utilisations de MCF sont en capacité de traiter sous l'angle des SHS des données de consommation du CPF, et ce pour le compte de l'Etat, les ponts avec l'informatique restent à construire. Les lacunes en la matière sont manifestes dans la littérature scientifique. Comme nous l'avons souligné en introduction, malgré l'existence de réflexions méthodologiques sur les traces [19] et de cas d'études ancrés empiriquement [12], le nombre de publications sur le sujet reste faible. C'est la raison pour laquelle nous avons proposé ce travail sur la conception d'indicateurs permettant d'aller au-delà des analyses de données d'achat de formation faites par la DARES, qui ne couvrent qu'un seul des aspects du comportement d'achat et de navigation des utilisateurs de la plateforme.

En effet, ces achats peuvent être faits « faute de mieux », pour dépenser le solde disponible dans les OF présents dans l'environnement immédiat, ou sous l'influence d'organismes peu scrupuleux réalisant des campagnes d'appels téléphoniques pour faire la promotion de leur offre. En mobilisant l'informatique, et plus précisément l'analyse de traces laissées via les moteurs de recherche, l'on peut avoir un accès plus direct à la nature des formations recherchées par les apprenants, et identifier si le catalogue contient des éléments correspondant aux recherches effectuées. Néanmoins, tout indicateur présente ses limites et nous nous proposons de mobiliser les notions de visites présentées en introduction pour suggérer des alternatives [2], en mobilisant celles de visite expérientielle et expéditive. Précisons notre propos, qui porte sur l'indicateur présenté en Figure 1, et toutes ses variantes. Ceci tient au fait que l'indicateur présenté en Figure 3, contrastant offre et demande, ne nécessite pas de prendre en compte cette notion de visite, puisqu'il est construit sur la base de l'agrégation d'un nombre important de requêtes.

L'indicateur centré sur les recherches positives porte le flanc à une critique en simplification excessive en ce sens qu'il met toutes les requêtes sur le même plan, aussi bien les premières recherches effectués au moment de la découverte de MCF, que celles réalisées après s'être approprié la plateforme, après plusieurs achats, et de nombreuses visites. Or l'on peut émettre l'hypothèse selon laquelle les premières thématiques recherchées par l'utilisateur sont celles pour lesquelles l'appétence est la plus marquée.

Dès lors, un utilisateur qui ne trouverait pas satisfaction aux premières tentatives se rabattra sans doute, pour dépenser le solde de son compte sur les formations disponibles dans son environnement immédiat, quitte à choisir des thématiques qui ne seraient pas son premier choix.

Il s'agit alors de modéliser les requêtes selon une séquence, et leur attribuer un rang dans cette séquence. On distinguerait alors les recherches infructueuses réalisées en début de séquence de celles qui ont lieu en fin de séquence, longtemps après, par exemple, avoir réalisé un premier achat. Lorsqu'une recherche est infructueuse, l'on ne devrait en effet sans doute pas même sur un même plan des sujets recherchés en priorité, relevant par exemple de visites expéditives où l'utilisateur se concentre sur un sujet et un seul, et des thématiques recherchées lors de visites expérientielles, par exemple. Dans ce dernier cas, l'utilisateur effectue dans un temps limité des requêtes sur des sujets variés, par exemple pour se former à devenir boulanger, instructeur de yoga, professeur des écoles ; la signification à attribuer à la requête est alors sensiblement différente. Dans la mesure où il s'agit de l'équivalent du lèche-vitrine, si pour un sujet donné, l'utilisateur ne trouve pas satisfaction, il y a manifestement pour lui d'autres thématiques qui pourraient l'intéresser suffisamment pour qu'il investisse le solde de son compte CPF.

Terminons cette discussion par une réflexion sur l'application de tels indicateurs. L'intérêt est manifeste pour des décideurs officiant au sein d'entreprises privées (OF, etc.) et jouant un rôle dans la construction des catalogues de formation présents sur MCF. Identifier l'intérêt que porte l'audience de la plateforme pour telle ou telle thématique permet d'orienter une offre, à plus forte raison si l'on peut identifier les aires géographiques où l'offre est lacunaire par rapport à la demande, ce qui permet une meilleure granularité dans la prise de décision. S'agissant du pilotage des politiques publiques, l'application est moins évidente, car une grande agentivité est laissée aux acteurs de la formation continue dans la construction de leur offre. La puissance publique intervient au niveau des certifications, pour déterminer qui peut, via l'inscription au RS ou au RNCP, bénéficier des fonds du CPF. Le fait d'être considéré comme un métier en tension aide pour ce faire, et pour le moment, cette dénomination de « métier en tension » est davantage basée sur des enquêtes auprès des entreprises que sur d'autres considérations. Mais peut-être à terme peut-on envisager que la demande ou l'intérêt que présente l'audience de la plateforme pour telle ou telle thématique devienne un critère de décision ; les cartes et les indicateurs que nous avons produits pourraient alors prendre tout leur sens.

## **Conclusion**

Au-delà des obstacles à la diffusion des résultats obtenus suite aux analyses de traces, la principale limite de ce type de recherche réside dans le fait qu'un utilisateur peut réaliser des recherches sur MCF sans se connecter, et donc sans qu'un identifiant utilisateur ne soit associé. Dans ce cas de figure, l'on ne dispose pas de données sur ses caractéristiques sociodémographiques et socioprofessionnelles. Cela constitue un obstacle inhérent à la réalisation des analyses que nous avons proposées.

En guise de perspective à cette recherche, nous suggérons de mettre en œuvre ces analyses en mobilisant des données plus précises que celles fournies dans le *Dossier*, ce qui passe par exemple par la mobilisation de la base nommée *Déclaration Sociale Nominative* (DSN) [8]. Au regard de la sensibilité des informations qu'elle porte, des accords contraignants devront être signés avec les chercheurs intéressés par de telles approches. Constituée par l'Etat, la base rassemble notamment des données sur les professions occupées, ainsi que les revenus associés, tout au long de la vie d'un individu.

En croisant traces et données de la DSN, l'on pourrait par exemple caractériser les comportements de recherche de formation avant ou après une transition professionnelle. Ainsi, si une recherche suit un changement de poste, peut-être est-ce le symptôme d'une volonté de développer ses compétences dans un secteur nouvellement rejoint. Si une recherche a précédé une transition, indépendamment d'un achat effectif, il est plus pertinent de l'interpréter comme une volonté d'anticiper une transition. L'on pourrait ainsi mieux appréhender les motivations des utilisateurs qui se connectent.

## References

1. Albert, C. (2021). Le recours au Compte personnel de formation par les salariés du secteur privé : quelles évolutions entre 2019 et 2020 ? *Questions Politiques Sociales - Les brèves* n°10, septembre 2021, disponible sous : <https://politiques-sociales.caissedesdepots.fr/qps-les-breves-ndeg10>
2. Cisel, M. (2018). Le choix d'un MOOC, un processus influencé par l'organisation des cours en catalogues?. *STICEF*, 25, 121-150
3. Wintermute, E. H., Cisel, M., & Lindner, A. B. (2021). A survival model for course-course interactions in a Massive Open Online Course platform. *PloS one*, 16(1), e0245718.
4. Cisel, M. (2022). De la collecte des traces d'interaction à la publication scientifique. Un passage à l'échelle semé d'embûches. *Distances et médiations des savoirs*. 37.
5. Balmat, C. (2018), Le compte personnel de formation : Une montée en charge rapide surtout au profit des personnes en recherche d'emploi, *Dares résultats*, n°004 ; Dares, janvier 2018. <https://dares.travail-emploi.gouv.fr/IMG/pdf/2018-004.pdf>
6. Bousquet, G., et Jaumont, L. (2020). Le compte personnel de formation pour les salariés : un retour sur les coûts de formation 2018, *Questions retraite et solidarité*, n°30, février 2020, disponible sous : [https://retraitesolidarite.caissedesdepots.fr/sites/default/files/qrs\\_29.pdf](https://retraitesolidarite.caissedesdepots.fr/sites/default/files/qrs_29.pdf)
7. Boutinet, J.-P. (2019). Dans le champ de la formation professionnelle, quel partage de responsabilités pour aujourd'hui ? *Savoirs*, 50(2), 129–153. <https://doi.org/10.3917/savo.050.0129>
8. Buhl, J. L. (2014). La Déclaration sociale nominative (DSN) et l'accès aux droits. *Regards*, (2), 135-140.
9. Bullich, V. (2018). La « plateformisation » de la formation. *Distances et médiations des savoirs*, 21.
10. Bullich, V. (2022). Plateforme, plateformiser, plateformisation: le péril des mots qui occultent ce qu'ils nomment. *Questions de communication*, 40, 47-70.
11. Caillaud, P. (2023). La formation est-elle devenue un objet de consommation?. *Éducation permanente*, (1), 137-146. <https://doi.org/10.3917/edpe.234.0137>
12. Hansen, J. D., et Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science*, 350(6265), 1245-1248.

13. Isaac, H. et Volle, P. (2008). *E-commerce : de la stratégie à la mise en œuvre opérationnelle*. Paris : Pearson Education France.
14. Jaumont, L. (2021). La montée en charge du nouveau Compte personnel de formation (CPF) : quels impacts sur la demande de formation ? Questions Politiques Sociales - Les brèves n°9, septembre 2021, disponible sous : <https://politiques-sociales.caissedesdepots.fr/qps-les-breves-ndeg9>
15. Langevin, G. (2022). Formations financées par le CPF des salariés en contrats courts ou au Smic, en 2019 et en 2020. Questions Politiques Sociales – Les études n°36, février 2022, disponible sous : <https://politiques-sociales.caissedesdepots.fr/qps-les-etudes-ndeg36>
16. Luttringer, J. M., et Soldini, D. (2018). L’ambition «universaliste» du nouveau droit de la formation tout au long de la vie. *Revue française des affaires sociales*, (4), 39-52. <https://doi.org/10.3917/rfas.184.0039>
17. Mesnard, O. (2018). La dépense nationale pour la formation professionnelle continue et l’apprentissage en 2015, Dares résultats, n°034 ; Dares, juillet 2018. <https://dares.travail-emploi.gouv.fr/IMG/pdf/2018-034.pdf>
18. Perez, C. (2019). Avec le Compte Personnel de Formation: L’avènement d’une logique marchande et désintermédiée. *Savoirs*, 50(2), 87–100. <https://doi.org/10.3917/savo.050.0087>
19. Perez, C., et Vourc’h, A. (2020). *L’individualisation des dispositifs d’accès à la formation: l’exemple français du Compte Personnel de Formation* (No. 245). OECD Publishing. <https://shorturl.at/szHVZ>
20. Robinson, A. C., Anderson, C. L., et Quinn, S. D. (2020). Evaluating geovisualization for spatial learning analytics. *International Journal of Cartography*, 6(3), 331-349.