

## Utilisation de données administratives pour étudier l'éducation au Canada

Martin Pantel<sup>1</sup>

### Résumé

Le Fichier maître de l'éducation (FME) est un système qui a été élaboré pour permettre l'analyse des programmes d'enseignement au Canada. Le noyau du système est constitué de fichiers administratifs comprenant toutes les inscriptions aux programmes postsecondaires et d'apprentissage au Canada. De nouveaux fichiers administratifs deviennent disponibles sur une base annuelle. Lorsqu'un nouveau fichier devient disponible, il est d'abord sujet à un traitement initial, qui comprend un couplage avec d'autres dossiers administratifs. Ce couplage produit de l'information qui peut améliorer la qualité du fichier; il permet d'autres couplages avec d'autres données décrivant le marché du travail, et il constitue la première étape pour ajouter le fichier au FME. Une fois qu'elle fait partie du FME, l'information du fichier peut être incluse dans des projets transversaux ou longitudinaux, afin d'étudier les cheminement scolaires et les résultats sur le marché du travail par suite de l'obtention du diplôme. À l'heure actuelle, le FME comprend des données de 2005 à 2013, mais il évolue à mesure que de nouvelles données deviennent disponibles. Le présent document donne un aperçu des mécanismes utilisés pour créer le FME, en mettant l'accent sur la structure du système final et une partie de son potentiel analytique.

Mots-clés : éducation; données administratives; fichiers couplés

## 1. Introduction

Le Fichier maître de l'éducation (FME) a été créé à Statistique Canada pour augmenter les possibilités analytiques des données sur les programmes d'enseignement au Canada. À cette fin, on crée un identificateur au niveau de l'étudiant, qui est uniforme pour tous les ensembles de données sur l'éducation, et on facilite les couplages avec d'autres fichiers administratifs. Le système comprend de nombreuses activités de couplage de données, mais celles-ci ne sont pas abordées de façon détaillée ici. On met plutôt l'accent sur la description de la structure du FME, ainsi que sur les processus liés à l'ajout d'un nouveau fichier dans le système ou à l'extraction d'un fichier analytique.

### 1.1 Raison motivant la création du FME

Les données administratives sont de plus en plus disponibles et faciles à utiliser. Lorsqu'une nouvelle question de recherche se pose, un analyste doit envisager la possibilité que les données existent déjà sous une forme ou une autre. Les enquêtes sur échantillon ont de toute évidence leur place dans la recherche scientifique : elles sont plus souples et peuvent être conçues pour un ensemble particulier de questions de recherche. Toutefois, si les données existent déjà sous une forme appropriée, leur utilisation pour la recherche présente un fardeau moins grand et est généralement plus opportune et rentable que la tenue d'une enquête sur échantillon.

Des études de l'enseignement postsecondaire ont été menées à Statistique Canada par le passé, mais la plupart étaient transversales. Parmi les sources récentes de données longitudinales sur l'éducation figurent l'Enquête nationale auprès des diplômés/l'Enquête de suivi auprès des diplômés (END/ESD). Les données les plus récentes de ce projet visaient les diplômés de 2000 : un échantillon de diplômés a été sélectionné pour l'END en 2002, et les répondants ont été interviewés une deuxième fois en 2005 pour l'ESD. Depuis, deux autres versions de l'END ont été menées (en 2007 et 2013), mais aucune enquête de suivi n'a été administrée pour ces années de collecte. Par conséquent, un besoin a été déterminé au chapitre de la recherche : obtenir des données sur les diplômés récents nous permettant d'étudier leur carrière universitaire et leur vie après l'obtention du diplôme.

---

<sup>1</sup>Martin Pantel, Statistique Canada (Ottawa, Canada, K1A 0T6). [Martin.Pantel@Canada.ca](mailto:Martin.Pantel@Canada.ca)

## 1.2 Sources des données

Deux sources de données constituent les composantes principales du FME : le Système d'information sur les étudiants postsecondaires (SIEP) et le Système d'information sur les apprentis inscrits (SIAI). Les deux sont des registres annuels qui suivent toutes les inscriptions dans des établissements postsecondaires publics ou des programmes d'apprentissage au Canada.

Le SIEP représente un recensement des inscriptions dans tous les établissements postsecondaires publics au Canada. Il recueille des données concernant les programmes, par exemple, leur classification et leur durée. Chaque inscription comprise dans le fichier comprend aussi des données sur l'étudiant, y compris certains renseignements démographiques (nom, date de naissance, adresse actuelle et numéro de téléphone). Un fichier complet du SIEP comprend habituellement environ trois millions d'enregistrements. Les données sont fournies par les établissements proprement dits ou, dans certains cas, par les ministères provinciaux de l'Éducation. Ainsi, elles ont des degrés divers de complétude et de qualité, qui doivent être pris en compte. Il existe certaines lacunes dans le FME, une province particulière pouvant être complètement absente pour une année dans le SIEP, ou comporter des données de qualité moindre. Des efforts soutenus sont déployés pour confirmer et améliorer la qualité des données recueillies.

Le SIAI comprend tous les apprentis inscrits qui suivent de la formation en salle de classe ou en milieu de travail, dans des métiers où la formation en apprentissage est à tout le moins volontaire (sinon obligatoire). Tout comme dans le SIEP, chaque inscription dans le fichier comprend des données sur le programme, ainsi que sur l'étudiant. Un fichier complet du SIAI comporte habituellement 450 000 enregistrements.

Afin d'évaluer les rapports entre l'éducation et le marché du travail, les fichiers du SIEP et du SIAI peuvent être couplés au fichier sur la famille T1 ou FFT1. Il s'agit d'une base de données élaborée et gérée par Statistique Canada, qui est tirée des fichiers de déclarations de revenu (T1) et d'autres fichiers administratifs. Pour un exercice donné, le FFT1 fournit des renseignements sur le revenu provenant de diverses sources, le domaine d'emploi, les frais de scolarité et les déductions au titre des études, les transferts gouvernementaux, ainsi que certaines données démographiques et géographiques. Le couplage avec le FFT1 est effectué au moyen du numéro d'assurance sociale (NAS).

Au moment de la rédaction du présent document, le FME comprenait des fichiers du SIEP de 2008 à 2013 et des fichiers du SIAI de 2008 à 2013, ainsi que des données fiscales de 2006 à 2013. Il comprenait aussi des données du SIEP pour 2005 à 2007 pour les quatre provinces de l'Atlantique au Canada. Toute combinaison de ces fichiers peut être utilisée pour l'extraction d'un fichier analytique à des fins de recherche.

## 2. Structure du FME

Au niveau conceptuel, le FME peut être décrit comme comprenant trois éléments :

- 1) Un identificateur au niveau de l'étudiant (appelé MASTERID), qui peut être lié à chaque fichier du SIEP ou du SIAI dans le système, ce qui permet le couplage de tous les fichiers. Il existe aussi un enregistrement de chaque NAS associé à chaque MASTERID, afin de permettre le couplage avec les données fiscales. Les MASTERID et les NAS sont considérés comme des données très délicates : leur accès se limite à quelques personnes qui sont responsables de la mise à jour du FME et de l'extraction des fichiers analytiques. Le système répond à toutes les exigences énoncées dans les politiques et lignes directrices strictes de Statistique Canada concernant le couplage d'enregistrements et la confidentialité.
- 2) Un mécanisme pour ajouter un nouveau fichier au système qui, parallèlement, permet des corrections et la mise à jour du système existant.
- 3) Un mécanisme pour extraire les fichiers analytiques anonymes. Il peut s'agir de fichiers transversaux ou longitudinaux, et cela peut comprendre plusieurs années du SIEP, du SIAI et des données du FFT1.

Au niveau informatique, le FME peut être considéré comme trois ensembles de fichiers, chacun étant entreposé séparément :

- 1) Les fichiers de contacts comprennent toutes les données de contact, comme les noms, adresses et numéros de téléphone. Lorsque l'on ajoute un nouveau fichier au FME, l'ensemble de fichiers de contact existants sert à déterminer les enregistrements qui existent déjà dans le système. Si de nouvelles données de contact deviennent disponibles (p. ex., une personne obtient un nouveau numéro de téléphone), les anciennes données sont conservées dans le système. Parmi les utilisations futures possibles de l'ensemble de fichiers de contact figure l'élaboration d'une base de sondage ou l'obtention de données de contact/dépistage.
- 2) Les fichiers de programmes comprennent tous les renseignements concernant les programmes d'enseignement, l'établissement, etc. – soit essentiellement, toutes les données sur l'enseignement qui feront partie du fichier extrait fourni à l'analyste.
- 3) Les fichiers de clés comprennent toutes les clés de couplage. Pour une année donnée, le fichier comprendra une variable d'identification de l'enregistrement (essentiellement le numéro de ligne), qui correspond directement à la même variable dans les fichiers de contacts et de programmes. Un MASTERID est associé à chaque enregistrement dans le fichier de clés, ce qui fait que le même étudiant peut être identifié ailleurs dans le même fichier et dans tout autre fichier du FME. Le fichier de clés comprend aussi tous les NAS associés à ce MASTERID, et pour chaque NAS, un score qui indique la qualité du NAS, c'est-à-dire la façon dont il a été obtenu et la valeur du couplage.

### **3. Traitement d'un nouveau fichier**

Une partie importante du traitement vient du fait que ce ne sont pas tous les enregistrements du fichier qui concernent des étudiants différents. Certains étudiants sont inscrits à plusieurs programmes la même année et, dans d'autres cas, un enregistrement figure en double dans le fichier en raison d'une erreur administrative.

Au moment de la réception d'un nouveau fichier du SIEP ou du SIAI, la première étape consiste à attribuer un numéro d'identification d'enregistrement à chaque ligne du fichier. Puis, le fichier est couplé aux dossiers administratifs pour obtenir le plus grand nombre de NAS possible pour chaque numéro d'identification d'enregistrement. Les renseignements figurant dans le fichier, y compris les NAS nouvellement obtenus, servent par la suite à déterminer les cas où deux enregistrements ou plus concernent la même personne.

Une fois que les enregistrements se rapportant à la même personne dans le nouveau fichier ont été identifiés, le fichier est prêt à être ajouté au FME. À partir de tous les renseignements disponibles, chaque personne du nouveau fichier est recherchée dans tous les fichiers qui existent déjà dans le système. Un identificateur au niveau de l'étudiant appelé MASTERID est attribué, et il est uniforme pour l'ensemble des fichiers qui constituent le FME. Enfin, tous les renseignements sont entreposés de façon sécuritaire et confidentielle, et ils sont prêts à être utilisés pour l'extraction des fichiers analytiques. Les sections qui suivent décrivent chaque étape de ce processus de façon plus détaillée.

#### **3.1 Identification des étudiants communs dans le nouveau fichier**

Après avoir reçu son numéro d'identification, chaque enregistrement est couplé à une base de données administratives qui, à nos fins, peut être décrite comme un répertoire des NAS pour tous les Canadiens (qui comprend aussi les noms et d'autres données de contact). Ce couplage est effectué au moyen d'un programme de couplage déterministe hiérarchique interne appelé MixMatch. Le logiciel permet à l'utilisateur de comparer les noms, les dates de naissance, les numéros de téléphone et les codes postaux des deux fichiers, ainsi que de quantifier les similitudes entre les enregistrements. MixMatch utilise un ensemble d'outils d'appariement qui sont axés particulièrement sur la comparaison de chaînes de caractères, comme les noms et dates de naissance.

Après ce couplage, un fichier typique du SIAI comportera des NAS pour plus de 98 % des enregistrements. Dans le cas du SIEP, les taux varient beaucoup; si le fichier du SIEP est complet (très peu ou aucune non-réponse à des

questions), les taux se situent autour de 91 %. Toutefois, certains fichiers comportent des renseignements très limités; sans le nom, par exemple, il devient difficile de coupler les enregistrements avec certitude.

Lorsque les NAS figurent dans le fichier, on peut rechercher les enregistrements multiples appartenant au même étudiant. À cette fin, on utilise trois outils principaux. À titre d'identificateur unique des Canadiens, le NAS est un outil excellent pour déterminer les paires d'enregistrements qui se rapportent à la même personne. Dans un établissement, le numéro d'identification d'étudiant représente aussi un identificateur unique au niveau de l'étudiant. Le troisième outil est MixMatch, encore une fois, mais dans ce cas, il est utilisé pour procéder à un « couplage interne »; plutôt que de chercher des enregistrements communs entre deux fichiers distincts, le couplage se fait entre le fichier et une copie exacte de celui-ci. Le résultat est la liste des numéros d'identification d'enregistrement du fichier qui décrivent la même personne.

Il existe deux autres raisons importantes pour obtenir le plus grand nombre de NAS possible. La valeur d'un fichier analytique extrait peut être améliorée en le couplant avec les données décrivant les activités sur le marché du travail. Le NAS permet de coupler les fichiers de l'éducation aux données du FFT1. L'autre raison importante de la présence du NAS est qu'il permet l'élaboration et la définition des stratégies de couplage de MixMatch : au moment de l'élaboration d'un algorithme qui compare les noms et les dates de naissance, les NAS disponibles peuvent servir de base pour calculer les taux de faux positifs et de faux négatifs.

Le tableau 3.1-1 fournit un exemple simplifié de l'apparence que pourrait avoir un fichier entrant, ainsi que de la façon dont les renseignements additionnels sont couplés. La première étape consiste à ajouter un numéro d'identification d'enregistrement (numéro de ligne). Puis, les NAS sont ajoutés au fichier; certains d'entre eux figurent déjà dans le fichier, mais la plupart devront être obtenus à partir d'un couplage avec le répertoire des NAS. Enfin, le NAS, le numéro d'identification d'étudiant et les données de contact (noms, dates de naissance, etc.) peuvent servir à identifier les enregistrements multiples appartenant à la même personne. Un numéro d'identification unique (ou UID) est établi comme identificateur au niveau de la personne.

Dans cet exemple, si on examinait uniquement les données entrantes (les quatre premières colonnes), il ne serait pas évident que les deux premiers enregistrements se rapportent à la même personne. Toutefois, le couplage avec les données fiscales pourrait avoir produit le même NAS pour ces deux enregistrements, ce qui permettrait d'établir le lien. L'UID attribué a révélé que les enregistrements L001 et L002 se rapportaient à la même personne, U001.

**Tableau 3.1-1**  
**Exemple d'un nouveau fichier du SIAP, avec le NAS et le MASTERID**

Établissement	Numéro d'identification d'étudiant	Nom	Programme		Numéro d'identification d'enregistrement	NAS	UID
Univ. de Montréal	(manquant)	Robert	Économie	←	L001	112233	U001
Univ. de Montréal	12345	Bob	Statistique		L002	112233	U001
Collège Western	GD89	Marie	Génie		L003	445588	U002

Le NAS et le numéro d'identification unique sont considérés comme très délicats et sont conservés très soigneusement. Afin de respecter les règles strictes régissant le couplage d'enregistrements et les politiques en matière de confidentialité, les données d'identification et les clés de couplage doivent être entreposées séparément. Ainsi, une fois que le fichier figurant dans le tableau ci-dessus est complet, il est fractionné en trois éléments, qui seront entreposés dans des emplacements distincts. Le numéro d'identification d'enregistrement est le trait commun qui permet de recréer le fichier original. Pour plus de concision, seulement une variable figure dans les fichiers de programmes et de contacts du tableau 3.1-2, mais ces fichiers comprendraient aussi de nombreuses autres variables de nature similaire.

**Tableau 3.1-2**  
**Fractionnement du nouveau fichier du SIEP pour l'entreposage**

Fichier de programmes		Fichier de contacts		Fichier de clés		
Numéro d'identification d'enregistrement	Programme	Numéro d'identification d'enregistrement	Nom	Numéro d'identification d'enregistrement	NAS	UID
L001	Économies	L001	Robert	L001	112233	U001
L002	Statistiques	L002	Bob	L002	112233	U001
L003	Génie	L003	Marie	L003	445588	U002

### 3.2. Ajout du fichier dans le FME

Les trois fichiers décrits dans le tableau 3.1-2 (Programmes, Contacts et Clés) se rapportent tous à l'état d'entreposage final d'un fichier (SIEP ou SIAI pour une année donnée). En fait, le système du FME représente l'ensemble de tous ces fichiers : tous les fichiers de programmes, pour tous les fichiers annuels du SIEP et du SIAI, sont entreposés ensemble à un endroit. Tous les fichiers de contacts et les fichiers de clés sont regroupés aussi, et les trois ensembles sont conservés dans des emplacements distincts.

Lorsque le traitement initial d'un nouveau fichier est terminé (numéro d'identification d'enregistrement, NAS et numéro d'identification unique ajouté), le fichier est prêt à être ajouté au FME. Le processus peut être difficile et long à exécuter, mais il est simple à décrire : le numéro d'identification unique est mis à jour dans un MASTERID qui est uniforme pour l'ensemble des fichiers. À cette fin, chaque nouvel enregistrement est comparé à tous les enregistrements qui existent déjà dans le FME. Si un appariement est trouvé, le MASTERID existant est attribué aux nouveaux enregistrements et remplace leur numéro d'identification unique. Une fois que tous les enregistrements couplés ont reçu des MASTERID existant au préalable, les numéros d'identification unique qui restent peuvent être remplacés par de nouveaux MASTERID.

Les couplages sont effectués selon la même série d'étapes que celle décrite à la section 3.1. Les NAS des nouveaux enregistrements sont comparés aux NAS appartenant à tous ceux qui existent déjà dans le FME. Les numéros d'identification d'étudiant sont comparés aussi, à condition que les établissements soient aussi les mêmes. Enfin, tous les fichiers de contacts existants sont combinés, et les noms/dates de naissance/numéros de téléphone/codes postaux sont comparés avec les mêmes renseignements provenant du nouveau fichier.

Lorsque des stratégies différentes mènent à des couplages conflictuels, seul l'appariement « le plus fort » est conservé. La mesure de la force dépend de la source du couplage. Si le couplage a été obtenu par MixMatch, le score rend compte de la similitude des enregistrements, selon la définition des règles de MixMatch. Si le couplage a été obtenu en appariant les NAS, la mesure de la force indique si le NAS figurait dans le fichier original, ou s'il a été obtenu par suite d'un ou de plusieurs processus de couplage.

### 3.3 Mises à jour et corrections du FME

L'ajout d'un nouveau fichier au FME fait entrer de nouvelles données dans le système, ce qui peut mettre en lumière certaines erreurs dans le nouveau fichier ou dans la version existante du FME. Il existe deux types d'erreurs :

- 1) Faux appariements : Deux personnes différentes sont couplées en raison de noms similaires ou parce qu'on leur a attribué le même NAS par inadvertance. Par exemple, une erreur pourrait être attribuable à des jumeaux (même date de naissance et même nom de famille) ayant des prénoms similaires, qui s'inscrivent au collège et décident de vivre ensemble (même numéro de téléphone et même code postal).
- 2) Appariements manqués : Deux enregistrements qui se rapportent à la même personne peuvent être laissés de côté si les données sur l'un des deux ou les deux enregistrements sont manquantes, erronées ou ont changé.

Par exemple, certains établissements n'ont pas fourni les noms des étudiants, ce qui complique extrêmement leur recherche dans le système.

Des corrections peuvent être apportées au FME (ou au fichier d'entrée) lorsque de telles erreurs sont décelées. La correction du FME consiste à mettre à jour les divers fichiers de clés, en modifiant ou en mettant à jour les MASTERID, et en supprimant les NAS au besoin. La majorité des mises à jour sont attribuables à des appariements manqués, et elles peuvent être résolues de façon automatique en établissant les MASTERID à la même valeur. Les faux appariements représentent un groupe beaucoup plus petit, mais ils peuvent quand même représenter des milliers de cas, et étant donné que leur résolution nécessite une vérification et une correction manuelles, la tâche peut être assez longue.

#### 4. Exemple : Extraction d'un fichier analytique du FME

Supposons que nous souhaitons étudier les personnes qui étaient inscrites à des études postsecondaires (c.-à-d. dans le SIEP) en 2010, 2011 et 2012, ainsi que leurs données fiscales de 2012. Nous commencerions par examiner les fichiers de clés pour les trois années d'intérêt du SIEP, et nous verrions quels MASTERID existent dans les trois fichiers. Le tableau 4.1-1 montre que deux MASTERID répondent à cette condition. Pour tous les cas sélectionnés comportant plus d'un NAS dans le fichier (cela se produit dans environ 4 % des cas), nous choisirions à ce moment-là le plus approprié pour le couplage avec les données fiscales de 2012. À l'occasion, un enregistrement comportant un NAS ne peut pas être couplé aux données fiscales, soit parce qu'on n'a pas été en mesure de le faire (une erreur dans le NAS inscrit, par exemple) ou que la personne n'a pas produit de déclaration de revenus pour cette année.

**Tableau 4.1-1**

##### Exemple de fichiers de clés des trois versions du SIEP

Clés du SIEP 2010			Clés du SIEP 2011			Clés du SIEP 2012		
N° d'identification d'enregistrement	MASTERID	NAS	N° d'identification d'enregistrement	MASTERID	NAS	N° d'identification d'enregistrement	MASTERID	NAS
L11	M01	101	L21	M06	201	L31	M09	301
<b>L12</b>	<b>M02</b>	<b>102</b>	L22	M07	202	<b>L32</b>	<b>M02</b>	<b>102</b>
<b>L13</b>	<b>M02</b>	<b>102</b>	<b>L23</b>	<b>M02</b>	<b>102</b>	L33	M10	302
L14	M03	103	L24	M08	203	L34	M11	303
<b>L15</b>	<b>M04</b>	<b>104</b>	<b>L25</b>	<b>M04</b>	<b>104</b>	<b>L35</b>	<b>M04</b>	<b>104</b>
L16	M05	105	L26	M05	105	L36	M12	304

Chaque fichier (année du SIEP) peut maintenant être traité de façon indépendante : les numéros d'identification d'enregistrement sélectionnés servent à tirer les données analytiques du fichier de programmes approprié, et les NAS servent à tirer les données appropriées du fichier des données fiscales.

Au besoin, des variables indicatrices peuvent aussi être créées et ajoutées aux fichiers. Certaines variables d'indicateur commun décrivent les caractéristiques entourant le NAS. Par exemple, des NAS provisoires sont parfois attribués aux immigrants récents ou aux victimes de vol d'identité. Ces NAS provinciaux peuvent facilement être identifiés grâce à leur premier chiffre. Ils peuvent présenter un intérêt pour les analystes, étant donné qu'il peut y avoir un NAS différent (permanent) dans un autre fichier se rapportant à la même personne. Comme les problèmes de confidentialité empêchent la diffusion du NAS proprement dit dans le fichier final, l'analyste peut avoir besoin d'une étiquette binaire indiquant simplement si le NAS est provisoire ou non.

Les fichiers définitifs fournis à l'analyste pourraient ressembler à l'ensemble de fichiers décrits ici :

**Tableau 4.1-2**  
**Exemple de fichiers analytiques**

<b>SIEP 2010</b>			
<b>N° d'identification de demande</b>	<b>Données de programme</b>	<b>Données fiscales de 2012</b>	<b>Indicateurs</b>
<b>R01</b>	<b>Programme d'enseignement, données sociodémographiques, etc., 2010</b>	<b>Revenu, frais de scolarité, etc., 2012</b>	<b>Étiquettes pour les NAS multiples, les NAS temporaires, etc.</b>
<b>R01</b>			
<b>R02</b>			

<b>SIEP 2011</b>			
<b>N° d'identification de demande</b>	<b>Données de programme</b>	<b>Données fiscales de 2012</b>	<b>Indicateurs</b>
<b>R01</b>	<b>Programme d'enseignement, données sociodémographiques, etc., 2011</b>	<b>Revenu, frais de scolarité, etc., 2012</b>	<b>Étiquettes pour les NAS multiples, les NAS temporaires, etc.</b>
<b>R02</b>			

<b>SIEP 2012</b>			
<b>N° d'identification de demande</b>	<b>Données de programme</b>	<b>Données fiscales de 2012</b>	<b>Indicateurs</b>
<b>R01</b>	<b>Programme d'enseignement, données sociodémographiques, etc., 2012</b>	<b>Revenu, frais de scolarité, etc., 2012</b>	<b>Étiquettes pour les NAS multiples, les NAS temporaires, etc.</b>
<b>R02</b>			

On peut souligner ici encore une fois l'anonymat des fichiers : les fichiers analytiques ne comprennent pas de NAS, ni d'identificateur personnel, mais uniquement des données analytiques et des variables indicatrices. Un numéro d'identification de demande est créé pour masquer le MASTERID réel : la mise en correspondance entre les deux peut être déchiffrée uniquement par les quelques personnes qui ont accès au FME. Par ailleurs, la mise en correspondance est valide uniquement pour cette demande; le même MASTERID obtiendrait un numéro d'identification de demande différent s'il était extrait d'un autre fichier analytique.

## 5. Travaux futurs

Dans le système actuel, chaque enregistrement couplé comporte un score qui rend compte de sa force, mais uniquement au moment où le couplage particulier a été établi. Ce score est particulièrement utile pour les mises à jour ou les corrections du FME. Toutefois, étant donné que le FME est dynamique, il serait utile que le score puisse changer, afin de rendre compte des corrections apportées à cet enregistrement. La forme de cet indicateur de qualité est difficile à définir, mais il pourrait s'agir d'un outil utile pour les analystes.

Le FME continuera d'évoluer, au fur et à mesure que des fichiers s'ajouteront au système. On s'attend à ce que les nouveaux SIEP, SIAI et FFT1 soient disponibles chaque année. Il est aussi possible que la valeur analytique du FME soit élargie, grâce au couplage de ses données avec celles provenant d'autres secteurs, comme le Recensement de la population, ou de bases de données décrivant les secteurs de la santé, de la justice ou d'autres secteurs.

## **Bibliographie**

Caron, P., Ellison, J. et Faucher, D. (2013) “Linkage of Education Files”, document interne, Ottawa, Canada: Statistique Canada.

Pantel, M. (2015) “Educational Master File (EMF) Methodology”, unpublished document, Ottawa, Canada: Statistics Canada.