



ETUDE DES TRAJECTOIRES DES JEUNES DANS LES DISPOSITIFS BRUXELLOIS FRANCOPHONES D'ALTERNANCE

Synthèse du rapport complémentaire

Février 2015

Commanditaires :

Région de Bruxelles Capitale

Fédération Wallonie Bruxelles

Bruxelles-Formation

efp/SFPME

CCFEE/ IB EFE Bruxelles

Conseil Zonal de l'Alternance de Bruxelles

Centre de Coordination et de Gestion des programmes européens

Altis

Table des matières

A. Contexte	3
B. Description succincte des bases de données des opérateurs	4
1. Bases de données des CEFA.....	4
2. Bases de données de l'efp/SFPME.....	5
C. Récolte des données	5
1. Récolte des données CEFA.....	6
2. Récolte des données efp/SFPME.....	7
D. Principales difficultés de traitement des données	8
1. Extraction des données.....	8
2. Coquilles et manque d'harmonisation de l'encodage.....	8
3. Identifiant personnel propre à chaque base de données.....	9
4. Identification des sortants des formations.....	9
5. Isolement des informations relatives aux trajectoires de formation.....	10
6. Manque d'harmonisation dans les codes relatifs au domaine de formation.....	10
7. Lacunes dans l'encodage du niveau à l'entrée en formation.....	10
8. Lacunes et difficultés d'interprétation de l'encodage du résultat de la formation.....	11
E. Recommandations	11
1. Créer un référentiel avec des consignes précises d'encodage.....	11
2. Harmoniser les nomenclatures.....	12
3. Scinder les valeurs regroupées en autant de variables que nécessaires.....	12
4. Ajouter des variables utiles au pilotage.....	12
5. Apporter un soin particulier à l'identifiant personnel de chaque individu.....	13
6. Créer une table avec les données agrégées du dernier état pédagogique.....	13
7. Créer une base de données centralisée.....	14
8. Intégrer la base de données centralisée dans le Datawarehouse Marché du Travail et Protection sociale (Datawarehouse MT&PS).....	14

A. Contexte

Sur une période de 2 ans, le centre de recherche METICES-ULB a mené une étude des trajectoires des jeunes dans les dispositifs bruxellois francophones d'alternance. Cette étude, commanditée par la Région de Bruxelles-Capitale, la Fédération Wallonie Bruxelles, Bruxelles Formation, l'efp/SFPME, les CEFA bruxellois rassemblés au sein du Conseil Zonal de l'Alternance et l'Instance Bassin enseignement qualifiant formation emploi¹, avait pour objectifs de :

- dénombrer les jeunes concernés par ces dispositifs
- déterminer une typologie de profils de ces jeunes
- analyser leurs trajectoires pendant la formation
- examiner leur entrée sur le marché du travail

Pour dépasser le comptage officiel à date fixe et pouvoir rendre compte des flux d'entrées et de sorties dans les dispositifs, l'équipe de recherche a rassemblé les données disponibles auprès des 5 CEFA bruxellois, de l'efp et du SFPME. L'idée de départ était de rassembler ces données dans une base de données commune qui servirait de point de départ à la recherche. Mais la récolte des données a rapidement révélé de nombreuses disparités entre les différentes bases, et l'optique d'une base de données commune a dû être abandonnée.

Au vu des nombreuses difficultés rencontrées, tant au niveau de la récolte, qu'au niveau de la tentative de compilation des données, le comité d'accompagnement de l'étude a considéré indispensable de disposer d'une analyse ad hoc de ces difficultés mais aussi de propositions d'amélioration des systèmes existants. Tirer de tels enseignements méthodologiques bénéficierait à l'ensemble des acteurs de l'alternance, notamment dans le cadre de la mise en place du futur Office Francophone de la Formation en Alternance (OFFA).

Le rapport complémentaire, synthétisé ici, poursuit cet objectif. Il a été réalisé de novembre 2014 à février 2015, grâce au financement complémentaire octroyé par Altis et le Centre de Coordination et de Gestion des programmes européens (CCG). Ces deux instances ont intégré le comité d'accompagnement de l'étude.

Dans ce document de synthèse, après une brève description des bases de données des opérateurs, nous nous attachons à décrire la récolte des données et les principales difficultés que METICES-ULB a rencontrées lors du traitement de celles-ci. Finalement, nous exposons les recommandations méthodologiques de METICES-ULB visant à améliorer et harmoniser les pratiques et les outils de collecte de données.

¹ À ce moment-là, la Commission Consultative Formation Emploi Enseignement.

B. Description succincte des bases de données des opérateurs

1. Bases de données des CEFA

Quatre CEFA sur cinq utilisent, depuis 2007, une base de données gérée par ProEco. Ce logiciel est utilisé pour la gestion administrative de la fiche signalétique de l'élève, les diplômes et historiques des diplômes, l'entrée/sortie du CEFA et les raisons de sorties prématurées (abandons/réorientations), ainsi que les contrats de stage (durées et entreprises partenaires)².

ProEco fonctionne sur deux années scolaires : l'année scolaire en cours et l'année scolaire précédente. Les données des années précédentes sont normalement « archivées » dans la base de données. L'opération d'archivage, de l'avant dernière période scolaire, est effectuée lors de la clôture de l'année scolaire en cours. Si l'élève ne quitte pas l'établissement, les données le concernant sont reportées pour l'inscription de l'année scolaire suivante.

Dans chaque CEFA, le logiciel est installé sur un serveur, dont l'accès est protégé et les droits d'utilisation sont modulés selon l'utilisateur qui le consulte. L'encodage est effectué par plusieurs personnes, selon un mode qui peut varier au fil des années. Par exemple, les systèmes de codification ont subi des modifications au cours des années. Ces modifications de codes sont conformes aux évolutions des nomenclatures propres au réseau auquel le CEFA est rattaché. Par ailleurs, l'encodage d'une variable peut devenir obligatoire d'une année à l'autre ; c'est le cas notamment de l'information sur la scolarité avant l'entrée au CEFA. Finalement, il faut garder en mémoire qu'en fonction de l'encodeur ou peut-être par manque de temps, certaines variables sont ou ne sont pas encodées.

Le 5^{ème} CEFA utilise deux programmes :

- CTI (relié au réseau ETNIC) : utilisé pour enregistrer la fiche signalétique de l'élève.
- ISIS : utilisé pour la gestion des absences et des bulletins. Le module « stage » de ce programme est prévu mais n'est actuellement pas fonctionnel.

Le programme CTI fonctionne en mode MS-DOS à partir d'une console reliée au réseau ETNIC. L'opérateur du CEFA, à partir de la console, peut obtenir et imprimer des listes, au moyen de requêtes prédéfinies.

Le programme ISIS est un programme propriétaire, géré par la société ECONCEPT. Sa base de données est hébergée et gérée sur un serveur de cette société.

Un transfert partiel de données (partie de la fiche signalétique de l'élève) est effectué entre les données CTI et le programme ISIS. Le résultat que l'opérateur du CEFA peut obtenir est une liste au format texte non formaté, dont l'extraction automatique des données s'est révélée impraticable.

² Notons que seuls 2 de ces 4 CEFA ont encodé les données relatives aux stages dans le logiciel. Les deux autres ont optés pour les dossiers papier.

2. Bases de données de l'efp/SFPME

La gestion informatique des jeunes en formation dans le dispositif des Classes moyennes est effectuée à l'aide de deux bases de données indépendantes :

- La base de données « Merlin » (créée en interne en 2007 par l'efp) : il s'agit d'une base de données utilisée par l'efp, qui organise les cours théoriques et pratiques des formations des classes moyennes. Ces bases de données répertorient tous les auditeurs inscrits à l'efp, et les distinguent selon la filière apprentissage (Merlin APP) ou la filière chef d'entreprise (Merlin CE). L'encodage des données est effectué par le staff administratif ou des opérateurs pédagogiques.
- La base de données « Gestion Tutelle » (développée par le CIRB³) : il s'agit de la base de données du SFPME, qui, entre autre, gère le suivi de la formation en entreprises des auditeurs en stage à Bruxelles. Y sont répertoriés tous les auditeurs qui sont sous contrat d'apprentissage ou convention de stage chez un employeur situé à Bruxelles. L'encodage des données est effectué par le staff administratif et des délégués à la tutelle.

Ces deux bases de données sont gérées indépendamment l'une de l'autre. Une partie des nomenclatures sont communes et des variables comparables sont utilisées dans les deux bases de données.

Certains auditeurs se retrouvent simultanément, ou successivement, dans les bases de données Gestion Tutelle et Merlin ; d'autres ne se retrouvent que dans l'une des deux bases de données. Cela s'explique par le fait que, dans les classes moyennes, différentes modalités de formation sont possibles : formation en alternance avec un volet en entreprise et un volet à l'efp, formation sans volet en entreprise, formation sans volet à l'efp, formation en alternance pour laquelle le volet formation en centre se fait à Bruxelles mais le volet en entreprise se fait dans une autre région etc.

Des informations sont échangées entre l'efp et le SFPME, mais les règles et les pratiques d'encodage étant spécifiques à chacune des deux bases de données, des divergences sont relevées entre les valeurs de variables similaires (erreurs d'encodage et/ou de mise à jour).

C. Récolte des données

La collecte de ces données a fait l'objet d'une demande d'utilisation de données non codées à des fins scientifiques envoyée à la Commission de Protection de la Vie Privée (CPVP), METICES-ULB agissant comme sous-traitant pour recueillir les données auprès des opérateurs et de les analyser. La CPVP a autorisé l'utilisation des données personnelles non codées nécessaires à la recherche.

³ Centre d'Information pour la Région Bruxelloise, le service informatique de la COCOF.

1. Récolte des données CEFA

1.1. Récolte des données ProEco

L'extraction des données nécessaire à cette étude a été réalisée par METICES. La formulation des requêtes a été mise au point avec des délégués des CEFA grâce au module de création de requêtes assisté propre au programme ProEco. Ce module, assez sophistiqué, génère des requêtes type dont l'opérateur n'a pas la maîtrise totale. Par cette méthode, METICES a donc pu extraire plusieurs tables pour chaque CEFA :

- Données administratives : nom ; prénom ; sexe ; date de naissance ; n° registre national ; résidence ; nationalité.
- Date d'entrée sortie de l'école.
- Données sur les diplômes : niveau scolaire à l'entrée au CEFA (en cas de disponibilité) ; diplôme obtenu en fin de formation.
- Historique de la scolarité : années scolaires ; orientation.
- Données sur les stages⁴ (en cas de disponibilité) : durée du stage ; information sur l'entreprise.

Ces différentes tables, obtenues pour chaque CEFA ont été combinées de façon à obtenir une seule base de données « ProEco ».

1.2. Récolte des données CTI/ISIS

Concernant le programme ISIS, METICE-ULB a essayé, en vain, d'obtenir une exportation des données par la société ECONCEPT.

Quant au programme CTI, il s'est rapidement révélé que l'exploitation des listes issues du programme était impraticable, puisque ces listes sont rendues au format texte non formaté, dont l'uniformisation aurait requis l'identification et l'encodage de toutes les valeurs.

Au final, les données que METICES a pu obtenir pour ce 5^{ème} CEFA, ont été fournies par les gestionnaires du réseau ETNIC, avec toutefois un nombre de variables limitées :

- Scolarité : dates d'entrée et de sortie de la formation ; année d'étude précédente ; année d'études en cours ; sortie anticipée et raison de la sortie ; année d'études d'attribution du CEB ; établissement scolaire d'origine.
- Formation : option de base ; article 45/49 ; statut de l'élève (libre ou régulier).
- La variable relative aux résultats de fin d'année n'a pas été traitée puisque encodée seulement à 28% du total des auditeurs uniques.
- Données sur les stages : non disponibles.

⁴ Il faut relever que souvent l'information disponible renseigne le siège social de l'entreprise et pas forcément le lieu de déroulement du stage.

La base de données ProEco a été combinée, dans la mesure du possible, aux données du 5^{ème} CEFA pour créer une seule base de données CEFA.

2. Récolte des données efp/SFPME

2.1. Récolte des données de l'efp

Pour cette étude, l'efp a livré deux extractions, dont METICES a pu exploiter les données suivantes :

- Dans la table « Merlin APP » :
 - Les données administratives : nom, prénom, sexe, date de naissance, numéro de registre national, résidence, nationalité.
 - La scolarité : année de formation, abandon de formation (déductible de la variable « resultat_inscription »), centre de formation PME. La variable spécifique « abandon de formation » n'est pas encodée.
 - Les contrats : dates de début et de fin, rupture de contrat (seulement la date de rupture), métier, localisation du lieu de l'entreprise.
- Dans la table « Merlin CE » :
 - Les données administratives : nom, prénom, sexe, date de naissance, numéro de registre national, résidence, nationalité.
 - La scolarité : année de formation, abandon de formation (date de l'abandon ; information aussi déductible de la variable « resultat_inscription »), examen d'entrée (seulement O/N).
 - Les conventions : dates de début et de fin, rupture de convention (seulement la date de rupture), métier, localisation du lieu de l'entreprise.

2.2. Récolte des données du SFPME

Après concertation avec le CIRB sur les données requises pour l'étude, 25 tables brutes ont été transmises à METICES. Ont été joints aux données un rapport succinct sur le programme et des informations techniques sur les variables, ainsi qu'une représentation graphique de la structure de la base de données.

L'application Gestion Tutelle permet de gérer principalement les contrats/conventions entre d'une part les apprentis et les chefs d'entreprise et d'autre part les entreprises qui les forment. Elle fournit également des informations complémentaires comme le passé scolaire, l'éventuel examen d'entrée à l'efp, les résultats aux cours généraux et professionnels, les absences, etc.

Concrètement, METICES a pu exploiter les données suivantes :

- De la table « auditeur » : la date d'une éventuelle d'exclusion de la formation et les données administratives : nom, sexe (déduit de la variable « titre »), date de naissance, numéro de registre national, résidence, nationalité et pays d'origine.
- De la table « contrat » : statut de l'auditeur (apprenti, stagiaire), début et fin de contrat, rupture, raison de la rupture (comme abandon, réorientation, non-respect contractuel par le jeune ou par l'entreprise, décision pédagogique ou disciplinaire), état du contrat, métier. Par des tables associées sont disponibles les informations sur les entreprises, le lieu de contrat, le centre de formation PME.
- Des tables « passescolaire » et « niveauetude » : information sur le niveau scolaire à l'entrée en formation à l'efp.
- De la table « resultat » et « deliberation » : résultat des cours généraux, des cours de technique professionnelle et de l'épreuve technique.
- De la table « examenentree » : résultat de l'examen d'entrée lorsque l'auditeur y a été soumis pour accéder à la formation.

D. Principales difficultés de traitement des données

1. Extraction des données

La difficulté essentielle a consisté à formuler la demande d'extraction des données sans avoir une connaissance concrète du contenu et de la structure des bases de données. Une demande identique a été formulée aux partenaires de l'étude et diverses réunions ont été nécessaires pour affiner les modalités de collecte et clarifier quelques incompréhensions. Les interlocuteurs de METICES-ULB sollicités à cette étape de la recherche n'avaient pas forcément l'habitude d'exploiter leurs données et leur système informatique à des fins de recherche. Les difficultés ont été résolues en bonne entente avec les opérateurs, par la transmission de compléments d'information ou de nouvelles extractions, mais le processus de collecte a pris beaucoup plus de temps que prévu.

2. Coquilles et manque d'harmonisation de l'encodage

Après la collecte des données auprès des opérateurs, un travail consistant de nettoyage des variables a été nécessaire dans l'ensemble des bases de données. Voici quelques exemples de corrections qui ont été effectuées:

- conversion des caractères accentués des extractions des tables efp/SFPME⁵ ;
- élimination de codes d'attente (les point d'interrogation, les astérisques, les tirets) ;
- suppression des doubles espaces, tirets (notamment dans les prénoms composés), points d'abrégié (de type « A.S. » pour « Anne-Sophie ») ;
- correction de dates improbables (un jeune né en 1894 par exemple)⁶ ;
- ajout de codes postaux manquants (sur base de la localité renseignée) ;
- harmonisation des valeurs des nationalités (ex : « Belgique », « bel », « B », « belge »).

Les traitements automatisés nécessitent une certaine rigueur dans l'encodage. Le nettoyage minutieux des variables a permis de diminuer le nombre de données intraitables.

3. Identifiant personnel propre à chaque base de données

Chacune des bases de données exploitée comporte un identifiant personnel permettant de suivre un jeune pendant son parcours. Ces identifiants sont suffisants pour identifier les jeunes dans leur parcours au sein du même opérateur, mais deviennent inadéquats lors de la mise en commun des données, même au sein du même dispositif (soit enseignement, soit formation en alternance). L'identifiant d'un jeune n'est pas le même dans la table de l'efp que dans celle du SFPME ; de même, il n'est pas identique d'un CEFA à l'autre.

Pour les besoins de l'étude, il était pourtant nécessaire que chaque jeune possède un seul et unique identifiant, de façon à pouvoir suivre son parcours d'un dispositif à l'autre. Un identifiant personnel pour chaque individu a donc été défini en une clé d'index composée de son nom, prénom et date de naissance. Grâce à cette clé d'index, la correction, l'harmonisation et la vérification des numéros de Registre National⁷ des diverses tables ont pu être réalisées, ce qui était nécessaire pour la phase relative au couplage des données avec la Banque Carrefour de la Sécurité Sociale (BCSS).

4. Identification des sortants des formations

Dans les bases de données telles qu'elles sont conçues actuellement, il est relativement difficile de voir si un jeune est sorti du dispositif. A défaut, METICES-ULB a identifié les sortants sur base de la dernière année de cours fréquentée ou du dernier contrat/stage effectué. Des ajustements ont été

⁵ Au moment de l'extraction, les variables, codées vraisemblablement en UFT-8, ont été altérées de sorte que tous les caractères accentués sont devenus illisibles (exemple : « Aurélien » devient « AurÃ%alien »). Ces caractères sont difficilement gérables lors des traitements automatisés ; ils ont donc dû être corrigés. Il serait opportun de se procurer un programme capable de reproduire les codes UFT-8 ou équivalent pour les extractions futures.

⁶ Au-delà des erreurs d'inattention, à cause des diverses versions de MS Excel utilisées, METICES-ULB a constaté un problème de conversion de dates au moment de l'importation de ces valeurs dans son logiciel. A l'avenir, pour éviter le décalage dans les dates, METICES-ULB recommande la procédure de conversion suivante : copier les valeurs dates de la feuille originale, les coller vers la feuille de destination en texte Unicode, puis formater la nouvelle variable en type date. Cet inconvénient peut être évité par l'exportation des données au format csv et pas au format xls, option d'exportation disponible dans ProEco.

⁷ Dans certaines bases de données, le numéro de registre national n'est encodé que pour moins de 20% des individus (voir détail dans le dossier complet pg 20, 37, 42, 48, 58).

nécessaires lorsque le même jeune a été retrouvé dans la base de données d'un autre opérateur d'alternance.

5. Isolement des informations relatives aux trajectoires de formation

Le traitement des données visant l'analyse des trajectoires en formation a soulevé deux types de difficultés : pour les CEFA ProEco, la difficulté a consisté à écarter les lignes n'ayant pas trait à la formation en alternance (lignes d'historique et de la scolarité avant l'entrée au CEFA) ; pour l'efp/SFPME, la difficulté a été d'identifier les auditeurs chef d'entreprise qui étaient dans une alternance effective⁸ (en tenant compte de l'âge et de la simultanéité de la formation au centre et en entreprise). L'analyse des trajectoires a nécessité l'écriture de programmes récursifs qui analysent la séquence complète des périodes de formations retenues. Les résultats ont été consignés dans des nouvelles variables pour lesquelles, au besoin, des regroupements ont été effectués pour synthétiser l'information.

6. Manque d'harmonisation dans les codes relatifs au domaine de formation

Les pratiques d'encodage, généralement transmises entre les encodeurs, peuvent varier dans le temps et selon l'encodeur. Au fil des années, les encodeurs peuvent changer, ce qui engendre facilement davantage de divergences d'encodage. Lors de l'encodage manuel, si aucune procédure de vérification n'est mise en œuvre, le traitement automatisé ultérieur nécessitera un recodage cohérent.

Pour les CEFA qui utilisent ProEco, les nomenclatures communes sont gérées par le concepteur du logiciel. Ces nomenclatures peuvent toutefois être enrichies par des codes propres à l'opérateur ou par des codes spécifiques au réseau auquel le CEFA est rattaché. METICES-ULB a constaté l'existence de nombreux codes non harmonisés, tant d'un CEFA à l'autre, qu'au sein du même CEFA. Un travail important de recodage a dû être effectué en concertation avec les opérateurs des CEFA et avec l'aide de la CCFFEE.

Pour l'efp et le SFPME, une nomenclature très stricte est utilisée pour encoder le métier visé par la formation. Aucun recodage n'a donc été nécessaire⁹.

7. Lacunes dans l'encodage du niveau à l'entrée en formation

Déterminer le niveau scolaire des élèves/auditeurs à l'entrée de la formation en CEFA ou à l'efp/SFPME n'a pas été évident. En effet, le niveau de diplôme n'est pas systématiquement encodé, soit il est encodé dans une variable contenant plusieurs valeurs, soit il est encodé dans une variable

⁸ Nous rappelons ici que la population de référence retenue pour l'étude est la suivante : tous les jeunes de 15 à 25 ans, sortant de l'un des CEFA bruxellois entre 2003 et 2012 ou sortant de l'efp/SFPME entre 2003 et 2011, après y avoir réalisé au moins une période de formation en alternance effective. Pour une définition exhaustive de la population de référence, voir le rapport final de l'Etude des trajectoires des jeunes dans les dispositifs bruxellois francophones d'alternance, pages 38 à 40.

⁹ La nomenclature « métiers » a été synthétisée selon la répartition par pôle telle que présentée dans le Rapport d'activités 2012 de l'efp. L'analyse nuancée restant par ailleurs possible par les codes métiers originaux, y compris l'analyse par année préparatoire.

« commentaire ». Dans les deux cas, la valeur pertinente doit être identifiée et recodée dans la variable adéquate.

Pour les CEFA ProEco, l'information est très inégale selon les CEFA, en particulier avant que cet encodage ne soit rendu obligatoire. Pour le 5^{ème} CEFA, si le passé scolaire n'est pas systématiquement encodé, l'obtention du CEB et, le cas échéant, l'année d'obtention de celui-ci est toujours mentionné.

Pour l'efp, l'information est encodée assez librement dans une variable texte, parmi d'autres valeurs. Pour le SFPME, trois variables sont prévues à cet effet.

De façon générale, des nomenclatures très diverses ont été utilisées (exemple de dernier diplôme obtenu : « CEB », « primaire », « 4^{ème} », « D2 », etc.). Elles ont nécessité un regroupement qui a été effectué en concertation avec la CCFEE.

8. Lacunes et difficultés d'interprétation de l'encodage du résultat de la formation

Le résultat de la formation n'est pas toujours disponible dans les bases de données des opérateurs ou est difficilement interprétable.

Pour les CEFA ProEco, l'information est encodée pour 66,5% des élèves. Cette information n'est pas nuancée : on encode « O » pour une réussite et « N » pour un échec. Une variable relative aux points obtenus est prévue, mais elle n'est pas encodée. Pour le 5^{ème} CEFA, l'information est connue pour 18,5% des élèves. Ici aussi, pas de nuance : on mentionne le passage dans l'année supérieure (avec le code « A »), un redoublement (« C »), ou un examen de passage (« E »). Dans la variable associée « Session », on est informé sur le moment de la délibération (session de juin ou de septembre).

Pour l'efp, l'information n'est pas disponible. Il existe bien une variable « résultat inscription », mais les valeurs qui s'y trouvent semblent peu fiables et ambiguës (présences de nombreuses antinomies). Pour le SFPME, il n'existe pas de variable relative au résultat final. On retrouve des variables relatives à la délibération de trois types d'épreuve (cours généraux, cours de technique professionnelle et maîtrise du métier), qui sont partiellement encodées. Pour qu'il y ait réussite final, il faut une délibération positive aux 3 épreuves. Dans la mesure du possible, et pour les besoins de l'étude, METICES-ULB a déduit les réussites/échecs finaux des auditeurs sur base des informations disponibles des 3 types de délibération, celles-ci pouvant être encodées sur des années de formation différentes.

E. Recommandations

1. Créer un référentiel avec des consignes précises d'encodage

Vu la multiplicité d'erreurs pouvant se produire à la saisie, il serait illusoire de programmer un contrôle de toutes les erreurs possibles. Cela ralentirait le programme, surtout sur des ordinateurs peu

performants, et génèrerait un inconfort pour l'encodeur. Par contre, il serait très utile que chaque opérateur ait à disposition un référentiel mentionnant les différentes consignes d'encodage.

Les variables qui nécessitent une valeur obligatoire devraient disposer de l'attribut de saisie obligatoire (nul non valide).

2. Harmoniser les nomenclatures

Pour la fiabilité de l'encodage, l'utilisation de listes déroulantes est vivement conseillée. Nous avons constaté que ce type de liste existe déjà chez certains opérateurs. Mais souvent, elles ne sont pas harmonisées. Il serait opportun de nettoyer ces listes pour rationaliser le nombre de code. Les nomenclatures de ces listes devraient également être développées de façon concertée au moins avec les opérateurs du même dispositif.

Il est toutefois possible qu'un opérateur doive disposer de codes propres ou conformes au réseau auquel il est attaché, dans ce cas, il serait utile de prévoir une table de conversion des codes équivalents des autres opérateurs du même dispositif.

3. Scinder les valeurs regroupées en autant de variables que nécessaires

Pour rendre possible le traitement automatisé des valeurs, les variables ne devraient contenir qu'une seule information. D'ailleurs, les variables « commentaires » devraient servir exclusivement à une information complémentaire et ne pas contenir de code valeur.

4. Ajouter des variables utiles au pilotage

Les variables suivantes seraient très utiles au traitement statistique en vue du pilotage de l'alternance :

- Le résultat final des formations (réussite, échec et éventuellement des informations nuancées en commentaire).
- Le dernier diplôme obtenu au moment de l'entrée en formation.
- La fin de la formation et son motif : formation achevée, abandonnée (raison de l'abandon : volonté du jeune ? suite à une mesure pédagogique ou administrative ?).
- Le type et les dates de début/fin du contrat de stage. Cette information serait particulièrement utile pour identifier les auditeurs chef d'entreprise pour lesquels le volet entreprise n'est pas obligatoire.
- Pour l'efp/SFPME particulièrement : le fait que les cours ou le stage soit ou ne soit pas obligatoire.
- Pour les CEFA particulièrement : la présence des élèves au 15 janvier¹⁰. Cette variable permettra, entre autre, de différencier les élèves pris en compte dans les comptages à date fixe.

¹⁰ Cette variable est disponible dans ProEco pour l'année en cours, mais pas pour les années archivées.

Il s'agirait de créer ces variables dans des tables spécifiques pour une manipulation plus aisée.

5. Apporter un soin particulier à l'identifiant personnel de chaque individu

L'identifiant personnel est un ressource absolument nécessaire à une analyse des trajectoires.

Au mieux, il faudrait utiliser comme identifiant personnel le numéro de Registre National de l'élève/auditeur. Ce procédé permettra entre autre de coupler plus aisément les données à celles de la BCSS. A minima, et si le numéro de Registre National ne pouvait pas être utilisé comme identifiant, il s'agirait, en particulier pour l'efp et le SFPME, d'effectuer un échange croisé d'identifiants personnels propres de façon à ce que les auditeurs puissent être identifiés dans les deux bases de données.

Quel que soit l'identifiant personnel, il est vivement recommandé :

- de donner aux encodeurs des consignes très strictes d'encodage des noms, prénoms et date de naissance (par exemple, pas d'accent, pas de tiret, pas d'espace et un modèle précis de date comme xx/xx/xxxx) ;
- d'instaurer des routines de vérification et des corrections systématiques pour s'assurer que 100% des numéros de Registre National soient encodés et corrects.

6. Créer une table avec les données agrégées du dernier état pédagogique¹¹

Afin de disposer d'un outil statistique rapide et permanent, METICES-ULB recommande la création d'une table spécifique destinée à collationner les données pédagogiques pertinentes sur le dernier état de la formation des jeunes. Cette table comprendrait des variables comme l'identifiant personnel, l'année de la 1^{ère} entrée en formation, l'année de la sortie de formation, le niveau scolaire à l'entrée, le dernier type de cours suivi, le dernier type de stage effectué, la résultat final de la formation, la dernière certification obtenue, et le cumul des durées de formation théorique et pratique, compte tenu des abandons de formation et des ruptures de contrat.

La table serait mise à jour de façon dynamique à la fin d'une période de formation ou lorsqu'une valeur de la dernière ligne de la séquence de formation d'un jeune vient à être modifiée :

- Dans ProEco, une routine générale pourrait être exécutée lors de la clôture de l'année scolaire en cours.
- Dans Merlin et Gestion Tutelle, des routines adéquates seraient à exécuter lors de la fin de l'année scolaire et, pour les contrats/stages, à solliciter en fin normale ou en cas de changement de statut d'une des variables concernées.

Cette nouvelle table permettrait de produire à tout moment une statistique qui reflète l'état de la formation de chaque élèves ou auditeur.

¹¹ Recommandations détaillées dans le rapport technique de l'Etude sur les trajectoires des jeunes dans les dispositifs bruxellois francophones d'alternance, pg 8 et 9.

7. Créer une base de données centralisée

Afin de faciliter le pilotage de l'alternance, il serait utile de créer une base de données centralisée, contenant des données pertinentes pour le suivi des trajectoires en formation et post-formation. Cette base de données, alimentée périodiquement par les opérateurs de chaque dispositif, donnerait une vision globale de l'état de l'alternance en région bruxelloise.

La création d'une telle base de données suppose :

- De trouver la couverture financière.
- De commander la programmation des logiciels de collecte et de la gestion des données.
- De désigner l'organisme préposé à la gestion de l'outil.
- De mettre en place les modalités de collecte des informations et d'exploitation des données centralisées.
- De définir les périodicités de collecte et actualisations de la base centralisée, de préférence de façon automatisée.
- De mettre en place les contrôles de la validité des transferts des données.
- De définir les modalités et droits de consultation, pour qui et à quelles conditions.

8. Intégrer la base de données centralisée dans le Datawarehouse Marché du Travail et Protection sociale (Datawarehouse MT&PS)¹²

L'intégration structurelle de cette base de données dans le Datawarehouse serait à envisager, de manière à pouvoir disposer d'un système permanent de suivi des jeunes ayant fréquenté la formation en alternance.

¹² Une information succincte est disponible à l'adresse suivante :
<https://www.socialsecurity.be/CMS/fr/news/citizen/archives/article.html?uid=409ef2e1555ba946007a7686f9778a1b&archivedate=201404>