

Article

Symposium 2008 :
Collecte des données : défis, réalisations et nouvelles orientations

Systeme de saisie et de codage en ligne de relevés de notes et de catalogues de cours d'études secondaires

par Donghui Wang, Tiffany Mattox, Brian Evans,
Daniel Pratt et Chinh Nguyen

2009



Système de saisie et de codage en ligne de relevés de notes et de catalogues de cours d'études secondaires

Donghui Wang, Tiffany Mattox, Brian Evans, Daniel Pratt et Chinh Nguyen¹

Résumé

RTI International mène une étude longitudinale sur l'éducation. L'un des volets de l'étude consistait à recueillir des relevés de notes et des catalogues de cours auprès des écoles secondaires fréquentées par les personnes comprises dans l'échantillon. Il fallait aussi saisir et coder les renseignements tirés de ces documents. Le défi était de taille puisque les relevés et les catalogues, dont le contenu et la présentation variaient grandement, ont été recueillis auprès de différents types d'écoles, dont des écoles publiques, privées et religieuses de tout le pays. La difficulté consistait à concevoir un système perfectionné pouvant être utilisé simultanément par de nombreux utilisateurs. RTI a mis au point un système de saisie et de codage des données tirées des relevés de notes et des catalogues de cours d'études secondaires. Doté de toutes les caractéristiques d'un système de saisie et de codage haut de gamme, évolué, multi-utilisateur, multitâche, convivial et d'entretien peu coûteux, le système est basé sur le Web et possède trois grandes fonctions : la saisie et le codage des données des relevés et des catalogues, le contrôle de la qualité des données à l'étape de la saisie (par les opérateurs) et le contrôle de la qualité des données à l'étape du codage (par les gestionnaires). Compte tenu de la nature complexe de la saisie et du codage des données des relevés et des catalogues, le système a été conçu pour être souple et pour permettre le transport des données saisies et codées dans tout le système afin de réduire le temps de saisie. Il peut aussi guider logiquement les utilisateurs dans toutes les pages liées à un type d'activité, afficher l'information nécessaire pour faciliter la saisie et suivre toutes les activités de saisie, de codage et de contrôle de la qualité. Les données de centaines de catalogues et de milliers de relevés de notes ont été saisies, codées et vérifiées à l'aide du système. La présente communication aborde les besoins et la conception du système, les problèmes de mise en œuvre et les solutions adoptées, ainsi que les leçons tirées de cette expérience.

1. Introduction

RTI International a été l'entrepreneur principal de la *Education Longitudinal Study* (ELS), étude conçue pour suivre la transition d'un échantillon national de jeunes à mesure qu'ils passent de la dixième année aux études secondaires, puis aux études postsecondaires et/ou au monde du travail. Un volet essentiel de l'étude consistait à recueillir des relevés de notes et des catalogues de cours auprès des écoles secondaires fréquentées par les personnes comprises dans l'échantillon, puis à saisir et à coder les renseignements tirés de ces documents dans un système électronique de collecte des données. Le défi était de taille puisque les relevés et les catalogues, dont le contenu et la présentation variaient grandement, ont été recueillis auprès de différents types d'écoles, dont des écoles publiques, privées et religieuses de tout le pays.

Pour relever ce défi, RTI a mis au point un système de saisie et de codage (*Keying and Coding System* ou KCS) des données de l'ELS. Il s'agit d'un système perfectionné pouvant être utilisé simultanément par de nombreux utilisateurs exerçant des activités de saisie et de codage ainsi que des activités de gestion. Outre son fonctionnement multitâche, le système de saisie et de codage est convivial et d'entretien peu coûteux. Il est basé sur le Web et possède trois grandes fonctions : la saisie et le codage des données, le contrôle de la qualité des données à l'étape de la saisie (par les opérateurs) et le contrôle de la qualité des données à l'étape du codage (par les gestionnaires). Compte tenu de la nature complexe de la saisie et du codage des données des relevés et des catalogues, l'ELS-KCS a été conçu pour être souple et pour permettre le transport des données saisies et codées dans tout le système afin de réduire le temps de saisie. Il peut aussi guider logiquement les utilisateurs dans toutes les pages liées à un type d'activité, afficher l'information nécessaire pour faciliter la saisie et suivre toutes les activités de saisie, de codage et de contrôle de la qualité. Les données de centaines de catalogues et de milliers de relevés de notes ont été saisies

¹ Donghui Wang, Tiffany Mattox, Brian Evans, Daniel Pratt et Chinh Nguyen. RTI International, RTP, NC, USA.

dans le système. La présente communication aborde les besoins et la conception du système, les problèmes de mise en œuvre et les solutions adoptées, ainsi que les leçons tirées de cette expérience.

2. Contexte de l'étude et préparation en vue de la saisie

2.1 Collecte des données dans les relevés de notes et les catalogues de cours

Pour donner une idée de la quantité de données à traiter par notre système, de décembre 2004 à juin 2005, nous avons envoyé des demandes de relevé de notes à 2 032 écoles fréquentées par les personnes comprises dans l'échantillon; de ce nombre, 759 écoles avaient participé pendant l'année de référence ou avaient répondu à la première enquête de suivi. Nous avons demandé aux écoles de fournir des renseignements de base sur l'inscription, les tests et les cours de chaque élève ainsi que sur les politiques et les exigences de l'école en matière de notation et de collation des grades. Entre autres types de données, nous demandions le type de diplôme et la date d'obtention, la date et la raison du départ de l'élève (obtention du diplôme ou transfert, par exemple), le titre et le numéro de chaque cours suivi, ainsi que la note et les crédits obtenus pour chaque cours suivi.

Nous avons recueilli les catalogues de cours pendant l'année de référence (2001-2002) au cours de laquelle les élèves échantillonnés étaient inscrits en dixième année, puis pendant la première année de suivi (2003-2004). En tout, nous avons recueilli les catalogues de cours de quatre années scolaires couvrant les années 2000 à 2004. Dans le cadre de cette activité, nous avons relancé les écoles dont nous n'avions pas reçu les catalogues. Dans l'ensemble, le nombre élevé de catalogues, de relevés de notes et de renseignements connexes ont constitué un facteur déterminant de la conception de notre système de saisie et de codage.

2.2 Résultats de la collecte des données

L'étude sur l'éducation a connu un franc succès : au chapitre de la collecte des relevés de notes, le taux de participation non pondéré des écoles a atteint 79,3 % et leur taux de réponse pondéré pour l'année de référence s'est élevé à 94,5 %. Le taux de couverture final des élèves a atteint 91 % (soit 90,7 %, pondéré) de l'ensemble de l'échantillon d'élèves pour lesquels nous avons reçu au moins un relevé relatif à un cours. Au chapitre de la collecte des catalogues de cours, le taux de réponse des écoles pour l'année de référence a atteint 88,0 %. Après avoir reçu, examiné et enregistré les données dans le système de contrôle, nous avons acheminé environ 15 700 relevés de notes et 700 catalogues de cours en vue de la saisie et du codage, soit une énorme quantité de données.

Tableau 2.2-1
Taux de réponse à la collecte des données

	Écoles (taux non pondéré*)	Écoles, année de référence (taux pondéré*)	Relevés de notes (taux pondéré*)	Catalogues de cours, année de référence
Taux de réponse (%)	79,3	94,5	90,7	88,0

Non pondéré* --- Les taux non pondérés sont fondés sur le compte des dossiers.

Pondéré* --- Les taux pondérés sont fondés sur les estimations pondérées.

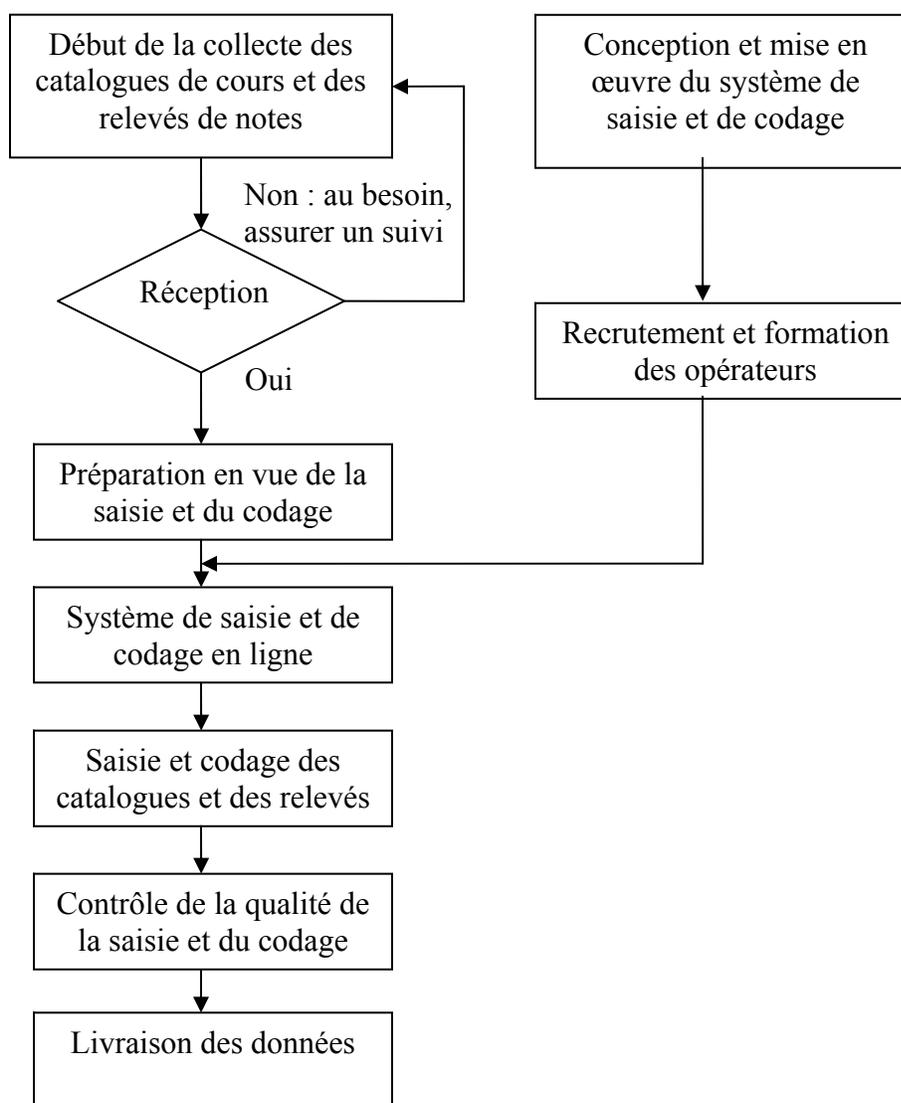
2.3 Préparation en vue de la saisie et du codage des données

En se préparant à la saisie et au codage des données des relevés de notes et des catalogues de cours, le personnel du projet a examiné les activités et les renseignements nécessaires à la conception du système. De plus, avant d'entrer des données dans le système, il a fallu coder chaque catalogue et chaque relevé selon un codage servant à classer et à regrouper les domaines de spécialisation et les cours. Ensuite, on a recruté des opérateurs spécialement formés pour extraire les données des catalogues de cours et des relevés de notes. Les opérateurs ont dû se familiariser avec des centaines d'éléments de données au niveau des écoles, des élèves et des cours. Puis, on a procédé au contrôle de la qualité de la saisie et du codage, activité essentielle. Enfin, le personnel du projet a dû produire de nombreux rapports complexes, notamment sur les affectations, la production et la qualité.

Tableau 2.3-1
Volume de la préparation des données

	Total	Pages par unité (moyenne)	Nombre total de pages
Relevés de notes	15 700	40	628 000
Catalogues de cours	700	150	105 000

2.4 Organigramme du système



3. Le système de saisie et de codage en ligne

3.1 Le système

Notre système de saisie et de codage en ligne constitue la pièce maîtresse de l'étude. Des opérateurs hautement qualifiés l'ont utilisé pour entrer des renseignements uniformisés dans la base de données. Les membres de l'équipe

de gestion de projet l'ont également utilisé quotidiennement pour suivre le déroulement des activités et résoudre les problèmes découverts grâce aux rapports sur le système en ligne.

3.1.1 Les caractéristiques du système

Vu la complexité de l'étude, nous avons conçu le système de saisie et de codage en ligne de manière à réduire les coûts en utilisant un langage de programmation Web et un serveur SQL communs. Pour les besoins de la conception et de la mise en œuvre, nous avons également créé des pages de saisie conviviales en les rendant nettes, simples et faciles à comprendre. Un autre aspect de la convivialité consistait à faciliter l'accès à la base de connaissances du système en dotant les pages de codage de documents de consultation, de listes déroulantes et de messages textuels utiles.

Pour maximiser l'efficacité du système en ligne, nous avons fait en sorte qu'une ouverture de session donne accès à trois composantes importantes du système : la saisie-codage, le contrôle de la qualité par l'utilisateur et le contrôle de la qualité par le gestionnaire. Le système orientait chaque utilisateur vers sa page de tâches en fonction du mode d'ouverture de chaque session.

Nous avons aussi accru l'efficacité du système en établissant au départ des données saisies et préchargées en vue des futures activités de saisie et de codage, ainsi qu'une gamme complète de contrôles pour réduire les erreurs de saisie. Pour gérer le déroulement des opérations, nous avons établi un ensemble complet de stades et d'états d'avancement ainsi qu'un ensemble complet de rapports papier et de rapports *ad hoc* à l'intention des gestionnaires et des opérateurs. De plus, nous avons déployé une gamme complète de déclencheurs pour gérer toutes les activités de la base de données et nous avons normalisé cette dernière pour réduire la redondance des données. Afin d'accroître la rapidité du système à mesure qu'on ajoutait des opérateurs et des gestionnaires, nous avons optimisé les interrogations et indexé les tableaux pour améliorer l'extraction et la mise à jour des données. L'équipe a également conçu et incorporé de nombreux écrans contextuels pour rendre le processus de saisie rapide et efficace. Le système effectuait des validations de la base de données afin d'éviter la saisie de données non valides et utilisait des états différents pour les relevés de notes et les catalogues afin d'optimiser l'attribution des différents rôles rattachés au processus de saisie. Enfin, le système intégrait le contrôle de la qualité de la saisie au processus de saisie et utilisait des algorithmes formulés pour sélectionner les cours en vue du contrôle de la qualité du codage.

3.1.2 L'efficacité du système

Au plus fort du déroulement de l'étude, jusqu'à 80 opérateurs utilisaient simultanément le système. De plus, les membres de l'équipe de gestion de projet étaient en ligne pour surveiller le système. Grâce aux outils fournis par le système, les opérateurs ont pu maintenir une productivité élevée. En outre, l'équipe de gestion de projet pouvait surveiller efficacement l'ensemble du système à l'aide des rapports produits par le système et d'autres outils. Par exemple, les gestionnaires pouvaient contrôler le nombre d'heures travaillées par un opérateur en une journée en fonction du travail accompli le même jour, pour découvrir si l'opérateur avait ou non un problème. Dans l'affirmative, l'opérateur recevait immédiatement de l'aide.

En raison du volume des données entrées, nous avons accordé une grande importance au contrôle de la qualité dès le début du processus de saisie. Notre système de saisie et de codage utilise tous les moyens nécessaires pour résoudre les problèmes de contrôle de la qualité, notamment la double saisie du contenu et la sélection automatique des opérations de vérification. En fin de compte, les opérateurs et les gestionnaires avaient l'assurance que les données entrées étaient de la plus haute qualité.

3.1.3 Le système de production de rapports

L'une des composantes essentielles du système en ligne était le système de production de rapports, programmé pour surveiller la totalité du processus de saisie et de codage et pour assurer la qualité du travail. Les rapports étaient conçus pour être produits le soir (rapports statiques) et de manière interactive (rapports dynamiques) pour aider les gestionnaires et les opérateurs à vérifier la qualité et la quantité des données entrées. Les gestionnaires ont utilisé les rapports pour suivre le travail et le nombre d'heures de chaque opérateur. Par exemple, nous avons mis au point un rapport sur le rendement des opérateurs pour aider les gestionnaires à respecter leur budget. Nous avons également

mis au point un rapport sommaire, produit chaque soir par un sous-programme imbriqué et comprenant plus de 200 mégaoctets de données, notamment sur la saisie et le codage des catalogues et des relevés de notes. Le personnel du projet utilisait des mots-clés pour accéder en mode descendant à l'information désirée. Pour les besoins du contrôle de la qualité, nous avons mis au point un rapport de contrôle de la qualité qui contre-vérfiait chaque jour plus d'un demi-million de champs de données dans la base de données pour déceler toutes différences.

4. Leçons tirées de l'expérience

Comme nous avons acquis une excellente connaissance de l'étude et des besoins du système, nous avons conçu un système robuste que nous avons pu mettre en œuvre avec succès. Comme dans le cas de tout système, nous avons eu notre part d'inconvénients et de contretemps. Notre plus gros problème tenait au rendement de la base de données. L'ajout d'opérateurs à la tâche et de dossiers aux tableaux a ralenti le système de saisie et de codage. Nous avons résolu ce problème en optimisant les interrogations et en indexant tous les tableaux. Avec le recul, nous avons compris que nous aurions dû le faire plus tôt pour prévenir le problème au départ.

5. Sommaire

La *Education Longitudinal Study* (ELS) est une étude multitâche et à plusieurs degrés, dont la collecte de relevés de notes et de catalogues de cours constituait un volet essentiel. Il a fallu recueillir un grand nombre de relevés de notes et de catalogues de cours auprès de diverses écoles, puis les saisir et les coder. Il a fallu, au préalable, planifier soigneusement la saisie et le codage des données pour assurer le bon déroulement du processus de saisie et de codage. Compte tenu de la complexité de l'étude, nous avons conçu le système de saisie et de codage en ligne pour qu'il soit convivial et peu coûteux. Il reposait sur une logique explicite, était facile à utiliser par les opérateurs et les gestionnaires et permettait le fonctionnement multitâche. Notre système de saisie et de codage nous a permis de bien gérer le déroulement de la saisie et du codage, le budget et la qualité des données entrées. Il a aussi permis à l'étude sur l'éducation de produire des données de haute qualité.