



L'Équipe

Chercheur principal

Chad Gaffield – *UOttawa*

Chefs d'équipe

Carl Amrhein – *Toronto/Alberta*

Peter Baskerville – *UVic*

Claude Bellavance – *UQTR*

Sean Cadigan – *MUN*

Gordon Darroch – *YorkU*

Charles Jones – *UToronto*

France Normand – *UQTR*

Evelyn Ruppert – *Trent/York*

Eric Sager – *UVic*

Marc St-Hilaire – *ULaval*

Lorne Tepperman – *UToronto*

Adam Green – *UOttawa*

Coordonnateurs/trices

Carmen Bauer – *UOttawa*

Nicola Farnworth – *YorkU*

Byron Moldofsky – *UToronto*

Terry Quinlan – *MUN*

Laurent Richard – *ULaval*

Doug Thompson – *UVic*

Martine Tremblay – *UQTR*

Universités partenaires

- University of Victoria
- York University
- University of Toronto
- Université du Québec à Trois-Rivières
- Université Laval
- Memorial University of Newfoundland
- Université d'Ottawa

Nos partenaires

- Statistique Canada
- Newfoundland & Labrador Statistics Agency
- Bibliothèque et Archives Canada
- IBM Canada
- Institut de la statistique du Québec
- International Microdata Access Group

TABLE DES MATIÈRES:

Message de la Coordinnatrice du projet 1

Annnonce de l'IRCS 1

La base de données contextuelles dans l'IRCS 2

Innovation de l'IRCS en matière de logiciels 3

Plans de sondage de l'IRCS 5

Événements de l'IRCS 6

VOLUME 1, NUMÉRO 4

PRINTEMPS 2007

Message de la Coordinnatrice du projet

Nous sommes heureux de vous présenter la quatrième édition du *Bulletin de l'IRCS*. Vous allez sans doute remarquer que le nombre de nouvelles a augmenté, signe de l'effervescence qui précède la mise en ligne prochaine de nos bases de données. Jusqu'à présent, nous avons complété la saisie des données des recensements des années 1911, 1921, 1931 et 1941. Le nettoyage et le codage des données du recensement de 1921 sont presque terminés, le traitement des données du recensement de 1951 a commencé et les préparatifs en vue du versement des données vers les Centres de données de recherche (CDR) sont bien amorcés. Comme dans les éditions précédentes du *Bulletin*, les membres de l'IRCS nous présentent leurs contributions au projet. Les chefs d'équipe Claude Bellavance et France Normand, et la coordonnatrice Martine Tremblay, tous trois de notre centre de l'UQTR, font le point sur les bases de données contextuelles. Ils nous expliquent comment ces données élargiront les possibilités offertes par l'IRCS. Par ailleurs, ce projet ne pourrait exister sans le concours de notre équipe informatique, dont le travail passe trop souvent inaperçu. Richard Smith, architecte en systèmes d'information pour l'IRCS, nous explique le logiciel SPIDER, notre programme-maison de saisie et d'échantillonnage de données. Enfin, Gordon Darroch, chef d'équipe à l'Université York et président du sous-groupe de l'échantillonnage, évoque les méthodes d'échantillonnage de l'IRCS.

J'espère que la lecture de ces articles vous intéressera. N'hésitez surtout pas à nous soumettre vos questions et suggestions. Et n'oubliez pas de consulter notre site régulièrement pour demeurer au fait des activités de l'IRCS.

Carmen Bauer
UOttawa

Annnonce de l'IRCS

À compter du mois de juin, Evelyn Ruppert sera en congé sabbatique de l'université de Trent. Elle occupera le poste de chercheuse principale invitée au Centre de recherche sur le changement socioculturel (CRSCS) de l'Open University (Milton Keynes, Angleterre). Elle y rejoindra son conjoint, Engin Isin (professeur et CRC à YorkU), qui vient d'obtenir une chaire sur la citoyenneté à l'Open University. Le CRSCS vise à développer des projets de recherche portant sur le changement culturel et ses implications économiques, sociales et politiques. Le centre agit en partenariat avec l'université de Manchester et implique des disciplines comme la sociologie, l'anthropologie sociale, la géographie, l'histoire et notre champ de recherche favori à l'IRCS... les statistiques du recensement! Evelyn travaillera sur son projet en cours, *Thinking and Making Population*. Elle travaillera également à établir des liens et des initiatives de collaboration sur des projets liés aux recensements avec des chercheurs du Royaume-Uni.



Chef d'équipe de l'IRCS, **Dr. Evelyn S. Ruppert**
Professeure adjointe, Département de sociologie
Trent University (Canada)
Chercheuse principale invitée, The Open University (R.-U.)

“Les données contextuelles représentent un matériau d’une richesse exceptionnelle pour comprendre les enjeux liés aux grands dénombremments de l’histoire contemporaine”

La base de données contextuelles dans l’IRCS : un outil pour mieux saisir les circonstances et les enjeux des recensements

Depuis quelques décennies, la valeur réelle des données des recensements a suscité bien des débats. Doit-on les aborder dans une perspective objectiviste ou considérer ces données comme un produit, une construction sociale?

L’équipe de l’IRCS a choisi, pour sa part, d’adopter une double approche, qui cherche à concilier la valeur des listes nominatives en tant que source d’informations sur un vaste éventail de phénomènes sociaux et culturels et l’organisation des recensements comme une initiative complexe, influencée par les grands enjeux politiques de l’heure, ainsi que par les rapports entre l’État et la société civile. Concrètement, cela nous a conduits à prendre en compte, outre les données

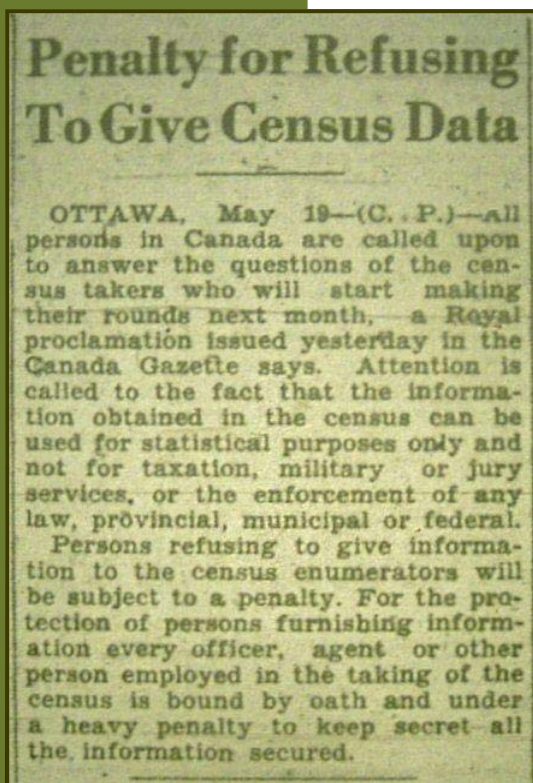
des listes nominatives, une foule de « données sur les données », plus communément appelées métadonnées. Les données contextuelles font partie de ces métadonnées. Mettant à profit une documentation variée, mais principalement la presse d’époque, nous avons pu recueillir des informations précieuses sur l’organisation, la tenue et la réception des recensements de 1911 à 1951 à travers l’ensemble du Canada.

La collecte des données contextuelles répond en pratique à une logique inverse de celle des sources sériées comme les listes nominatives de recensements. L’information se présente de

Le locuteur et le récepteur ne sauraient être négligés dans une stratégie de dépouillement adaptée à de tels types de sources. Qui parle? À qui? Et au nom de qui? Autant d’interrogations qui nous ont amenés à appréhender les médias et les institutions non pas comme de simples acteurs, mais en tant que véritables intermédiaires culturels, tour à tour porte-parole de l’État et de ses appareils, des groupes d’intérêts et des citoyens.

Pour mener à bien le traitement d’informations non sérielles, l’élaboration d’une grille thématique de dépouillement s’avère, dès le départ, une opération cruciale. Il en va non seulement de l’efficacité de la base de données en mode d’interrogation, mais aussi de la qualité et de l’homogénéité des informations recueillies, la grille servant de cadre de référence à l’équipe d’assistants chargée de relever les données en fonction de leur pertinence. La grille a ainsi été conçue pour permettre d’identifier le locuteur, de replacer temporellement l’information par rapport au recensement de même que pour situer le recensement national canadien par rapport à d’autres formes de dénombrement. Celle-ci intègre également les grands groupes de variables du recensement.

L’organisation de l’information textuelle et iconographique a nécessité l’élaboration de solutions informatiques novatrices. La base de données *ContextData* est exploitée en réseau selon une approche client/serveur, et un mode d’accès par fureteur Internet est opérationnel pour la consultation. L’interface privilégie la souplesse tout en offrant des outils assurant l’intégrité et la qualité de l’information colligée. Cette base a été conçue pour être exploitée de manière autonome pendant la période de construction de l’IRCS. À terme, cependant, l’information sera intégrée à la base de données principale, de sorte que le chercheur qui consultera les don-



The Montreal Daily Star, le 19 mai 1931, p. 33.

manière tout à fait disparate, aussi bien sur le plan de la forme que du contenu.

nées tirées des listes nominatives pourra s'informer sur leur contexte de production.

En somme, les données contextuelles représentent un matériau d'une richesse exceptionnelle pour comprendre les enjeux liés aux grands dénombrements de l'histoire contemporaine. Elles nous renseignent sur des questions aussi fondamentales que la construction de l'État, les mécanismes assurant sa légitimité, le rôle des intermédiaires culturels, etc. En même temps, elles viennent compléter les autres sources de « données sur les données » pour la validation des informations sérielles tirées des listes nominatives. Intégrées aux bases de données centrales de l'IRCS, elles seront également une sorte d'invitation au voyage pour le chercheur curieux de se replonger dans une époque aujourd'hui révolue.

**Claude Bellavance, France Normand
et Martine Tremblay**
CIEQ-UQTR

La Presse, le 12
novembre
1921, p. 11



Innovation de l'IRCS en matière de logiciels

À mesure que la qualité des documents historiques commence à se détériorer, les archivistes se tournent vers la technologie afin d'assurer la sauvegarde de leurs sources pour les générations à venir. La solution technologique au problème de sauvegarde associe généralement un nouveau support de données qui, à son tour, mène à de nouveaux outils de recherche. L'IRCS a élaboré un programme informatique novateur, surnommé SPIDER, qui tire pleinement parti de la dernière forme de sauvegarde des documents : l'imagerie numérique.

Au départ, l'IRCS devait obtenir ses résultats sur les recensements de 1911 à 1951 de microfilms que Statistique Canada avaient produits dans les années 1950 à partir de documents manuscrits. Lorsque Bibliothèque et Archives Canada a annoncé en 2003 un projet de numérisation des bobines du recensement de 1911, l'IRCS a émis une proposition de subvention de la numérisation de toutes les bobines des années qui entrent dans le cadre du projet. BAC a accepté l'offre de l'IRCS et débuté en septembre 2004 la conversion des bobines de 1911.

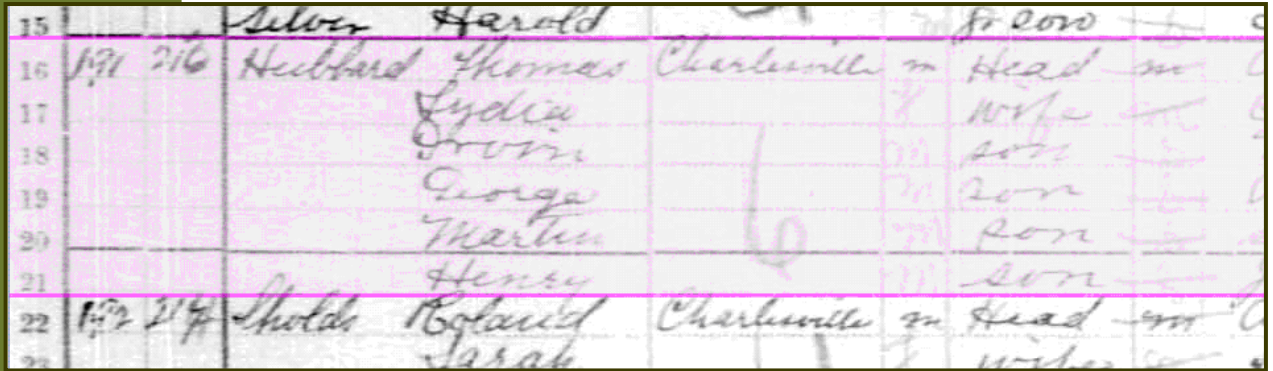
Quant au recensement de 1911, l'IRCS a adopté une méthode de saisie de données qui a permis à l'équipe de profiter immédiatement de l'utilisation d'images numériques. Un programme informatique de société indépendante a servi à visionner les documents manuscrits

du recensement et un second programme, élaboré à l'interne, a servi simultanément à entrer les données. La localisation de chaque image numérique a été beaucoup plus rapide que la recherche d'information équivalente sur microfilm. De surcroît, le programme d'affichage permet de modifier la couleur, la netteté et la taille des images, ce qui peut faciliter leur lecture considérablement.

Le programme SPIDER (« Sample Point Identification, Data Entry and Reporting » ou « programme d'identification des points échantillons, de saisie et de communication des données ») a été élaboré par l'IRCS pour les recensements de 1921 à 1951 en faisant fond sur les avantages de la méthode employée pour le recensement de 1911 tout en éliminant le goulot d'étranglement qui se trouve dans la plupart des systèmes fondés sur les microfilms : le besoin d'un nombre généré qui permet d'identifier un ménage sur un formulaire de recensement. Cet identificateur, appelé numéro repère, est obligatoire pour qu'un ménage destiné à l'échantillon par une personne (l'échantillonneur) puisse être localisé par une autre personne (opérateur de saisie de données) qui aurait à saisir les données sur ce ménage à une date ultérieure¹.

¹Il est possible d'éliminer le problème du numéro repère en faisant saisir les données à la personne qui effectue l'échantillonnage à mesure qu'elle relève les points échantillons. Cependant, cette méthode n'est pas pratique pour plusieurs raisons.

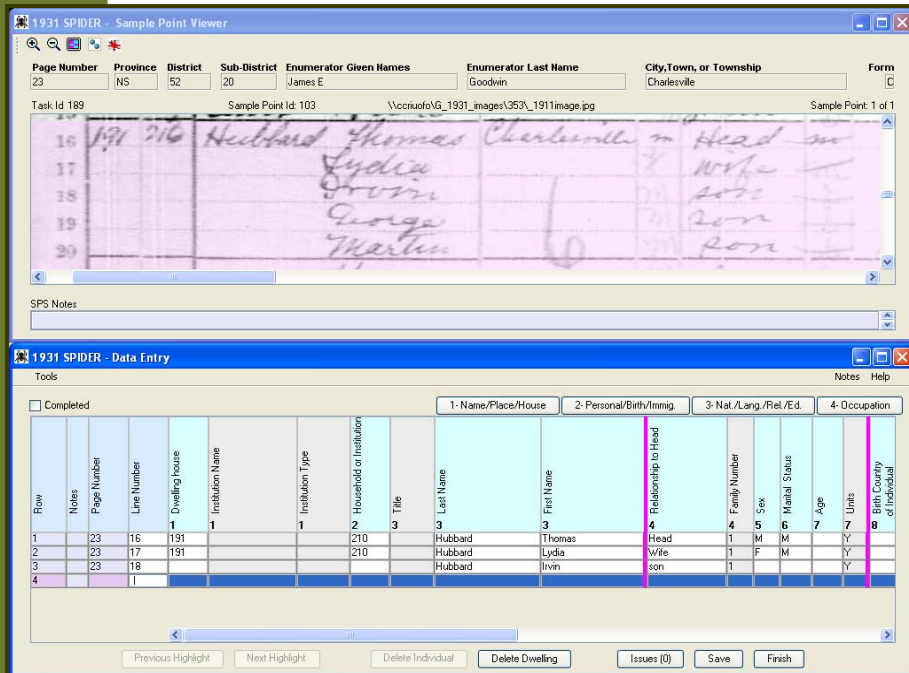
L'expérience a montré que la transcription et l'analyse du numéro repère sont non seulement chronophages mais aussi sujettes aux erreurs en raison des différentes numérotations qui peuvent apparaître sur les formulaires de recensement. La solution de SPIDER permet d'exploiter le potentiel du support : plutôt que de faire passer un numéro repère entre l'échantillonneur et l'opérateur, le logiciel fait passer une image numérique.



L'image représentée dans cet article provient d'un document du recensement de 1911 qui a été intégré au domaine public en 2003.

Figure 1 - La fonction de mise en évidence de SPIDER sert à marquer les rangées 16 à 21 de l'image destinée à l'échantillon.

L'échantillonneur utilise le module de visionnement de SPIDER afin d'afficher le formulaire de recensement numérisé. Utilisant la fonction de mise en évidence incorporée au programme, l'échantillonneur marque sur l'image les individus qu'il faut inclure dans l'échantillon (figure 1). Plus tard, le module de saisie de données de SPIDER présentera la même image marquée (figure 2) à un opérateur qui peut alors rapidement, et avec précision, repérer les données à saisir.



Le système SPIDER permet d'exploiter le potentiel de l'imagerie numérique. En incluant les caractéristiques de l'image dans sa définition des points échantillons et en intégrant les tâches de sélection et de saisie, l'IRCS a élaboré un nouvel outil qui simplifie considérablement le processus de saisie des résultats de recensements.

Figure 2 -Module de saisie de données de SPIDER. Il faut saisir dans le tableau de la fenêtre inférieure les données mises en évidence dans la fenêtre supérieure.

Rick Smith
UOttawa

Plans de sondage de l'IRCS

Le plan de sondage décrit ici est le résultat de consultations menées dans un sous-groupe de l'équipe de l'IRCS, qui comprend Michael Ornstein, directeur du York Institute for Social Research

Les plans de sondage de l'IRCS visent à fournir de grands échantillons, représentatifs à l'échelle nationale et facilement accessibles, de chacun des recensements réalisés de 1911 à 1951. Les résultats des recensements de 1921 à aujourd'hui étant strictement confidentiels, les échantillons ne seront disponibles que dans un format qui empêche l'identification des personnes. Les analyses plus détaillées ne seront autorisées que dans les centres de

Statistique Canada. La définition de famille statistique, de ménage et de logement varie d'un recensement à l'autre, mais nous permet tout de même de prélever des échantillons comparables. Nous échantillonons les logements et leurs membres. Plusieurs des analyses intéressantes établissent un rapport entre les caractéristiques du logement, d'une part, et les familles et les personnes qui y résident, d'autre part. On prélève deux échantillons parallèles pour chaque année de recensement, un premier composé de logements de « taille normale » et un second, spécialisé, formé de « logements collectifs », la plupart étant des institutions telles que les hôpitaux, les asiles ou les orphelinats, ou encore des camps de travail. Par définition, les logements de taille normale comptent 30 résidents ou moins et les logements collectifs comptent 31 résidents ou plus. Bien qu'arbitraire, cette taille minimum sert à séparer les logements occupés par des familles maitresses et les résidents qui leurs sont associés (parents ou pensionnaires) des endroits occupés par des personnes sans lien de parenté. L'IRCS prélève des échantillons suivant la méthode de l'équiprobabilité. Elle les prélève en choisissant une origine au hasard dans les régions géographiques représentées par les bobines de microfilms originales. Bien que les micro-données sur les logements et les personnes aient été converties en images numériques pour nos besoins, cette stratification géographique efficace a tendance à améliorer la représentativité géographique des échantillons à l'échelle nationale, comparativement aux échantillons de logements prélevés au hasard. Dans chaque région géographique, une sélection systématique des logements suit l'origine choisie au hasard de manière à produire la densité d'échantillon souhaitée dans les enregistrements. Puisque la population canadienne croît pendant ces années et que la proportion de la population choisie pour l'analyse est inférieure à la proportion que représente la taille réelle de l'échantillon, nous prélevons des échantillons composés à 5 p. 100 de logements (et donc de

personnes) en 1911, à 4 p. 100 en 1921 et à 3 p. 100 en 1931, 1941 et 1951. Cette méthode produit des échantillons d'environ 320 000 à 420 000 individus pour chacune des cinq années de recensement.

Les enregistrements de toutes les personnes résidant dans un logement choisi de taille normale sont transcrits fidèlement. En principe, cette méthode génère des « échantillons en grappes » des personnes et des familles qui résident dans les logements (puisque l'échantillon inclut l'ensemble des personnes et des familles de tous les logements prélevés). À certaines fins, les estimations d'intervalles de confiance et d'erreurs types doivent tenir compte des effets de tels regroupements. Il est possible d'y arriver facilement avec certains logiciels d'analyse statistique, mais pas tous. Dans bien des analyses, les échantillons sont si grands que l'agrégation statistique ne préoccupera pas beaucoup.

La façon dont nous traitons le nombre relativement petit de « logements collectifs » est une caractéristique inhabituelle des plans de sondage. La solution conventionnelle à ce problème consiste à réaliser un échantillon séparé de cas inhabituels pour réduire leurs effets sur les estimations des caractéristiques de la population. Notre plan y arrive grâce à un logiciel qui nous permet d'examiner tous les logements au pays et de dresser l'inventaire de ceux qui comptent plus de 30 résidents. Nous échantillonons les unités d'habitation ou les personnes au sein de ces logements collectifs dans des proportions plus grandes que dans l'échantillon principal. Cette méthode donne de rares occasions d'examiner en détail les logements institutionnels et les logements collectifs qui y sont liés. Elle permet d'examiner par exemple la croissance et l'évolution de la composition des populations placées en établissements. L'analyse de ces résultats est régie par les exigences strictes de Statistique Canada en matière de confidentialité. L'IRCS fournira le nettoyage et le codage intégrés et documentés des micro-données de tous les échantillons. Les nouveaux échantillons pourront ainsi s'intégrer dans les échantillons existants du XIX^e et du début du XX^e siècles et dans ceux que Statistique Canada a publiés depuis 1971.

Ensemble, ces deux formes d'échantillon produiront des interprétations nouvelles et révisées des années de formation du Canada dans la première moitié du XX^e siècle. L'IRCS encouragera aussi l'analyse comparative dans un nombre croissant de pays qui réalisent des bases de données similaires, et notamment avec l'IPUMS (Integrated Public Use Microdata Series) des États-Unis¹.

Gordon Darroch
York University

¹Consultez le web site de l'University of Minnesota Population Center: <http://www.ipums.umn.edu/>



INFRASTRUCTURE DE RECHERCHE SUR LE CANADA AU 20^e SIECLE

Institut d'Études canadiennes
Université d'Ottawa
52 Université
Ottawa, Ontario Canada K1N 6N5

Conception graphique et rédaction

Mirela Matiu

Adjointe du projet de l'IRCS

Téléphone: 613-562-5800 poste 3657

Télécopieur: 613-562-5216

Courriel: mmatiu@uottawa.ca

www.canada.uottawa.ca/ccri/



Agence universitaire de la Francophonie

chercheurs en sciences sociales dont les travaux portent sur les dynamiques démographiques en privilégiant une dimension historique. Plus d'une centaine de participants provenant d'une trentaine de pays différents participeront aux divers ateliers.

L'une des séances portera sur «Les expériences québécoises et canadiennes» et sera présidée par Rosemary Bender, Directrice Générale, Statistique sociale et démographique, Statistique Canada. Chad Gaffield, président du Conseil de recherche en sciences humaines du Canada (CRSH) et chercheur principal de l'IRCS, sera le conférencier invité pour l'ouverture de cette séance. Marc St-Hilaire présentera une des communications et un kiosque de l'IRCS accueillera les participants désirant obtenir plus d'informations au sujet du projet. L'organisation de cette rencontre scientifique est sous la responsabilité de Richard Marcoux, Coordonnateur du Réseau Démographie de l'AUF et Professeur titulaire au Département de sociologie de l'Université Laval.

<http://www.demographie.auf.org/>

Événements de l'IRCS

La conférence *Health in Families, Healthy Families: Gendered Explorations* a été organisée par l'International Sociological Association Research Committee on Family Studies. La conférence s'est tenue à Toronto du 8 au 11 mai. Lorne Tepperman (chef d'équipe de l'IRCS, University of Toronto) a été coorganisateur de cette conférence. Le 9 mai de 11:45 AM à 1:00 P.M., a eu une session sur les recensements au Canada et les études sur la famille. Cette session a été organisée par Peter Baskerville (chef d'équipe de l'IRCS, University of Victoria), et a été présidée par Dr. Gordon Darroch (chef d'équipe de l'IRCS, York University).

Les présentateurs pour cette session ont été :

- Dr. Chad Gaffield

Titre de la présentation : *The Canadian Century Research Infrastructure & Canadian Research*

- Terry Quinlan

Titre de la présentation : *Using Census Records to Advance Genetic Research*

- Dr. Peter Baskerville

Titre de la présentation : *The changing material welfare of women in urban Canada in the late 19th century*

Les 7^{es} Journées scientifiques du Réseau Démographie de l'Agence universitaire de la francophonie (AUF) se tiendront à l'Université Laval du 19 au 22 juin 2007. Elles ont pour thème «Mémoires et démographie. Regards croisés au Sud et au Nord». L'objectif de cette rencontre est de créer un espace de réflexion et d'échange afin de proposer des lectures différentes et plurielles des transformations sociales et démographiques. Les 7^{èmes} JS s'adressent à tous les intervenants et

