

**Conférence
nord-américaine sur
l'Initiative de
documentation des
données
2019**

*Avantages de décrire la
production statistique et les
variables*

Du 24 au 26 avril 2019

Statistique Canada

Ottawa (Ontario)



Statistique
Canada

Statistics
Canada

Canada

Conférence nord-américaine sur l'Initiative de documentation des données

La conférence nord-américaine sur l'Initiative de documentation des données (*North American Data Documentation Initiative, NADDI*) est l'occasion pour les utilisateurs de normes de métadonnées et les personnes souhaitant en savoir plus de se réunir et d'apprendre les uns des autres. Inspirée de la Conférence européenne des utilisateurs de la DDI, laquelle a remporté beaucoup de succès, la NADDI est une conférence de deux jours proposant des communications sollicitées et des communications libres; elle devrait intéresser à la fois les chercheurs et les spécialistes des données du domaine des sciences sociales et d'autres domaines. Une journée entière de séances de formation précède la conférence.

Le thème de la conférence NADDI 2019, Les avantages de décrire la production statistique et les variables, met l'accent sur les avantages de l'utilisation de métadonnées pour accroître l'efficacité du cycle de vie des données de recherche et promouvoir la réutilisation ultérieure de produits de données finaux.

Initiative de documentation des données



L'Initiative de documentation de données (*Data Documentation Initiative, DDI*) est une norme ouverte relative aux métadonnées qui est employée pour décrire des données et préciser des activités de collecte de données. Le principal objectif de la DDI est de permettre le traitement automatique de métadonnées de recherche. La spécification peut permettre de documenter et de gérer différentes étapes de cycles de vie des données, comme la conceptualisation, la collecte, le traitement, l'analyse, la distribution, la découverte, la réorientation et l'archivage.

Organisme hôte



Statistique
Canada

Statistics
Canada

L'organisme hôte de la conférence de cette année est Statistique Canada, l'organisme national de la statistique du Canada. Ce dernier veille à fournir aux Canadiens tous les renseignements clés sur l'économie, la société et l'environnement du pays dont ils ont besoin pour assumer efficacement leur rôle de citoyens et de décideurs.

Mercredi 24 avril 2019

Atelier—salle de conférences du Jean-Talon

De 8 h à 9 h

Café et déjeuner

De 9 h à midi

Atelier 1

Président de l'atelier : Stéphane Crête, Statistique Canada

Présentation de la DDI

Formateur : Jane Fry, Université Carleton

Savoir ce qu'est la DDI et comprendre son fonctionnement ainsi que la façon de l'utiliser peut être intimidant, particulièrement si vous n'êtes pas programmeur ni codeur. La DDI peut être un schéma particulièrement puissant pour les sciences sociales, mais en comprendre les bases est une condition préalable pour apprendre à s'en servir.

Cet atelier vous présentera le contexte de la DDI, les différents outils utilisés pour l'interpréter et son intégration au déroulement du cycle de vie des données. Des exercices vous aideront à comprendre les différentes facettes de la DDI, vous permettant ainsi de constater son potentiel et sa puissance.

Il ne s'agit pas d'un atelier technique; cet atelier s'adresse aux personnes n'étant pas familières avec la DDI. Cependant, il peut également vous être utile si vous souhaitez mettre à niveau vos connaissances sur l'utilité et le fonctionnement de la DDI.

De midi à 13 h

Dîner

Veuillez prendre note que le dîner n'est pas fourni par l'hôte.

De 13 h à 16 h

Atelier 2

Documenter des questionnaires et des ensembles de données grâce à la DDI : présentation pratique de Colectica

Formateurs : Jeremy Iverson et Dan Smith, Colectica

Cet atelier offre une approche pratique de création et de documentation d'enquêtes et d'ensembles de données à l'aide de la DDI et de Colectica. Les participants élaboreront et appliqueront sur le terrain une enquête fondée sur la DDI à l'aide de leurs propres questions ou au moyen d'échantillons fournis dans le cadre de l'atelier. Ensuite, ils assimileront, annoteront et publieront les descriptions d'ensembles de données DDI à l'aide des données d'enquête recueillies.

Les participants pourront apporter leur propre ordinateur portable Windows, afin de participer aux exercices pratiques.

De 18 h à 20 h

Rencontre informelle

Restaurant Johnny Farina
216, rue Elgin, Ottawa (Ontario) K2P 1L6

Jeudi 25 avril 2019

Première journée—salle de conférences du Jean-Talon

De 8 h à 9 h

Inscription—café et déjeuner

De 9 h à 10 h 30

Bienvenue

Statistique Canada

Kathryn Stevenson, Statistique Canada

Mot de bienvenue de la DDI Alliance et présentation par l'animateur principal

Jared Lyle, directeur, DDI Alliance, Inter-university Consortium for Political and Social Research (ICPSR)

Discours principal : Les normes de métadonnées à l'appui de la modernisation des statistiques officielles

Anil Arora, statisticien en chef du Canada

Monsieur Arora a mené d'importantes initiatives transformationnelles tout au long de sa carrière, ayant acquis de l'expérience et établi des partenariats auprès des trois ordres de gouvernement, du secteur privé et d'organismes internationaux, notamment les Nations Unies et l'Organisation de coopération et de développement économiques. Il a dirigé des projets relatifs à d'importants enjeux de politique ainsi qu'à des réformes législatives et réglementaires et a supervisé de vastes programmes nationaux. Anil Arora a été nommé statisticien en chef du Canada en septembre 2016.

Analyse des données en tant que service

Robert McLellan, Statistique Canada

De 10 h 30 à 11 h

Pause

De 11 h à 12 h 20

Première séance

Président de séance : Dan Smith, Colectica

Démonstration de l'évaluation de la conférence sur la DDI

Barry T. Radler, Université du Wisconsin à Madison

Utilisation de la DDI pour documenter les microdonnées à grande diffusion annuelles pour les enquêtes sur les dépenses des consommateurs

Dan Gillman, Bureau de la statistique du travail des États-Unis

Le Bureau de la statistique du travail des États-Unis a recours à la norme DDI pour documenter les enquêtes sur les dépenses des consommateurs. Dans le cadre de ces enquêtes, des fichiers de microdonnées à grande diffusion sont produits

chaque année. L'objectif immédiat du déploiement du système est de permettre aux utilisateurs de relier des variables des fichiers de microdonnées à grande diffusion à leurs questions. La base de données autonome actuelle contenant les données relatives à toutes les variables est incomplète, et il n'est pas simple d'effectuer le suivi des modifications apportées aux codes, aux définitions, aux variables et aux fichiers au fil du temps et pour toutes les enquêtes. Notre plan est d'intégrer des fonctions d'affichage de la correspondance entre les variables et les questions et d'effectuer le suivi des modifications. De plus, nous souhaitons enregistrer des documents méthodologiques, tels que des questionnaires, des pondérations et des plans d'échantillonnage. Dans la version du système de métadonnées actuellement en cours d'élaboration, nous établissons la correspondance entre les variables des fichiers de microdonnées à grande diffusion de 2017 et leurs questions, lesquelles leur fournissent entièrement ou partiellement des données d'entrée. Dans cet exposé, nous décrivons l'effort d'élaboration en soulignant en particulier la mise en correspondance des variables et des questions. Dans de nombreuses discussions sur les modèles de métadonnées, on ne fait que survoler la mise en correspondance des variables et des questions, ce qui laisse entendre que cette opération est simple. Il en est autrement selon notre expérience, et nous décrivons ici les difficultés rencontrées.

Utilisation de la DDI pour soutenir la conservation de données dans Dataverse

Amber Leahey, Scholars Portal, Conseil des bibliothèques universitaires de l'Ontario

Dataverse est une plateforme de répertoire de données ouvertes élaborée par l'Institute for Quantitative Social Science de l'université Harvard (comptant des contributeurs et adopteurs dans le monde entier). Dataverse est utile aux chercheurs pour la publication de données ouvertes et médiatisées, le partage de données, la promotion de la recherche et la réutilisation de données d'autrui. La plateforme Dataverse facilite le dépôt et la publication de données de recherche; cependant, des données de recherche mises en dépôts nécessitent souvent plus de préparation et de conservation pour pouvoir être réutilisées de façon utile. Les données doivent être accompagnées de documents et de métadonnées exacts et riches; sans quoi elles sont dénuées de sens. Des métadonnées normalisées, comme c'est le cas de la DDI, sont utiles pour décrire les données de Dataverse. Cet exposé vise à montrer la façon dont Dataverse prend en charge la création de métadonnées pour améliorer la découverte de données, leur accès et leur réutilisation ainsi que la préservation des données de recherche dans Dataverse.

Création de la documentation DDI Lifecycle pour les données longitudinales à l'Inter-university Consortium for Political and Social Research (ICPSR)
Sanda Ionescu, Inter-university Consortium for Political and Social Research

Conformément à l'engagement de l'ICPSR d'adopter DDI Lifecycle pour documenter certaines de ses études longitudinales, nous avons récemment entrepris un nouveau projet contribuant à créer des métadonnées DDI 3.2 pour l'une de nos collections les plus populaires du projet national sur la vie sociale, la santé et le vieillissement, qui est gérée et distribuée par notre archive nationale de données informatiques sur le vieillissement. Nous présenterons brièvement cette collection et fournirons des précisions sur les étapes suivies pour migrer ses métadonnées vers DDI Lifecycle à l'aide de Colectica Designer et les rendre publiques sur le portail de Colectica. Nous comptons nous concentrer sur des défis particuliers qui se sont posés au cours de ce processus et les solutions pratiques permettant de les surmonter. Nous souhaitons également présenter certains des avantages que comporte l'utilisation de DDI-L, comme la comparabilité et l'utilité améliorées des données.

De 12 h 20 à 13 h 30

Dîner

De 13 h 30 à 15 h

Deuxième séance

Président de séance : Barry T. Radler, Université du Wisconsin à Madison

Résolution pour X(XML) : transformer des métadonnées pour transformer l'accès aux données

John Huck, bibliothèques de l'Université de l'Alberta

À l'Université de l'Alberta, l'inscription à des ressources de données du Linguistic Data Consortium (LDC) présentait des difficultés de catalogage pour le personnel des bibliothèques et d'accès pour les utilisateurs : les ressources étaient publiées en formats numérique et physique, mais l'accès numérique était difficile à cataloguer, et certains ensembles de données étaient uniquement disponibles en format physique. Par conséquent, aucun emplacement unique ne permettait de découvrir les ensembles de données achetés par la bibliothèque, et accéder aux données nécessitait un processus de récupération des données différent pour chaque format.

Cet exposé décrit un projet entrepris par le personnel de trois unités de bibliothèque visant à élaborer un modèle d'accès simplifié à des ressources, lequel devrait être lancé en 2019. Des métadonnées du LDC disponibles auprès d'un hôte OAI-PMH mis à jour activement ont établi la base d'un flux de travail transformationnel, afin de générer des enregistrements MARC à feuilles de style XSLT personnalisées. Cela a permis à l'équipe de réfléchir aux objectifs des utilisateurs plutôt que de cataloguer les flux de travail, et d'imaginer un modèle d'accès selon lequel les utilisateurs n'auraient pas

besoin de renseignements sur le format des données. Dans le cadre de ce nouveau modèle d'accès, chaque enregistrement MARC comprend un lien vers un formulaire de demande de récupération. Le personnel de bibliothèque choisit le meilleur format d'extraction et livre les données par voie électronique aux utilisateurs autorisés.

metajelo + DDI : Dérivation d'un ensemble de métadonnées pour que des revues puissent prendre en charge des objets couplés en externe à partir de la DDI

Lars Vilhuber, Institut de la dynamique du travail, Université Cornell

Nous avons proposé un ensemble de métadonnées (appelé « metajelo ») visant à fournir aux revues scientifiques un moyen facile d'enregistrer l'existence et la disponibilité de contenu complémentaire au moment de la publication. Les renseignements relatifs à ce contenu complémentaire sont, dans la plupart des cas, essentiels à la reproductibilité de résultats d'études. Dans de nombreux cas, ce contenu est traité par une tierce partie qui peut ou non suivre des normes d'élaboration pour déterminer et décrire ce contenu. Les chercheurs rencontrent des difficultés lorsqu'ils essaient de se conformer pleinement aux normes relatives à la documentation des données et à leur provenance.

Cependant, bon nombre des éléments requis figurent dans la DDI et lorsque des données y sont entrées correctement par des administrateurs de données utilisant la DDI, la génération de l'ensemble metajelo est simple. Dans cet exposé, nous décrivons l'intérêt de metajelo et l'aisance avec laquelle des archives ayant recours à la DDI fournissent aux chercheurs un ensemble de métadonnées compact qui permet d'améliorer la reproductibilité tout en réduisant les efforts des chercheurs.

Paramètres de plans d'échantillonnage complexes : rapprocher la documentation technique et l'utilisation de logiciels de recherche appliqués

Stanislav Kolenikov, Abt Associates

Ces travaux rendent compte de notre compréhension en matière de définition de paramètres de plans d'échantillonnage complexes dans les logiciels statistiques utilisés pour analyser des données d'enquêtes fondées sur le plan. Ils consignent également les pratiques exemplaires que nous recommandons en la matière. Les fonctionnalités de données d'enquête complexes, comme la stratification, la catégorisation des résultats, les probabilités inégales de sélection et le calage, y sont passées en revue. Nous soulignons leur incidence sur des procédures d'estimation et démontrons les dangers de toute analyse incorrecte qui n'en tient pas compte. Nous

montrons de quelle façon un logiciel statistique gère ces fonctionnalités. Nous analysons la documentation accompagnant plusieurs enquêtes recueillies et distribuées par des bureaux nationaux de statistique, des organismes statistiques et des organisations universitaires et évaluons la facilité ou la difficulté de préciser correctement le plan d'échantillonnage dans le logiciel en fonction de la documentation fournie sur les ensembles de données. Nous concluons par nos recommandations aux fournisseurs de données d'enquête quant à la façon de créer de la documentation transparente, exacte et efficace dont ont besoin les utilisateurs de données d'enquête pour tenir compte des caractéristiques du plan d'échantillonnage dans leurs analyses de données d'enquête complexes. Les autres coauteurs de ces travaux sont Brady T. West (Université du Michigan) et Peter Lugtig (université d'Utrecht).

Utilisation de la DDI et des métadonnées dans l'analyse des travaux de recherche, la production de rapports de recherche et la diffusion de ces travaux de recherche : cas d'utilisation pour l'élaboration d'algorithmes de prédiction du risque dans le domaine de la santé

Doug Manuel, Statistique Canada

Malgré la nécessité de travaux de recherche ouverts et reproductibles, les métadonnées de la DDI survivent rarement à la diffusion des travaux de recherche.

Nous décrivons un flux de travail de métadonnées de la DDI pour les analyses et la diffusion de travaux de recherche au moyen du cas d'utilisation d'algorithmes prédictifs reposant sur l'Enquête sur la santé dans les collectivités canadiennes de Statistique Canada (ESCC) couplée aux registres longitudinaux de mortalité et de maladie. Ce flux de travail commence par une feuille de travail conforme à la DDI permettant aux chercheurs de collaborer entre établissements pour préciser à l'avance des paramètres du modèle prédictif. Des fonctions routinières nettoient et transforment la cohorte à l'étude en fonction des spécifications de la feuille de travail, en ajoutant des métadonnées au document de la DDI de la cohorte à l'étude. Les métadonnées de la DDI sont alors réutilisées dans un éventail de contextes de diffusion, y compris des tableaux manuscrits et des outils de visualisation d'algorithmes Web. Le document de la DDI est alors publié. Les modèles sont décrits à l'aide du langage PMML (de balisage de modélisation prédictive), un schéma XML complémentaire. Les algorithmes sont déployés à l'aide de moteurs de notation DDI, PMML et TensorFlow.

De 15 h à 15 h 15

Pause

Élaboration d'un flux de travail afin de saisir, d'améliorer et de diffuser des métadonnées d'étude longitudinale

Jon Johnson, UK Data Archive

CLOSER rassemble huit études longitudinales britanniques de calibre mondial, afin de maximiser leur utilisation, leur valeur et leur incidence. Un produit important est CLOSER Discovery, qui permet aux utilisateurs de rechercher et de parcourir des métadonnées de questionnaires et d'ensemble de données. La gestion efficace de données d'études longitudinales complexes est à la fois souhaitable et de plus en plus essentielle. Des normes relatives aux métadonnées sont vitales pour conserver ces renseignements tout au long du cycle de vie des données. De nombreuses études à long terme font face à un arriéré historique les empêchant de passer aux meilleures normes en matière de métadonnées. Le projet CLOSER a reçu un important financement du Conseil de la recherche économique et sociale ainsi que du Conseil de la recherche médicale pour ces huit études, qui forment la base d'une ressource de recherche durable de haute qualité.

CLOSER a permis d'élaborer une série d'outils et de logiciels à partir de solutions internes et disponibles sur le marché, grâce auxquels on commence à surmonter certains des obstacles que comportent la consignation et l'utilisation de métadonnées longitudinales. Cette communication mentionnera les réussites et les difficultés rencontrées lors de l'utilisation de la norme de métadonnées DDI Lifecycle pour réaliser ces projets ambitieux. Elle se déroulera sous la forme de trois exposés sur :

- la saisie et l'annotation de métadonnées de questionnaire (Charles De Barros, CLOSER);
- la mise en correspondance de questions, de variables, de mots clés et de concepts (Charles De Barros, CLOSER);
- une démonstration de la plus récente fonctionnalité à la disposition de la communauté des chercheurs (Jon Johnson, CLOSER).

Utilisation de la DDI pour documenter une étude longitudinale complexe

Barry T. Radler, Université du Wisconsin à Madison

Des métadonnées de recherche complètes aident grandement à clarifier les méthodes et les processus utilisés pour saisir des données et produire des ensembles de données. Ce faisant, elles fournissent aux utilisateurs de données les renseignements nécessaires pour mieux analyser, interpréter, conserver et partager ces données. Les métadonnées à structure riche sont encore plus importantes, étant donné que les études longitudinales complexes comportent

des milliers de variables, différents types de données et de nombreuses vagues de collecte de données. MIDUS est une étude longitudinale nationale menée auprès d'environ 12 000 États-Uniens, qui porte sur le vieillissement en tant que processus biopsychosocial intégré. MIDUS comprend une vaste et unique combinaison de données de recherche d'enquête, expérimentales et de laboratoire, recueillies pendant 20 ans de différentes façons. Au cours de la dernière décennie, MIDUS s'est de plus en plus appuyée sur la norme DDI pour gérer ces données de recherche complexes et permettre leur traitement automatique dans un format convivial. Plus récemment, MIDUS a eu recours à des outils Colectica, afin d'améliorer l'infrastructure de la DDI et de créer un système harmonisé d'extraction de données fondé sur la DDI. Un tel système permet aux chercheurs de rechercher des variables d'intérêt au sein d'ensembles de données, de recenser et d'harmoniser des versions longitudinales apparentées de variables et de créer facilement des exportations de données personnalisées.

De 16 h 15 à 16 h 45

Quatrième séance

Président de séance : James Doiron, bibliothèques de l'Université de l'Alberta

Engagement communautaire de la DDI : discussion et commentaires

William Block, Cornell Institute for Social and Economic Research, université Cornell

Barry T. Radler, Université du Wisconsin à Madison

Jared Lyle, directeur, DDI Alliance, Inter-university Consortium for Political and Social Research

De 16 h 45 à 17 h

Synthèse

James Doiron, bibliothèques de l'Université de l'Alberta

De 19 h à 21 h

Banquet

Bier Markt Ottawa

156, rue Sparks, Ottawa (Ontario) K1P 5C1

Vendredi 26 avril 2019

Deuxième journée—salle de conférences du Jean-Talon

De 8 h à 9 h

Café et déjeuner

De 8 h 30 à 9 h 30

Séance de présentation par affiche

Président de séance : Flavio Rizzolo, Statistique Canada

Guide de pratiques exemplaires sur les métadonnées Dataverse North

Amber Leahey, Alexandra Cooper, Martine Gagnon et coll.

L'un des avantages d'utiliser Dataverse est qu'il fournit un ensemble complet de métadonnées permettant de décrire des données de recherche propres à un projet et à un ensemble de données. Basé sur une combinaison de normes Dublin Core et DDI, Dataverse fournit un ensemble de métadonnées bien adapté aux sciences sociales, aux sciences du comportement, aux sciences économiques et aux sciences de la santé. Cependant, les chercheurs ne connaissent pas tous nécessairement les champs de métadonnées à utiliser ou la meilleure façon de les interpréter. Cette affiche présente le guide de pratiques exemplaires sur les métadonnées Dataverse North, un guide d'introduction pour les utilisateurs novices à intermédiaires de la plateforme de répertoire de données Dataverse, lequel fournit une orientation quant à la façon de décrire des ensembles de données de recherche à l'aide du champ de métadonnées standard dans Dataverse.

Étude pratique de la tendance méthodologique de DDI 4

Dan Gillman, Bureau de la statistique du travail des États-Unis

La structure méthodologique de la DDI 4 est une structure générale conçue pour tenir compte des descriptions de la méthodologie de l'étude à n'importe quel niveau de détail. Cette structure tient compte de deux points indépendants : 1) elle distingue la conception (le quoi), l'algorithme (le comment théorique) et le processus (le comment pratique); 2) elle permet une description récursive en sections plus détaillées de la cascade de traitement. Distinguer la conception, l'algorithme et le processus permet la réutilisation et fournit une riche plage descriptive. La distinction entre un algorithme à critère d'arrêt non défini et un processus comportant une limite précise, par exemple, est manifeste. De plus, une conception se distingue nettement de l'algorithme. De nombreux algorithmes peuvent satisfaire à la même conception de base. À l'aide de l'exemple des enquêtes états-uniennes sur les dépenses de consommateurs, nous illustrons cette approche à plusieurs niveaux. Le traitement de ces enquêtes a lieu en quatre étapes ou sous-systèmes distincts. Chaque étape peut être généralement décrite à un niveau général, à un niveau moyen en décrivant de façon générale chaque sous-processus et à un niveau inférieur en fournissant des renseignements détaillés sur chaque sous-processus étape par étape. Nous illustrons également les raisons pour

lesquelles les descriptions du niveau inférieur peuvent ne pas respecter les exigences en matière de confidentialité.

Maintenir l'activité des métadonnées pendant les analyses de la recherche sur la santé et la production de rapports de recherche sur la santé : utilisation d'un impératif de science ouverte pour mettre à jour l'approche actuelle d'analyse limitée de métadonnées

Doug Manuel, Statistique Canada

Mise en correspondance de DDI 2 et de DDI 4

Larry Hoyle, Institute for Policy and Social Research, Université du Kansas

Joachim Wackerow, GESIS, Institut Leibniz pour les sciences sociales

Cette affiche décrit les efforts entrepris pour ajouter une fonction d'importation DDI-Codebook (DDI-C) au progiciel DDI 4 R.

Le DDI 4 Codebook Group a effectué une grande partie de la modélisation de l'une des sections de DDI 4 à l'aide d'un tableur établissant la correspondance entre des éléments de DDI-C et des propriétés DDI 4. Ce processus a commencé par une liste d'éléments utilisés par le CESSDA (Consortium d'archives européennes de données de sciences sociales) et a été raffiné au « sprint » de Norvège à Knutholmen en mai 2016. Malheureusement, ces correspondances n'étaient pas toujours au niveau du nœud terminal.

Un programme R a également importé DDI-C XML de l'Enquête sociale européenne et a généré une liste de XPath uniques d'éléments terminaux utilisés dans cet ensemble de métadonnées. Ces éléments, ainsi que les chemins terminaux DDI 4 correspondants, ont été utilisés pour mettre le tableur à jour.

Ce tableur a ensuite été amélioré de nouveau pour créer un tableau exploitable de mise en correspondance entre des valeurs terminales DDI-C et des propriétés terminales dans DDI 4. La rédaction du code permettant d'importer DDI-C a nécessité les renseignements supplémentaires suivants :

- la mise en correspondance des sous-chemins DDI-C avec les catégories DDI 4 identifiables (p. ex. correspondance de tous les renseignements pour une variable DDI-C correspondant à une variable DDI 4 identifiable);
- la mise en correspondance des catégories de cibles abstraites avec des extensions particulières;
- d'autres valeurs de propriétés sémantiques, telles que le type de méthodologie.

Importer DDI-C dans une version de DDI au niveau du cycle de vie (p. ex. DDI 4) nécessite également de repérer des métadonnées répétées, comme des domaines de valeurs réutilisés (p. ex. des listes de codes de style Likert réutilisés) se répétant pour de multiples variables. Une fonction R a permis ce type de mise en correspondance en utilisant la fonction R « all.equal », excluant les différences en matière d'organisme, d'identification et de version.

Service de vocabulaire du CESSDA (Consortium d'archives européennes de données de sciences sociales) : nouvel outil dernier cri permettant de créer et de publier des listes de termes contrôlés

Sanda Ionescu, Inter-university Consortium for Political and Social Research
Taina Jaaskelainen, Archive finlandaise de données de sciences sociales

DDI Alliance crée et publie ses propres vocabulaires contrôlés depuis 2005. On les cible pour des catégories de la DDI particulières, mais ils proviennent de l'extérieur de la DDI et peuvent être utilisés avec d'autres normes de métadonnées.

Dans son nouveau modèle de métadonnées, le CESSDA recommande l'utilisation de vocabulaires contrôlés par la DDI lorsque cela est possible, ainsi que la création de nouvelles listes en fonction des besoins. Pour soutenir ces efforts, le CESSDA a financé l'élaboration d'un outil sophistiqué facilitant la création, la traduction et la publication de vocabulaires contrôlés et automatisant une part importante du processus. L'outil en est maintenant aux dernières étapes d'essai et devrait être lancé ce printemps.

Notre affiche présentera ce nouvel outil Web dans le contexte de notre travail continu sur les vocabulaires contrôlés et souligne certaines de ses fonctionnalités les plus essentielles, comme des recherches transvocabulaires, le partage et la modification en groupe, la prise en charge de la traduction en plusieurs langues, le versionnage automatisé, la publication en un clic et les téléchargements multiformat.

Exploitation des codes de langue dans une transformation de feuille de style : Open Language Archives Community (OLAC) (ISO 639-3) en MARC

John Huck, bibliothèques de l'Université de l'Alberta

Pouvoir générer des enregistrements MARC au moyen d'une transformation de feuille de style était l'un des éléments nécessaires d'une solution récemment présentée permettant la découverte d'une collection de ressources de données linguistiques des bibliothèques de l'Université de l'Alberta et l'accès à cette collection. Cette affiche montre la façon dont les codes de langue ISO 639-3 des métadonnées sources ont été transformés en vedettes-matières de la Bibliothèque du Congrès et en codages de langue MARC pour environ 750 enregistrements.

Crosswalk 4.0 ajoute DDI 3.2

Florio Arguillas Jr. et William Block, Cornell Institute for Social and Economic Research

Lors de la NADDI de l'année dernière à Washington, Crosswalk 3.0 a démontré la façon dont il redonnait vie à des données presque mortes (des ensembles ASCII sans fichier de configuration joint, mais seulement une copie papier du dictionnaire de données) en simplifiant le processus de création de fichiers et d'ensembles de données de configuration à l'aide d'un tableur Excel contenant des renseignements

sur les variables et leur emplacement et en utilisant l'ensemble de données ASCII original comme données d'entrée. Parallèlement à la création de fichiers et d'ensembles de données de configuration SAS, SPSS et Stata, Crosswalk 3.0 publie également le livre des codes DDI 2.5. Nous présentons cette année Crosswalk 4.0, la dernière version du logiciel, lors d'une séance de présentation par affiche. Outre les fonctionnalités figurant dans la version 3.0, nous avons ajouté la capacité de créer des fichiers de configuration R, des espaces de travail R ainsi que DDI 3.2 Lifecycle.

Répertoire de données du Bureau du recensement des États-Unis

Jared Lyle, Inter-university Consortium for Political and Social Research

Nous fournirons un aperçu du répertoire de données du Bureau du recensement des États-Unis, qui conserve et diffuse des instruments d'enquête, des spécifications, des dictionnaires de données, des livres de codes et d'autres documents fournis par le Bureau du recensement des États-Unis. ICPSR, l'hôte de ce répertoire de données, a également dressé la liste d'autres collectes de données relatives au recensement de ses fonds de données plus vastes. Nous présenterons également du contenu DDI produit par le recensement utilisé pour l'archivage de collections au sein du répertoire.

De 9 h 35 à 10 h 45

Cinquième séance

Président de séance : Amber Leahey, Scholars Portal

Mise en œuvre de la prise en charge de DDI 3.3 dans Colectica, version 6

Dan Smith, Colectica

Colectica version 6 comportera un grand nombre de nouveaux domaines de contenu de DDI Lifecycle 3.3 et adoptera bon nombre des nouvelles fonctionnalités, tout en maintenant la compatibilité avec toutes les versions précédentes de la norme.

Dans cet exposé, on passera en revue le nouveau contenu disponible dans DDI 3.3 et la façon dont il est ajouté à Colectica version 6 dans les domaines suivants : gestion de la classification basée sur GSIM et Neuchâtel et migration de Colectica d'une mise en correspondance Copenhague à DDI 3.3; ajout aux questionnaires Colectica de mesures de collecte de données ne provenant pas d'enquêtes; échantillonnage et pondération; conception de questionnaire; soutien au travail avec la DDI comme graphique de propriétés comprenant des propriétés d'éléments et de référence par services RDF de DDI 3.3 et par Colectica pour le répertoire Colectica; et améliorations de l'énoncé de la qualité. De plus, dans cet exposé, on mettra l'accent sur le modèle officiel désormais terminé qui sous-tend DDI 3.3 et son utilité en matière de développement de logiciels ainsi que sur la façon dont Colectica aborde la mise en œuvre de DDI pour assurer la rétrocompatibilité.

DDI 4 dans R : nouvelles possibilités

Larry Hoyle, Institute for Policy and Social Research, Université du Kansas
Joachim Wackerow, GESIS, Institut Leibniz pour les sciences sociales

Nous travaillons à une représentation du modèle DDI 4 dans R, réalisé sous forme d'un ensemble. Nous disposons désormais d'une catégorie axée sur les objets R pour chaque catégorie DDI 4 s'accompagnant des fonctions connexes pour valider et imprimer des objets, gérer un registre d'objets DDI 4, gérer des URN DDI 4 ainsi qu'importer et exporter du contenu DDI 4 XML.

Notre objectif initial était d'améliorer la capacité des chercheurs à saisir des métadonnées à la source et à produire des rapports ainsi qu'à intégrer des références aux métadonnées.

Dans cet exposé, nous discuterons de la perspective intéressante d'effectuer des calculs directement sur des métadonnées. Que pourrait-on faire avec les objets de métadonnées pour faciliter la comparaison et l'harmonisation? De quelles manières pourrait-on visualiser les métadonnées?

DDI 4 propose de nouvelles capacités puissantes en matière de tendance de collecte. Pour les catégories réalisant une collecte, les opérateurs pourraient être définis pour donner leur intersection, leur union et leur différence. Il serait également possible de définir les bornes internes et externes. Les relations au sein de la collecte pourraient être visualisées à l'aide de diagrammes de réseau. Ces opérateurs pourraient fournir des outils efficaces d'harmonisation.

Des opérateurs pourraient également être définis pour des paires d'objets d'une même catégorie. Des mesures de similarité pourraient être calculées à l'aide d'attributs correspondants. Elles pourraient être utilisées pour créer des visualisations (p. ex. des visualisations de similarités entre les variables).

Projet Picasso

Bureau de gestion de la protection de la vie privée et de coordination de l'information, Statistique Canada

Picasso est une solution pour entreprise aux fins de gestion de données statistiques et de métadonnées. Des règles opérationnelles automatisées permettront aux métadonnées d'être rassemblées de manière uniforme,

en respectant des instruments communs d'architecture, de gouvernance et de politique.

Picasso remplace les solutions locales par un centre unique de gestion de métadonnées pour l'ensemble des enquêtes, des fichiers administratifs et des projets de couplage d'enregistrements. Il s'agit d'un point d'accès unique pour tous les fichiers de données prêts à l'utilisation et d'un moteur de recherche et de découverte pour entreprises utilisant des métadonnées afin de faciliter la réutilisation de l'information. Son interface conviviale comprend un déroulement des tâches automatisé et une gestion intégrée du cycle de vie.

Picasso rassemble la puissance et le potentiel de l'information en une seule plateforme pour promouvoir la réutilisation de données et de métadonnées statistiques, afin de soutenir la production et l'analyse de statistiques. Son registre et son dépôt bien organisés ainsi que ses outils conviviaux sont conçus pour soutenir la croissance rapide des actifs de données et le passage à une approche donnant la priorité aux données administratives. Des principes de gestion de l'information sont intégrés au flux de travail automatisé, afin d'assurer une bonne gestion des données et de faciliter la recherche, l'accès, la réutilisation, la gestion et la production de rapports en matière de données et de métadonnées statistiques pour tous les employés.

Picasso respecte des modèles de référence (p. ex. GSIM et GSBPM) et prend en charge les vocabulaires standards (p. ex. DDI, SDMX, DCAT, SKOS/XKOS, PROV-O) dans divers formats (p. ex. JSON, XML, RDF).

De 10 h 45 à 11 h

Pause

De 11 h à 12 h 15

Sixième séance

Président de séance : Bill Block, Université Cornell

Consignation de la provenance de variables avec la DDI et Colectica

Jeremy Iverson, Colectica

La DDI prend en charge la consignation de données en offrant des renseignements détaillés au niveau des variables. Normalement de tels renseignements comprennent un nom de variable, un libellé et un type de données, mais la DDI permet d'inclure plus de précisions (p. ex. l'historique de la variable). Les données correspondant à une variable peuvent provenir d'une enquête, de données administratives ou d'un autre ensemble de données. La variable peut avoir été copiée directement d'une source ou avoir été calculée à l'aide d'une formule ou d'un algorithme. Pour qu'un chercheur puisse comprendre et analyser correctement les données, l'historique de chaque variable devrait être consigné en comportant autant de renseignements détaillés que possible.

Des renseignements sur la provenance au niveau de la variable peuvent être enregistrés manuellement par le personnel chargé de la gestion des données. Cependant, dans certains cas, il est également possible d'extraire les transformations utilisées pour créer une variable en fonction du code source statistique. Le projet C²Metadata, commandité par la Fondation nationale des sciences des États-Unis, élabore actuellement des outils pour accomplir cela et enregistrer de façon structurée les transformations pouvant être incluses dans la documentation des données.

Cet exposé montrera la documentation de données d'études réelles qui fournissent des renseignements sur la provenance au niveau des variables et traitera des méthodes employées pour créer et publier la documentation.

Avantage de la DDI en matière de conformité pour un modèle générique du processus de production statistique

Division de l'accès aux microdonnées, Statistique Canada

La Division de l'accès aux microdonnées est chargée de fournir aux chercheurs à l'extérieur de Statistique Canada un accès aux microdonnées. La Division offre également aux secteurs spécialisés de Statistique Canada du soutien et de l'expertise et les renseigne sur les normes régissant la diffusion des données.

La Division de l'accès aux microdonnées présentera le projet de Laboratoire virtuel de données et la façon dont la division profiterait d'un modèle générique du processus de production statistique en utilisant la DDI, de la conception à la diffusion.

Enquête sur la population active : amélioration de l'accès pour les chercheurs

Susan Mowers, bibliothèques de l'université d'Ottawa

L'Enquête sur la population active (EPA) est une enquête mensuelle produite par Statistique Canada. Elle génère un certain nombre de produits de données utilisés par des chercheurs de l'ensemble du pays. Les fichiers mensuels de microdonnées à grande diffusion distribués dans le cadre de l'Initiative de démocratisation des données en coordination avec les services de données fournis par les bibliothèques universitaires exigent exactitude et cohérence des métadonnées en format DDI pour offrir un accès en ligne. On traitera, au cours de cet exposé, du processus de balisage de l'EPA entre les diverses séries de fichiers de données (de 1976 à nos jours) en anglais et en français ainsi que des différents auteurs de métadonnées de divers établissements au fil des années.

De 12 h 15 à 13 h 30

Dîner

De 13 h 30 à 14 h 30

Septième séance

Séance plénière de fermeture—Présentation : James Doiron, bibliothèques de l'Université de l'Alberta

Jeffrey Moon, directeur, Association des bibliothèques de recherche du Canada, réseau Portage

De 14 h 30 à 15 h

Synthèse et évaluation

Résultats de l'évaluation de la conférence sur la DDI

Barry T. Radler, Université du Wisconsin à Madison

Synthèse

Jared Lyle, directeur, DDI Alliance, Inter-university Consortium for Political and Social Research

Cory Chobanik, Statistique Canada

Commanditaires et partenaires

Platine



Or



Argent



