

Initiative de démocratisation des données (IDD)



Présentation sur l'Initiative de démocratisation des données

Michel B. Séguin - Statistique Canada

Gaston Quirion - Université Laval

Mémoires & démographie, 19-22 juin 2007



Initiative de démocratisation des données (IDD)

*Le besoin de
démocratiser les
données*



Initiative de démocratisation des données

Le besoin de démocratiser les données

- Historiquement Statistique Canada rendait les données disponibles via le programme de dépôt
- C'était des publications papiers et aucun fichier numérique ne faisait partie de ce programme
- Des fichiers de données numériques étaient disponibles aux chercheurs à coûts marginaux
- Des tableaux personnalisés étaient une autre façon, mais plus coûteuse, d'obtenir des données non publiées.



Initiative de démocratisation des données

Le besoin de démocratiser les données

- Années 80 – Les coupures du budget fédéral ont mis de l’emphase sur le recouvrement des coûts à Statistique Canada
- Années 90 -- Coût des fichiers de microdonnées publics a augmenté de façon dramatique
- La majorité des fichiers sont devenus trop dispendieux pour la majorité des chercheurs et surtout pour les étudiants



Initiative de démocratisation des données

Le besoin de démocratiser les données

- 1986 – un consortium d'universités a été créé afin de fournir un meilleur accès aux données du recensement de 1986
- Cette idée a été bien reçue par Statistique Canada et un mouvement a été créé pour la démocratisation des données sur fichiers électroniques



Initiative de démocratisation des données

Le besoin de démocratiser les données

- Des champions de la communauté académique et de Statistique Canada se sont réunis pour mettre de l'avant l'idée d'un consortium
- L'approbation informelle a été reçue en 1995 pour créer l'IDD – l'Initiative de démocratisation des données
- Cette approbation fut suivie de la création:
 - D'un comité de direction interne à Statistique Canada
 - D'une équipe de projet
 - D'un comité consultatif externe à Statistique Canada



Initiative de démocratisation des données

Le besoin de démocratiser les données

- Une licence a été préparée et approuvée
- On a demandé aux divisions auteures de fournir des données à l'IDD
- Une invitation pour se joindre à l'IDD a été envoyée aux institutions académiques
- L'approbation formelle du Conseil du Trésor a été obtenue au début de l'année 1996



Initiative de démocratisation des données

Le besoin de démocratiser les données

- L'IDD utilise principalement l'Internet comme outil de diffusion
- Création d'un site FTP à Statistique Canada
- Création d'un « listserv » – un forum pour partager les questions et réponses



INITIATIVE DE DÉMOCRATISATION DES DONNÉES

Le besoin de démocratiser les données

- Avant l'IDD, environ 15 institutions offraient un service de données
- La formation co-opérative fut identifiée comme première priorité compte tenu du degré d'expérience variée de chacun des membres de l'IDD
- Un comité de formation a été mis en place et un curriculum, des formateurs et un budget ont été identifiés
- La formation régionale a débuté en 1997



Initiative de démocratisation des données

Le besoin de démocratiser les données

- En 1996, il y avait 50 membres post-secondaires
- En 1998, il y avait 61 membres
- Aujourd'hui, il y a 71 membres
- Plus de 19,000 fichiers dans la collection, tous accessibles via l'Internet et le site FTP



Initiative de démocratisation des données

Le besoin de démocratiser les données

- L'IDD est maintenant un programme permanent à Statistique Canada
- Les diplômés d'aujourd'hui ont eu l'opportunité d'utiliser des données canadiennes durant toute la durée de leurs études
- L'IDD est décrit comme l'un des plus importants développements pour les sciences sociales au Canada dans les dernières 50 années.



Initiative de démocratisation des données

2006

10 années de succès

- Licence
- Produits
- Formation
- Communauté



Initiative de démocratisation des données

Le besoin de démocratiser les données

L'accès aux données (FMGD)

- Données déposées sur le site ftp de l'IDD
- Seul le représentant IDD y a accès



Initiative de démocratisation des données

Le besoin de démocratiser les données

Deux contraintes pour le chercheur

- **Accès par un tiers**
- **Besoin de connaissances particulières** (logiciels d'exploitation statistique : SPSS, SAS, etc.)



Initiative de démocratisation des données

Le besoin de démocratiser les données :
un pas de plus...

- **Plusieurs initiatives régionales et locales pour faciliter l'accès** (téléchargements directs : services de données centralisés)
- **Au Québec : regroupement de certaines universités** (partage des expertises)
- **Création d'une plateforme électronique de gestion et distribution des FMGD : Sherlock**
- **Sherlock :**
<http://sherlock.crepuq.qc.ca/public/francais/accueil.html>



Initiative de démocratisation des données

Sherlock :

un accès privilégié aux données de l'IDD

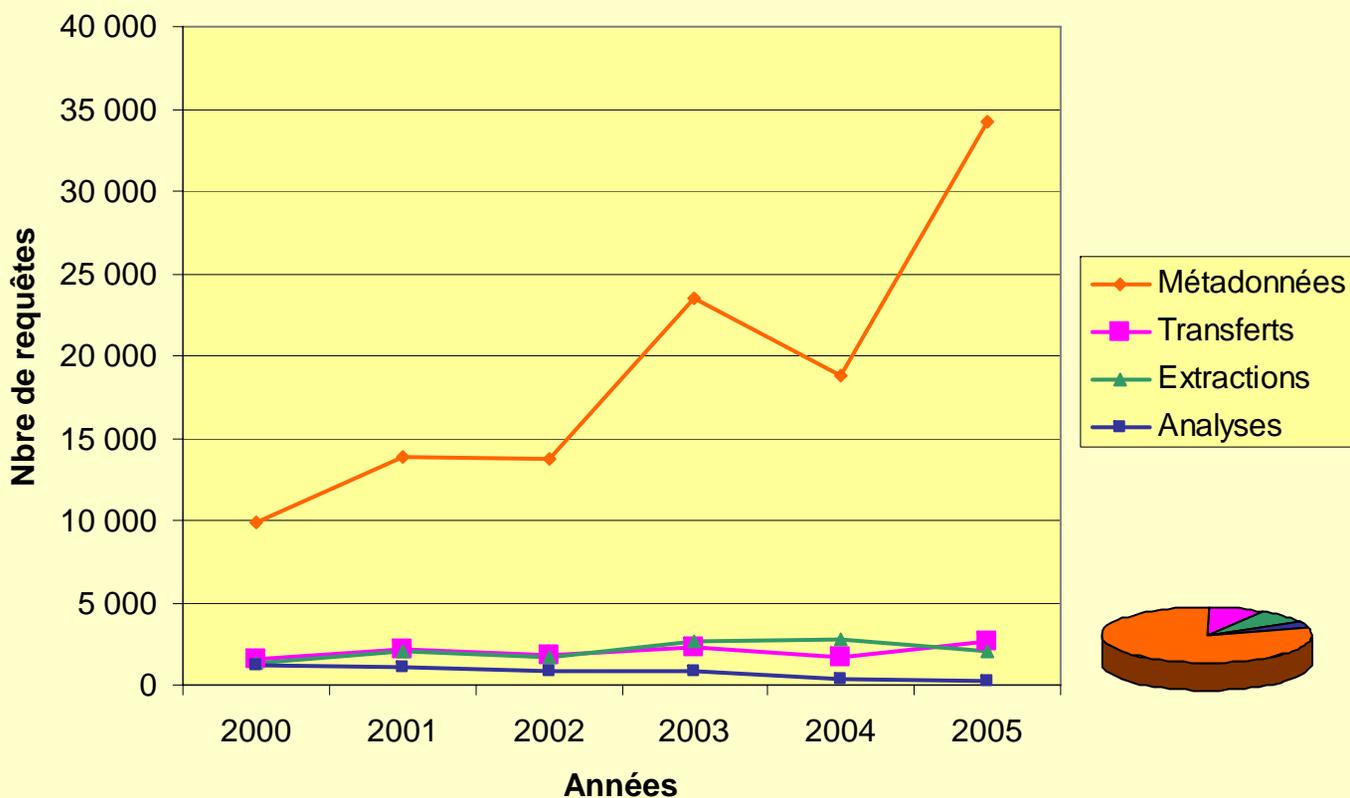
- **Principales fonctionnalités :**
 - Transfert direct des FMGD (reconnaissance IP)
 - Extractions de sous -ensembles (variable)
 - Analyse et exploration (de base)
- **Accès libre à la documentation**
- **Possibilités de recherche avancée (variable)**
- **12 établissements participent à Sherlock**



Initiative de démocratisation des données

Coup d'œil sur l'utilisation de Sherlock

Évolution des requêtes, 2000-2005





Initiative de démocratisation des données

Réflexions...

- **Importance de la documentation** (métadonnées)
76.5% de l'ensemble des requêtes
- **On cherche et on trouve les métadonnées**
(culture des données)
- **Importance de métadonnées décrites selon des normes** (DDI)



Initiative de démocratisation des données

Défis

- S'assurer de la prise en charge des nouvelles normes de description des métadonnées par nos plateformes de distribution
- Garantir la pérennité des données (réelle démocratisation des données)
- Diffuser sans cesse de manière plus efficace les données de l'IDD/accroître nos collections