

Secteur du développement des télécommunications



7ème réunion sur les indicateurs des télécommunications/TIC dans le monde
Le Caire (Egypte), 3-5 mars 2009

Document 017-F
27 février 2009

Original: anglais

SOURCE: Malaysian Communications and Multimedia Commission, Malaisie

TITRE: Données primaires, données secondaires et données administratives dans le secteur des télécommunications



Données primaires, données secondaires et données administratives dans le secteur des télécommunications

- Le cas de la Malaisie -

Koay Hock Eng
Malaysian Communications and Multimedia Commission
SKMM



Ces données revêtent diverses formes...

- Par exemple
 - Données primaires (dont des données en “piggy-backs”)
 - Données secondaires
 - Données administratives
- Mais aussi, selon Guven (2001), de la Division statistique de la CESAP
 - Données sur le suivi
 - Méthodes de calcul
 - Données indirectes

Données primaires

- Définition
 - Données collectées directement auprès d'organismes statistiques, conformément à des objectifs de recherche
- Avantages
 - Données sur mesure, avec tous les avantages correspondants
- Inconvénients
 - Coût
 - Temps nécessaire à leur extraction
 - Risques d'erreur, pas toujours quantifiables
- Données en "piggy-backs"
 - Quelquefois moins onéreuses et plus rapides à obtenir
 - Leur communication représente une charge; taux de réponse global assez faible

Données secondaires



- Définition
 - Statistiques ou données extraites enquêtes/recensements ayant d'autres objectifs que l'objectif directement visé.
 - Au Chapitre 2 de son ouvrage, **Marketing Research and Information Systems**, publié par le Bureau régional de la FAO pour l'Afrique, Crawford (1997) s'est prononcé énergiquement pour que l'on étudie les données secondaires avant de commencer à étudier les données primaires.
- Avantages par rapport aux données primaires
 - Plus facilement disponibles et tout à fait adéquates.
 - Moins onéreuses.
 - Plus rapides à obtenir.



Données secondaires

- Inconvénients
 - Différences au niveau de la définition
 - Différences au niveau de la période de référence
 - Différences au niveau de la population cible
 - Erreur dans les mesures
 - Biais à la source
 - Fiabilité
 - Echelle de temps



Données administratives

- Définition
 - Données recueillies couramment dans un but administratif précis, par exemple l'enregistrement des abonnés par les fournisseurs de services
 - De l'avis de Marquis, Palacios & Wetrogan (1996), "Dans l'avenir, les systèmes statistiques publics utiliseront intensivement les données administratives. La demande d'informations augmentera dans de telles proportions que les enquêtes et recensements habituels ne permettront pas d'y répondre".



Données administratives

- Avantages
 - Précision
 - Peuvent être d'une grande efficacité lors de la phase préparatoire
 - Contribuent à définir la population ciblée
 - Peu onéreuses
 - Peuvent être obtenues rapidement
 - Moins contraignantes pour les administrés et pour les organisations
 - Statistiques plus précises et mieux ciblées
 - Il est possible d'obtenir davantage d'informations, sans la contrainte d'avoir à remplir un long questionnaire
 - Faible pourcentage de données manquantes, faible nombre de corrections pour non-réponse, faible nombre d'imputations et moindres biais pour non-observation
 - Pas de possibilité d'être influencé par la personne qui conduit l'entretien
 - Pas d'erreurs humaines telles que oubli, etc.



Données administratives

- Inconvénients
 - Sont généralement recueillies par des personnes qui ne sont pas véritablement intéressées par les statistiques (par exemple, fonctionnaires de l'Administration des douanes) et leur qualité peut donc être sujette à caution
 - Difficulté ou impossibilité à se les procurer
 - Coûts parfois très élevés pour localiser, acquérir, faire correspondre et revoir les données, remédier aux erreurs et établir des estimations
 - Le champ de l'étude n'est pas toujours complet
 - Les données recueillies ne coïncident que rarement avec les concepts
 - Les informations ne sont pas nécessairement à jour
 - Des problèmes de mise en correspondance se posent pour évaluer les relations entre les variables
 - Problèmes liés au respect de la vie privée, à la confidentialité et à la divulgation

Données administratives



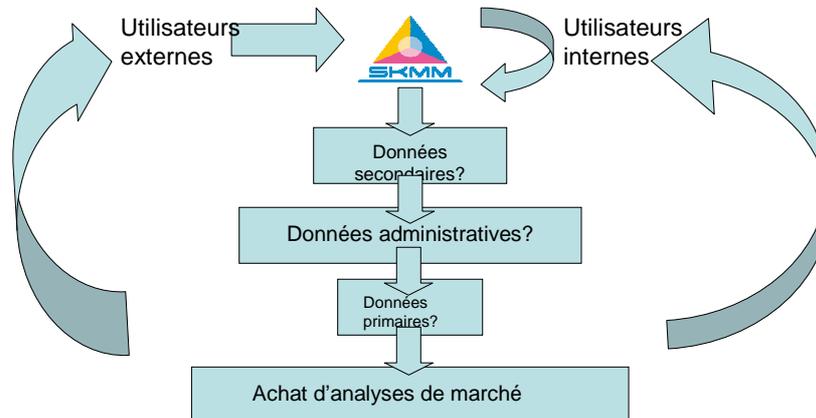
- Selon Marquis et d'autres auteurs, des bases de données administratives sont utilisées avec succès:
 - au Canada (Standish et Ravindra, 1998)
 - en Finlande (Myrskylä, 1991 et Lankinen, 1995)

Le cas de la Malaisie



- La SKMM est la Malaysian Communications and Multimedia Commission.
- Depuis sa création, la SKMM a un rôle en matière de statistiques puisqu'elle collecte et utilise des données administratives sur les détenteurs de licence.
- A l'heure actuelle, la SKMM a étendu ses activités en matière de statistiques.

Activités statistiques de la SKMM

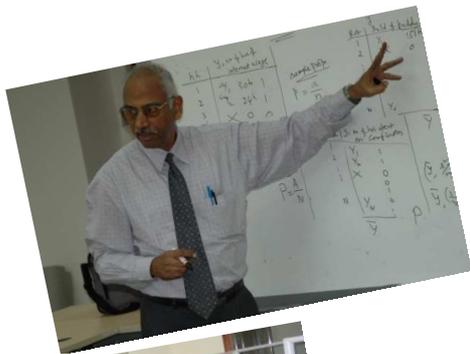


Collecte de données primaires

- Cette collecte a réellement débuté en 2004 avec la création du Centre CATI de la SKMM.
- Ce centre a, entre autres, réalisé les études suivantes:
 - Enquête sur les utilisateurs de téléphone mobile (enquête annuelle depuis 2004)
 - Enquête sur l'utilisation à domicile de l'Internet (2005, 2006, 2008)
 - Enquête sur le large bande mobile (enquête ponctuelle 2009)
- Ces enquêtes permettent de collecter des données primaires sur l'utilisation (par exemple, SMS, opérations bancaires sur Internet, etc.), que l'on n'obtient pas avec les données administratives.
- Le régulateur est-il bien placé pour effectuer des enquêtes sur des données primaires?



Un aperçu du
Centre CATI de la
SKMM



Un aperçu du
Centre CATI de la
SKMM



Collecte de données primaires



- A partir de 2005, collecte de données primaires sur le web
 - Enquête 2005 sur l'utilisation de l'Internet au domicile
 - Enquête annuelle sur les détenteurs de licence
- Mise en garde contre une exposition abusive sur le web

Exemples d'utilisation de données secondaires

- Recensements et enquêtes auprès des ménages, avec l'aimable autorisation de l'Organisme national de la statistique de Malaisie.
- Examinées sous l'angle des similitudes entre les concepts, les définitions et les périodes de référence.
 - Par exemple, sous l'angle de la population pour une année donnée
- Utilisées pour obtenir:
 - le pourcentage de ménages ayant un ordinateur.
 - le pourcentage de ménages ayant une connexion Internet.
 - le pourcentage de ménages ayant une télévision.
 - le pourcentage de ménages ayant une radio.
 - le pourcentage de ménages ayant une ligne téléphonique d'abonné.
- Longue période de préparation.
- Il convient d'améliorer la régularité.

Utilisation des données administratives



- Les données administratives collectées auprès des détenteurs de licence concernent par exemple les aspects suivants:
 - Abonnements
 - Commutateurs
 - Passerelles internationales
 - Points d'interconnexion
 - Minutes de trafic et recettes
- Tous les services
 - Lignes téléphoniques d'abonné
 - Téléphonie cellulaire mobile
 - Large bande

Règles régissant la tenue à jour de ces données



- Loi de 1998 "Record Keeping Rules, Communications and Multimedia"; Section 268
- Le ministre peut établir des règles qui sont publiées dans la Gazette officielle, concernant la tenue des fichiers et peut exiger qu'un ou plusieurs détenteurs de licence ou particuliers tiennent à jour et conservent ces données.
- Les champs des données ainsi conservées dans le cadre d'une procédure administrative peuvent être modifiés ou les données peuvent être transmises dans le cadre d'une procédure de collecte de données primaires.
- Il est prévu que les données ainsi conservées fassent l'objet d'un audit.
- Cette démarche permet d'améliorer la portée et la qualité des données administratives et de les uniformiser.



Synergie entre les données primaires, les données secondaires et les données administratives

- A l'heure actuelle, les données primaires, les données secondaires et les données administratives sont collectées et utilisées de manière à établir des synergies.
- Lors de la planification d'études sur les données primaires
 - Les détenteurs de licence recherchent dans leurs données administratives le nombre d'abonnements par réseau.
 - Des échantillons à composition aléatoire sont extraits, avec une probabilité proportionnelle à la dimension du réseau.
 - Des estimations sont faites quant au nombre d'utilisateurs.
 - Les estimations liées aux données démographiques secondaires fournies par les offices nationaux de la statistique donnent des taux de pénétration variés



Synergie entre les données primaires, les données secondaires et les données administratives

- Principaux points
 - Evite d'avoir à établir un cadre pour une population infinie et un unique domaine d'étude.
 - Permet même de calculer les erreurs types relatives.
 - Efficace dans une population où le taux de pénétration approche de la saturation, par exemple, en Malaisie, pour la téléphonie mobile cellulaire.
 - Même les données administratives de base contiennent un numéro NRID (date de naissance, âge et code pour indiquer le sexe).
 - Il est possible de faire des vérifications croisées avec la base de données de l'office national de la statistique pour situer l'abonné (zone urbaine ou zone rurale).
 - La contrainte est moindre pour ceux qui doivent répondre aux question et on évite ainsi des erreurs dues à l'ignorance.



Exemple d'utilisation efficace des données administratives

- Identification des zones mal desservies
 - Les détenteurs de licence recherchent dans leurs dossiers administratifs le nombre d'abonnements dans une localité donnée.
 - Le chiffre de la population de cette localité est donné par l'Office national de la statistique.
 - On calcule le taux de pénétration dans cette localité.
 - Les localités sont considérées comme mal desservies si le taux de pénétration est inférieur à un pourcentage préalablement déterminé.



Exemple d'utilisation efficace des données administratives

- Principaux points
 - L'autre solution serait d'effectuer un recensement ou un échantillonnage dans chaque localité en indiquant le taux de pénétration dans chaque cas. Cette procédure serait trop longue à réaliser et trop onéreuse.
 - Il est nécessaire d'effectuer des mesures à intervalles réguliers pour évaluer les progrès du déploiement, ce qui implique l'utilisation de données administratives plutôt que de données primaires obtenues au terme d'enquêtes.
 - Les données recueillies n'ont pas à être extrêmement précises. Si le critère de référence est un taux de pénétration de 14%, que ce taux soit de 8,6% ou de 8,1% dans une localité n'a pas grande importance puisqu'elle est de toute manière mal desservie.

Conclusion



- Les discussions sur les avantages et les inconvénients de l'utilisation des données primaires, des données secondaires et des données administratives ne permettent pas de tirer de conclusions.
- Les données secondaires et les données administratives peuvent être très utiles dans certains cas, en fonction des éléments suivants:
 - Finalité
 - Utilisation
 - Exactitude
- Faire abstraction de ces deux catégories de données au profit des données primaires équivaldrait à «réinventer la roue».
- Faire totalement abstraction des données administratives reviendrait à «jeter le bébé avec l'eau du bain».

Conclusion



- Les données primaires obtenues au terme d'enquêtes peuvent être utiles pour mesurer l'utilisation (par exemple utilisation réelle de l'Internet, caractéristiques des internautes, activités en ligne, etc.).
- Les données administratives sont utiles, mais ne peuvent servir à mesurer l'utilisation réelle (par exemple, il arrive que les abonnés au téléphone mobile soient comptés en double (cartes SIM à prépaiement inactives, etc.).
- Néanmoins, les enquêtes basées sur des échantillons sont sujettes à des erreurs d'échantillonnage, qui peuvent être évaluées et quantifiées.
- Les problèmes les plus insidieux sont la partialité, les erreurs de couverture, les erreurs de non-réponse, les fraudes et les mensonges délibérés dont on ne connaît pas précisément l'étendue.



Références

- Marketing research and information systems
www.fao.org/docrep/w3241E/w3241e03.htm
- Data Types and Sources of Environment Statistics
www.unescap.org/stat/envstat/stwes-mo1.5.pdf
- TOWARDS A U.S. POPULATION DATABASE FROM ADMINISTRATIVE RECORDS
www.amstat.org/sections/sgovt/marquis.pdf



Merci

S T A T I S T I C S

Thinking big • Starting small • Scaling fast