Statistics Canada
www.statcan.gc.ca

Reducing Response Burden with alternative data, analysis and collection methods

24th PACIOLI-workshop

Presented by

Martin S. Beaulieu (Agriculture Division)

Outline

- Purpose
- Background
- New data ecosystem
- Ongoing projects and potential development in
  - Data
  - Analysis
  - Collection methods
- Challenges
- Conclusion

Purpose

- We have made significant progress in replacing survey data with administrative and remote sensing data
- What else can be done to fill data gaps in agriculture statistics?
- What are the challenges to overcome?

Background

- Several data gaps identified during the 2016 Census of Agriculture Content Consultation
  - http://www5.statcan.gc.ca/olc-cel/olc.action?lang=en&ObjId=95-635-X&ObjType=2
  - Users want more detailed related to existing content
  - Users want new content on such as:
    - Agri-tourism
    - Animal welfare
    - Value chain and marketing channels
    - GMO
    - Grain storage on farm

Background

- Statistics Canada Agricultural Statistics Program
  - Census of Agriculture every 5 years
  - 17 surveys directly to farmers
  - Remote Sensing and Geospatial Analysis
  - Agricultural Taxation Data (more than 25 years)
  - Over 200 administrative data sources
  - New administrative data (crop insurance, income stabilisation program…)

A new data ecosystem

- A wave of technological changes :
  - New ways of generating and sharing data: new technological platforms
  - New data sources: Big Data and alternative data sources
  - New actors producing and consuming data: local, national, international, public, and private
  - New data governance models: Open Data
- More data at lower cost, accessible to more people, on common platforms/standards, from anywhere in the world

Alternative data in agriculture

- Data sources we should keep an eye on:
  - Internet of Things (IoT) – "Things" are sensors and devices that capture data and flow data to other devices: e.g. biochip transponders on farm animals, steer and control equipment, manage inputs at very precise levels across fields, detect crop stress
    - Precision agriculture
    - Track food along supply chain - sensor based technology
    - Smartphones "The pioneer species of the IoT is the smartphones. Every time we take a smartphone with us in a car, it beams information on our location and speed to Google. The result is real-time traffic information that can be used by everyone."
      Source: http://qz.com/156075/internet-of-things-will-replace-the-web/
  - Smart meters
  - Social media – food consumption behaviours, trends and attitudes

Earth observation & remote sensing

- Very high resolution remote sensing
  - E.g.: Sentinel-2, 20-60m resolution, global coverage, every 5 days; this is now Open Data
- Derived data products
  - Model-based Principal Field Crop Estimates
    - developed and tested on the crops that are typically published within the September Farm Survey
    - accounts for approximately 98% of the agricultural land in Canada
    - Alberta, Saskatchewan, Manitoba, Ontario and Quebec
    - 3 weeks in advance of the September Farm Survey, and
    - 11 weeks in advance of the November Farm Survey
- The Daily: http://www.statcan.gc.ca/daily-quotidien/150917/dq150917c-eng.htm
- CANSIM Table 001-0075:
  http://www5.statcan.gc.ca/cansim/home-accueil?lang=eng&p2=50&HPA

Earth observation & remote sensing

- Very high resolution remote sensing
  - E.g.: Sentinel-2, 20-60m resolution, global coverage, every 5 days; this is now Open Data
- Derived data products
  - Model-based Principal Field Crop Estimates
    - developed and tested on the crops that are typically published within the September Farm Survey
    - accounts for approximately 98% of the agricultural land in Canada
    - Alberta, Saskatchewan, Manitoba, Ontario and Quebec
    - 3 weeks in advance of the September Farm Survey, and
    - 11 weeks in advance of the November Farm Survey
- The Daily: http://www.statcan.gc.ca/daily-quotidien/150917/dq150917c-eng.htm
- CANSIM Table 001-0075:
  http://www5.statcan.gc.ca/cansim/home-accueil?lang=eng&p2=50&HPA

Alternative analytical methods

- Record linkage implies combining two or more micro-records to form a composite record combining information about the same entity (e.g. a business or an individual)
- Linkages are potentially important sources of valuable statistical information. But there are privacy concerns
- Canada has developed a policy framework permitting data linkage

Examples with farm taxation data

- 2015 Farm Financial Survey
  - Replacement selected revenues and farm expenses
- Agriculture Taxation Data
  - Uses the Census of taxation records of farm operations (unincorporated and incorporated)
- Longitudinal Census – taxation data
  - Productivity study

Business microdata linkage using the LFE: potentials and challenges

- For business data (including farm businesses), the Agency has developed a Linkable File Environment (LFE) facilitating the linkage of pre-approved databases
- The LFE brings together microdata holdings of Statistics Canada from both **administrative and survey sources**
- At the core of the LFE there is the Business Register (BR), which is used to identify the Business Number and implement all linkages
- Currently, LFE is intended to link records and to extract databases for analytical purposes

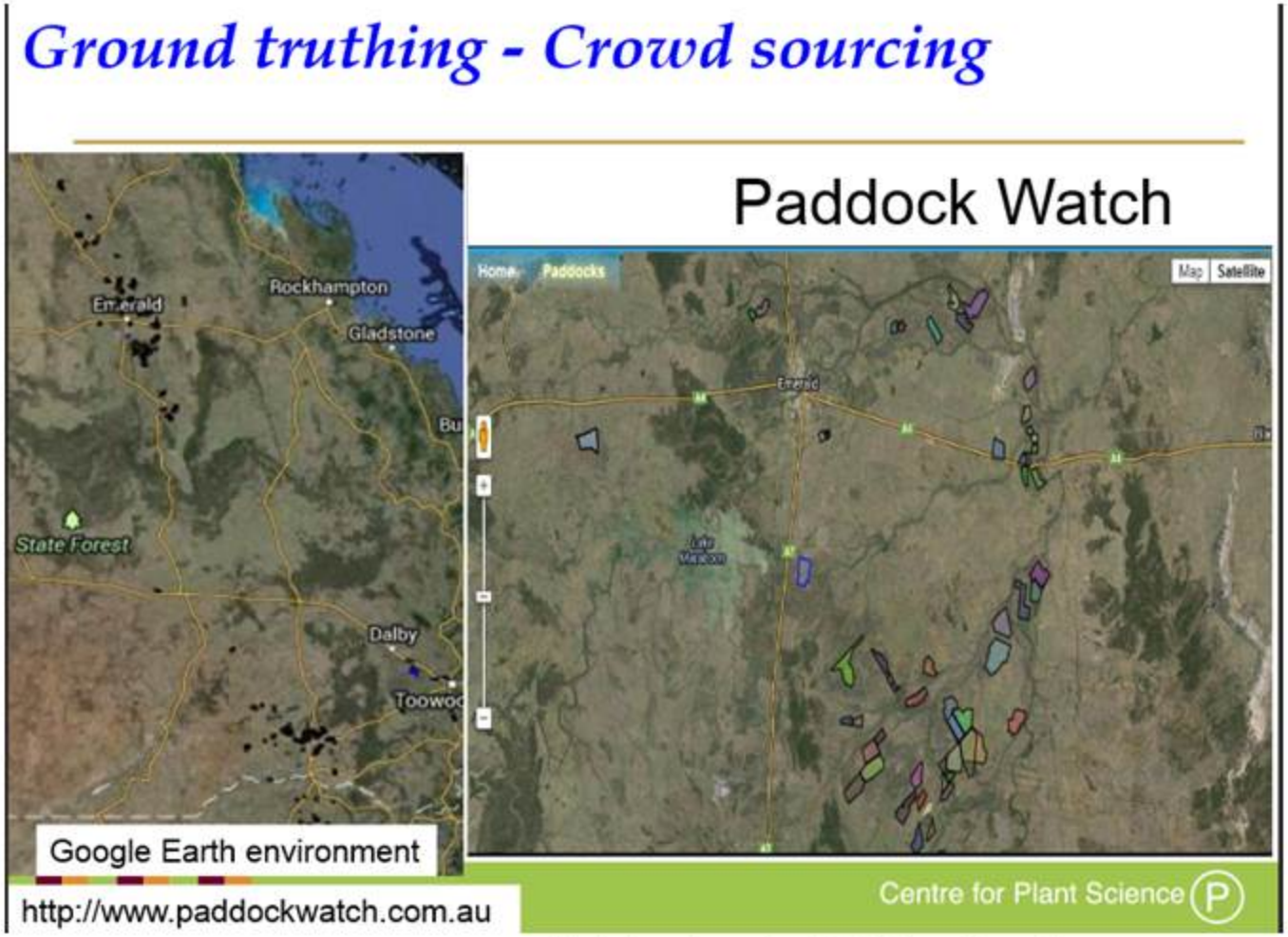
Crowdsourcing

- Much of the work on crowdsourcing agricultural data conducted in developing countries where there is lack of data
- Can we use crowdsourcing to reduce response burden and fill data gaps?
- Recently, Statistics Canada initiated a pilot project aimed at understanding the potential of crowdsourcing for statistical purposes – to describe characteristics of buildings

Crowdsourcing

- Collection will officially launch on October 2016
- The project is making use of OpenStreetMap as a platform for inviting contributors to crowdsource information on buildings
- Contributor will be able to input the location, physical attributes and other features of buildings
- A customized OSM editor will be used for this purpose
- The pilot will focus on the Ottawa-Gatineau region for now

Crowdsourcing in agriculture – an Australian example



A centralised hub for collecting and recording 'Ground Truthing' data.

Utilised to calculate total crop area planted and total crop volume.

By contributing information, users can win an min iPAD.

Takes 30-45 minutes, depending on the number of fields that the user wants to record.

"Click on the Paddock Watch link and start entering your fields to go into the draw for a mini iPAD."

Challenges to crowdsourcing

- Accessibility
  - The Government of Canada has standards that may not be applied by third party providers
- Acceptability and willingness to participate
  - Voluntary
  - Incentives to participate?
- Privacy and confidentiality protection
  - Informed consent for research and statistical uses of shareable information
  - Not for collecting sensitive information (could be viewed by the crowd)

Challenges to crowdsourcing

- Security
  - Third service providers standards (before sent to NSOs)
  - Agreement under Section 12 *Statistics Act*
    - *enter into an agreement with any … or other corporation for the sharing of information collected from a respondent by either Statistics Canada or .. corporation on behalf of both of them and for the subsequent tabulation or publication based on that information*
- Data quality and selection bias
  - Not every person with smart phone, mobile devices, GPS…
  - Crowdsourcing and potential false answer
  - Selected crowd (collection not fully outsourced)
- Building a business case for sharing
  - Proprietary value of privately held data

Conclusion

- Change in users expectations for timeliness, detail, frequency
- NSO's roles
  - Trusted party to certify alternative data quality
  - Promote the use of best practices
  - Partner and collaborate with other (private/public) data providers
- Explore, learn, adapt new technologies for data collection, processing, analysis and dissemination

Questions/contacts

Analysis and Outreach, Agriculture Division

Martin S. Beaulieu
martin-s.beaulieu@canada.ca

Centre for Special Business Projects, Crowdsourcing data collection pilot

Alessandro Alasia
alessandro.alasia@canada.ca

Remote Sensing and Geospatial Analysis, Agriculture Division

Gordon Reichert
gordon.reichert@canada.ca