Recommendations on a Geospatial Information Reference Framework for Public Health (GIRF)

How to address gaps, and support the public health community in the GIRF context.

Gail Kucera
Pierre Lafond
30 September 2009

HOLONICS

1

Background on the GIRF

- Geoconnections project April 16 - September 30, 2009
  - Part of an initiative to lay a foundation for public health community to exploit geospatial information in decision-making.
- What is the GIRF?
  - Categorical scheme *tailored to the public health community*.
  - Inventory of data sources mapped to the categorical scheme.
- Objectives of the GIRF
  - Improve the ability to search for geospatial data.
  - Facilitate communication between data users and data providers via an intuitive structured terminology.
  - Facilitate browsing for semantically proximal information.

HOLONICS

2

Development of the GIRF

- Development methodology
  - Drive out information requirements using a "strawman" categorical scheme.
  - Consult with public health community via questionnaire survey.
    - 123 stakeholders were invited to participate.
    - 52 stakeholders completed the questionnaire.
    - Extensive telephone/email follow-up.
  - Revise categorical scheme.
  - Adopt a "keyword" approach to incorporate existing terms & indices.
  - Locate data sources, map to categorical scheme.
  - Validate with stakeholders.
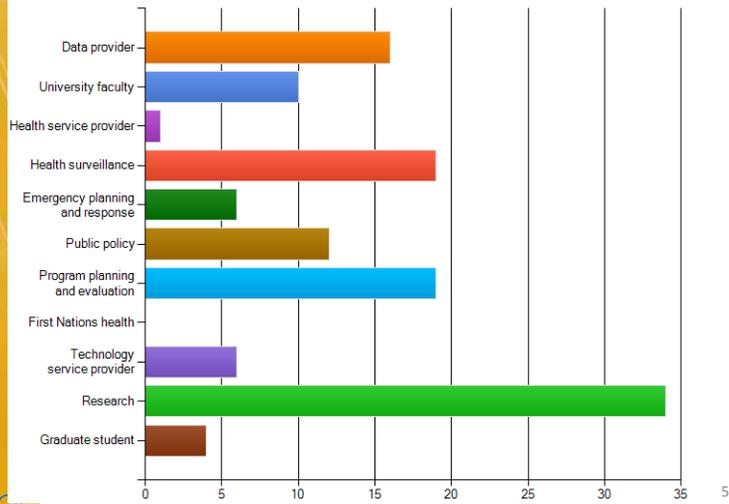- Strong support from stakeholders for the completed categorical scheme.

HOLONICS

3

Stakeholder participation:
areal distribution

If you work at a provincial or sub-provincial level, which province? (choose all that apply)

HOLONICS

4
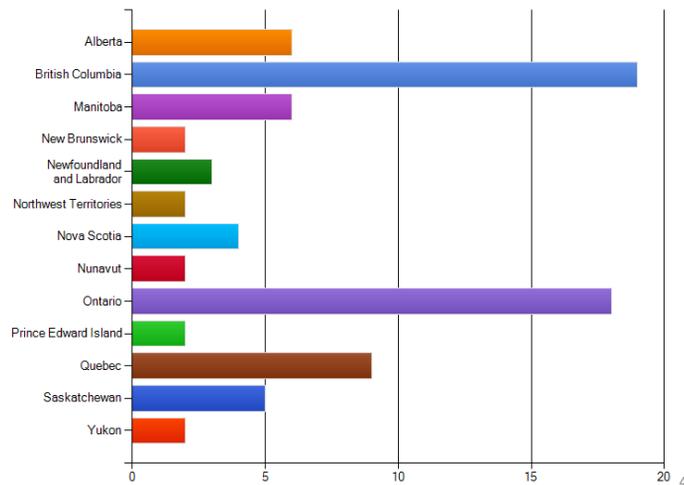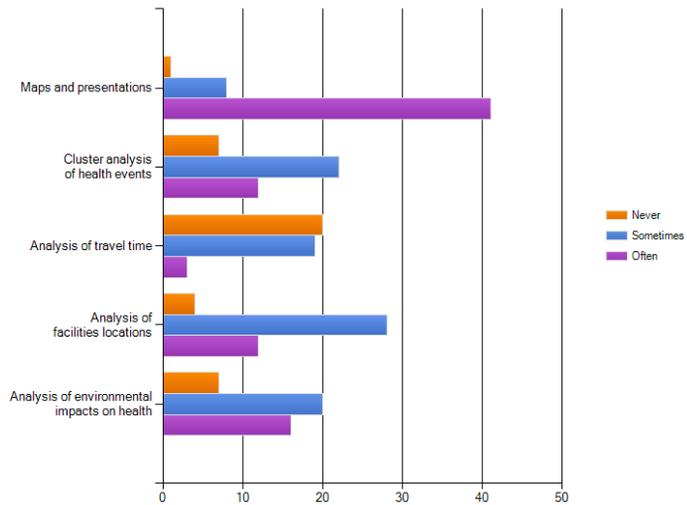
Stakeholder participation:
type of job

What is the nature of your job? (choose all that apply)

HOLONICS

5

Stakeholder participation:
geospatial data usage

How do you use geospatial information in your job?

HOLONICS

Nine classes in the GIRF

- Health Status
- Health Events
- Health Facilities and Services
- Health Hazard, Exposure, and Risk
- Population Demographics
- Natural Environment
- Built Environment
- Socio-economic Environment
- Geocoding Reference

HOLONICS

7

Details of GIRF classes (page 1)

Class	Subclasses	Relevant frameworks & standards	National data sources
Health Status	Death, Health Condition, Injury, Human Function, Well-Being, Maternal and Child Health, Use of Health Care System, Use of Pharmaceuticals.	ICD-10, CIHI Health Indicators, APHEO core indicators, PHAC Inventory of Injury Surveillance Data Sources and Surveillance Activities, PHAC Infectious and Chronic Disease categories.	Discharge Abstract DB, National Trauma Registry, National Ambulatory Care Registry.
Health Events	Outbreak, Intervention, Notification, Observation.	Based on terminology used by WHO Global Alert and Response Network, US CDC, and throughout Canadian and International public health communities.	Integrated Public Health Information System, Real-time Outbreak and Disease Surveillance System, Provincial health surveillance centres.
Health Facilities & Services	Facility description, Service delivery perspective, Care level, Service details, Functional perspective, Mobile, temporary or periodic facilities or services, Funding source.	National Infrastructure Data Model.	Mostly provincial and municipal.
Health Hazard, Exposure, and Risk	Health behaviours, Occupational, Environmental, Infectious or contagious disease, Vector-borne disease.	CIHI's "Non-medical determinants of health", APHEO "Health Behaviours," PHAC infectious disease reporting, Briggs classification for WHO.	Generally need very large-scale data.
Population Demographics	Age, Gender, Marital status, Education, Income, Household members, Clients of social programs, Employment, Ethnicity, New Immigrants, Language skills, Household spending and saving, Body description and functions, Personal resources, Time activity pattern	StatCan 2006 Census, CIHI Health indicators, Quality of Life Reporting System, Socio-Economic Risk Indicators.	Statistics Canada

Details of GIRF classes (page 2)

Class	Subclasses	Relevant frameworks & standards	National data sources
Natural Environment	Land cover and land use, Geology, Soils, Hydrography, Climate and weather, Elevations and landforms.	WHO Health and Environmental Linkages, NRCan land cover legend.	NRCan Geobase, GeoGratis.
Built Environment	Transportation systems, Energy, Agriculture, Recreational water sites, Buildings, Industrial sites, Water supply, Food supply, Solid waste, Wastewater and sewage.		NRCan Geobase, GeoGratis.
Socio-economic Environment	Neighbourhood character, Living conditions, Working conditions, Traffic safety, Crime, Property values, Economic opportunities, Education opportunities, Childcare services, Retail services, Recreation and sports, Arts and culture, Civic engagement.	CIHI Health Indicators, Canadian Index of Well-Being, Quality of Life Reporting System, WHO Commission on Socail Determinants of Health	Generally need very large-scale data.
Geocoding Reference	Core geocoding references, StatCan geographies, Health-related administrative areas, Other administrative areas, Other locational references.		StatCan Road Network File, Postal Code Conversion File, Statcan cross-references to Health Geographies.

Recommendations

1. Support use of a grid-based, "common area unit" for ease of analysis via OLAP, SOLAP, geospatial tools.
2. Support analysis of trends and temporal patterns.
3. Establish a national dataset of health facility locations.
4. Pursue the specification of a socio-economic data product.
5. Use a Wiki approach for the GIRF.
6. Facilitate access to data subsets or special compilations.
7. Define community standards to describe data quality and lineage (within ISO 19115).

HOLONICS

10

5

1   Support use of raster "common area units"

- Brings decision-makers closer to the data, without dependence on complex technology (and technologists).
  - **Analysis of raster data is conceptually simpler, more efficient, and many open-source options exist.**
- Many public health models were designed to operate on a grid, e.g., contagion, kriging, interpolation, autocorrelation.
- Minimizes issues of correlating spatial data aggregated to different regions.
  - **Data integration is done by geospatial IT professionals.**
- Provides new options to safeguard confidentiality.
  - **Each cell is sized to ensure it holds a minimum number of cases.**
  - **Finer grid in urban areas, coarser grid in low-density areas.**
- Supports spatial time series with *relative* ease.
  - **Fewer issues, more efficient than vector-based methods.**

HOLONICS®

11

2   Support analysis of trends and temporal patterns

- The basis of epidemiology is person(s), place and time.
- 37 of the 50 survey respondents need to analyze trends or patterns in space and time using geospatial datasets.
- Examples:
  - Analyze the change in clustering of communicable disease cases over time, including movement of clusters.
  - Analyze changes or trends in contaminant concentrations in air, water, soil, food, etc. in space and time.
  - Analyze how environmental change affect health or health resource usage.

HOLONICS®

12

How to support spatiotemporal needs?

- Longitudinal spatial data exist, but in separate "editions" that have not been compiled into integrated spatiotemporal datasets.
- A simple approach is to build a spatiotemporal "data cube" based on regular or nested cells.
  - **A data cube would support multidimensional analysis via OLAP and SOLAP tools.**
  - **OLAP/SOLAP require minimal technology expertise; the domain expert explores & analyzes the data without an intermediary.**
- Issue: must define the correspondence between attributes of different data "editions".
  - **Same name, different collection parameters?**
  - **Imprecise mappings should be described via metadata.**
- For consistency, geospatial professionals should define the data integration methods to create multi-temporal geospatial datasets.

HOLONICS

13

3 Compile a national dataset of health facilities

- This would support needs of:
  - **National emergency response.**
  - **Comparison of health system performance across Canada.**
  - **Analysis of the success of different spatial models of health service access.**
- Begin with a data model that describes all aspects of facilities and services.
  - **The national data would describe stable information, e.g., location, size, service level(s).**
  - **Provinces and health regions could maintain volatile info (e.g., points of contact, hours, staffing, available beds, treatments).**
- Transform and merge data from provincial/territorial and regional databases.
- Define update strategy as for other national-level datasets.

HOLONICS

14

4 Develop a hi-res socio-economic data product specification.

- Many prospective uses within and outside the public health community.
  - **Non-medical determinants of health are highly related to socio-economic environment.**
  - **Integrated Land Management also needs socio-economic data.**
- Unlike health facilities, many of these data do not exist in any consistent form.
  - **Can be derived from other data (e.g., quality of life, neighbourhood involvement) but requires analysis and discussion.**
  - **Statistics Canada is a core source, but spatial descriptors must be derived from geospatial datasets (e.g., proximity to recreation, education, retail opportunities.)**
- Gridded data could be appropriate--finer in urban areas, coarser in rural areas.

HOLONICS◈

15

Socio-economic subclasses include complex relationships

- Examples:
  - **Air quality: airsheds (movement of air), hazards (pollutant sources and types), monitoring stations, measurements, historic issues with air quality.**
  - **Drinking water quality: water source, storage, distribution, recipients, hazards, monitoring stations, measurements, historic issues with water quality**
  - **Food safety: food production (animal and vegetable), food distribution sources, "foodsheds" to define likelihood of exposure, occupations that involve food handling, historic occurrences of unsafe food.**
  - **Roads: affect the "walkability" of neighbourhoods and traffic safety.**
- Need to model these characteristics as part of dataset specification.

HOLONICS◈

16

5   Use a Wiki approach for the GIRF

- A Wiki is a website for creating, browsing, and searching through information (structured & unstructured).
  - **More interactive, more conversational than a traditional catalogue.**
- Builds collaboration and community.
  - **Users can perform searches, post questions, and suggest revisions to content.**
  - **Users could post new data created via special project.**
  - **Revisions would be approved by a moderator, and published as the accepted version.**
  - **Content evolves organically over time (e.g., Wikipedia).**
- Sample Wiki: http://wikiph.org/index.php?title=Special:Categories
  - **Jumping off point to browse or search for data, find others with common interests.**
- Sample Wiki: http://wikiph.org/index.php?title=Wiki_Public_Health
  - **Illustrates "extras" that can evolve from discussions among users.**

HOLONICS

17

Steps to Wiki creation

- Must be well-designed or it will not be used. Some ideas:
  - **Conceptually based on GIRF categorical scheme.**
  - **Present metadata (managed by data providers) from a user perspective.**
  - **Associate GIRF keywords to authoritative metadata and datasets.**
  - **Wiki may be underpinned by a forms-based application (i.e., a catalogue). Forms would support query/update of GIRF elements, including cross references to data, standards, indicators, authorities, activities, and research.**
- Free Wiki tools include MediaWiki (used for Wikipedia), TikiWiki (open-source), and DokuWiki.
  - **Wiki tools enable the creation of Wiki pages from Word or Excel docs, creation of links within the Wiki, registration of new data.**
- Many services exist to host Wiki sites.
- Need a moderator & custodian from within the community.

HOLONICS

18

6 Facilitate access to custom combinations/subsets of data

- Users need to extract and combine data to address special topics.
  - Ex: *to study recreational opportunities, need lakes with swimming beaches, services offered at each beach, access via public or pedestrian transit, and frequency of prohibitions due to health concerns.*
- Data portals and custodians should offer ETL tools for customized data downloads.
- Consider standard data subsets common to the community.
- Define relationships between GIRF classes/subclasses and data elements. Ex: National Road Network
  - **Roads are in the class "Built Environment"**
  - **Street addresses are in "Geocoding References"**
  - **Speed limit, traffic volume are in "Socio-economic environment"**
  - **Surface type, traffic volume are used to define "Hazard, Exposure, Risk" of air or noise pollution**

HOLONICS

7 Define community standards to describe data quality and lineage (within ISO 19115)

- Quality information should include:
  - **For geocoded data, describe records that could not be geocoded, anticipated fidelity of geocoding.**
  - **For aggregated data, describe the anticipated degree of heterogeneity within the spatial unit.**
  - **For temporal data, describe uncertainty introduced by data integration.**
- Lineage reporting should include references to research papers
  - **Many high-res, special-purpose datasets emerge from special projects (ex: quality of life, neighbourhood amenities) described in detail in research papers.**
  - **A Wiki approach supports external references.**
- Fitness for use could be qualitatively described by comments, additions to the Wiki.

HOLONICS

20