

The Socially Smart Community

Practical Uses of Community-level Socio-economic Indicators

Donald Cowan, Paulo Alencar, Kyle Young
Email: dcowan, palencar, kyoung@csg.uwaterloo.ca
David R. Cheriton School of Computer Science,
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Bryan Smale
Email: smale@uwaterloo.ca
Recreation and Leisure Studies,
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Ryan Erb
Email: tbirtch@socialresearchandplanning.ca, rerb@perthhuron.unitedway.ca
United Way-Perth-Huron,
Stratford, Ontario, Canada N5A 2M4

Fred McGarry
Email: mcgarry@comap.ca
Centre for Community Mapping,
Waterloo, Ontario, Canada N2L 2R5

October 11, 2017

Technical Report CS-2017-06

Executive Summary

There is a large amount of discussion in the literature about smart cities where the focus of the discourse is on gathering and analyzing real-time data from smartphones or other sensors to support public services such as vehicular traffic flow and utility consumption or to infer human behaviour. There does not appear to be any discussion of ‘socially’ smart cities where the focus is on using citizens as ‘smart sensors.’ Here the citizens’ interactions with a community’s services are captured in a timely fashion to derive socio-economic indicators about characteristics of the population relevant to sectors such as education, food security, health, housing, community participation, community safety, income levels and government and to use those as a basis for monitoring community well-being or the effectiveness of government, social service and economic policies designed to produce community improvement.

This paper provides the motivation for and description of a community-level socio-economic indicator system to support the socially smart community. The system will accept timely indicator base data from many different community sources and operate on that data using various software tools and maps. The data can be combined in various ways to show single indicators and relationships among indicators. In addition, multiple layers of data can be displayed on a map showing various geographic relationships. An initial version of this approach and related system to collect community-level social and economic data and display appropriate socio-economic indicators while protecting individual privacy, is being deployed in a mixed urban-rural community in Southwestern Ontario, Canada. The web site can be found at myPerthHuron.ca.

1 Introduction

“Change is not merely necessary to life - it is life.” - Toffler [14]. How do 21st-century communities measure, analyze and react to the constant socio-economic change, adapt their social and economic structures, services and policies to cope with this change and measure the impact of such adaptation?

There is a large amount of discussion in the literature about smart cities, Internet of Things (IoT) and related infrastructure and data gathering [6, 4, 8, 9]. However the focus of this discourse appears to be on gathering and analyzing real-time data from smartphones and other sensors where the data is used in public services such as vehicular traffic flow, impact of weather, tracking population movement to infer human behaviour [9] and utility consumption as well as inferring human behaviour [9]. There is little discussion of management of socio-economic factors.

There does not appear to be any discussion of ‘socially’ smart cities where the focus is on collecting data over time to derive socio-economic indicators¹ about characteristics of the population relevant to sectors such as education, food security, health, community participation, safety and services, income levels and government, and to use those as a basis for determining current community socio-economic well being and the impact of change. Nor does there appear to be any attempt to collect data over time that would allow longitudinal studies to monitor government, social service and business policies designed to produce community improvement. Collecting data over time would provide the information that would assist governments and other organizations in monitoring and assessing the effectiveness of a law, service or policy in a timely manner and possibly indicate directions for change.

Current socio-economic indicators can be used by lower level governments such as municipalities or social agencies and can be examined at the level of the neighbourhood or census tract, but the underlying data is not usually collected in such a way as to reflect a community’s character, values and specific needs. Most social indicator research and implementation uses the census or similar aggregated data to draw conclusions about a segment of society and its interaction with the social system [7]. Such data is collected regularly but often infrequently,² making it difficult to measure change that may happen over a short period of time because of an occurrence such as a plant or school closure or a new local government service or community social policy or service.

Although valuable and respecting anonymity, aggregation of data through a census or similar instrument also hides much critical information about a community. Further, aggregation makes it difficult to measure community changes particularly in smaller communities outside the major metropolitan areas.

This paper provides the motivation for and description of a community-level socio-economic indicator system to support the socially smart community. The system will accept socio-economic indicator base data from many different community sources gathered over time and operate on that data using various software tools and maps. The data can be combined in various ways to show single indicators and relationships among those indicators. In addition, longitudinal studies are possible as the same data is gathered over time and multiple layers of data can be displayed on a map showing various geographic relationships. An initial version of this approach and related system to collect timely community-level data repeatedly and display appropriate socio-economic indicators while protecting individual privacy is called myPerthHuron (myPerthHuron.ca), and is being deployed in a mixed urban-rural community in Southwestern Ontario, Canada.

¹“As statistical time series, social indicators are used to monitor the social system, helping to identify changes and to guide intervention to alter the course of social change.” [5] Work on social indicators is usually performed in academic institutions and large government agencies such as those found at the country or state/province level [7].

²In many countries the census occurs every five or ten years making it adequate for high-level government planning but inadequate for community policy development, planning and monitoring.

2 Monitoring and Modelling a Community

Collecting information about the citizens and their interaction with the community and its support systems and legal framework over time would provide much valuable intelligence and could direct policy and planning by local government, social agencies, NGOs and businesses. Collecting data repeatedly in a timely manner can be viewed as longitudinal analysis or community monitoring, where what is happening in the community can be viewed through indicators shortly after events have occurred. To use an analogy, individuals, governments and agencies are acting as socio-economic sensors, not just obtaining movement data through devices, but tracking interactions of individuals with the community, its government, social and business services.

2.1 Communities as People

People and communities are similar; people's health can change rapidly either because of bacteria, virus or structural change, while a large business closing, extreme weather or medical event can affect the health of a community. Perhaps we can find new ways of looking at community socio-economic health by examining how the medical community manages a person's health.

Apart from periodic checkups, a visit to the doctor is usually because of the appearance of symptoms of an illness. The doctor measures some vital signs, asks some questions, perhaps orders more tests, and makes a diagnosis. Based on the diagnosis an intervention is usually required and a return visit to the doctor may be scheduled to determine the effect of that intervention. Depending on this latest outcome a version of this procedure could be repeated. The key point of this approach is that the data about the patient is gathered repeatedly in a timely manner; the patient's temperature, heart rate or other vital signs are current and not assessed five years ago as might be the case with a census.³

Can we take the same approach with communities where community data is gathered in a timely manner rather than delayed? The approach would be to gather a community's current socio-economic data repeatedly over time, use this data to diagnose the community's current condition, and then propose and implement new approaches to community governance and/or development. As the governance or development plans are enacted, newer community data can be gathered and these plans and structures can possibly be modified. This cycle of measure, diagnose and implement, called adaptive management in the environmental field [1] can be repeated to test the effectiveness of laws, services, policies, various governance rules, development plans and structures and to modify them as appropriate.

Communities through their governments or various agencies pass laws or enact new social service or business policies to create what they believe to be positive change. Creating such laws, services or policies to change a situation in a community is akin to modelling or predicting the future. The group, based on experiences and data from the past makes a change, which they expect will improve the community, but the change could just as easily have unintended consequences [18, 10].

If data about the community were captured in a timely manner and repeated, there would be opportunities to try policies or services and measure their impact in the short-term and modify them appropriately. Thus, a community would have the ability to formulate a policy based on a vision of the future, monitor the results through relevant social indicators and take corrective action as required. With the appropriate tools it is not only possible to describe the current state of the community, but to prescribe changes that would change that state for the better.

Most social indicator platforms are designed to measure the same indicator across all the communities in which the system is applied. Such an approach allows intercommunity comparisons such as average wage, level of education or community engagement. However, each community has a unique character that should also be measured, because this knowledge can be used to benefit and improve the community even though cross-community comparisons are not possible. For example, a community with a strong industrial base has

³A census or other major survey of a jurisdiction such as a country or state may occur periodically such as every five years. The outcome may take quite a significant amount of time before it is published and available for use.

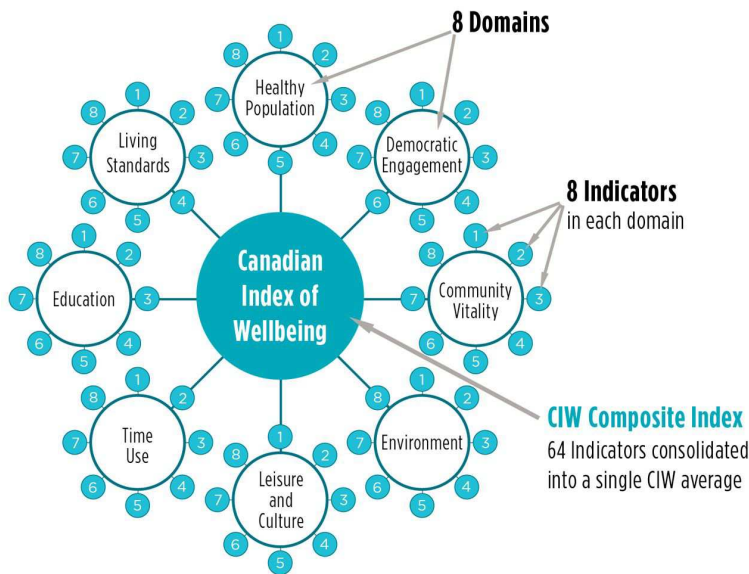


Figure 1: The Eight Domains of the CIW

quite different characteristics than a community with a significant higher education component or high-tech sector. For example, the types of skills needed or socialization requirements in the two types of communities may be quite different.

3 Community Well Being - State of the Art

We have argued that monitoring a community's well-being in a timely manner is important to ensure that the community can respond effectively to various issues and crises that may occur and to ensure that new or modified laws, services and policies may have the expected effect. Such monitoring is possible with current information technology infrastructure, but not necessarily implemented.

What is the current situation with respect to community monitoring of well being? We describe the Canadian Index of Well Being (CIW) [11, 12] as one example of the state of the art. Although similar systems have been developed for other jurisdictions, CIW is relatively new having first been investigated through public consultation across Canada starting in 2000.

Canadians were asked what societal values mattered and what factors contributed to and detracted from their quality of life. Based on these answers a panel of 60 experts and others on social indicators was convened to sift through the answers to focus on domains that were central to overall quality of life. By 2008 eight interconnected domains had been identified as shown in Figure 1. From 2009 through 2011 comprehensive discussions and literature reviews were used to identify the eight indicators that composed each domain. Once the indicators for each domain were computed they were consolidated into a single CIW average.

Details of the various domains of the CIW are provided in Table 1 while the indicators that compose the Community Vitality domain are shown in Table 2. Figure 1 and both tables are from papers by Smale and Hilbrecht [11, 12].

Once a broad array of indicators was chosen, they were narrowed down based on a set of four criteria [12]:

Validity – the extent to which the indicator was directly related to well-being based on compelling evidence in the literature;

Quality – the extent to which the indicator could be derived from credible sources and was easy to define and understand;

Domain	Description
Community Vitality	Vital communities are those that have strong, active and inclusive relationships among residents, the private and public sectors, and civil society organisations relationships that promote individual and collective well-being.
Democratic Engagement	Engagement democratically is the state of being involved in advancing democracy through political institutions, organisations, and activities. A healthy democracy requires ongoing democratic engagement both during and between elections.
Education	Education is the systematic instruction, schooling, or training given to the young in preparation for the work of life, and by extension, similar instruction or training obtained in adult age. Societies that thrive encourage that thirst for knowledge at every age and stage.
Environment	The environment is the foundation upon which human societies are built it is the basis for our health, our communities, and our economy, and is the source of our sustained well-being. It involves our efforts to prevent waste and damage, and to revitalise the quality and sustainability of all our resources.
Healthy Populations	Healthy populations considers not only the physical, mental, and social well-being of the population, but also examines life expectancy, lifestyle and behaviours, and the circumstances that influence health as well as health care quality, access, and public health services.
Leisure and Culture	By participating in leisure and cultural activities, whether arts, culture, or recreation, we contribute to our well-being as individuals, to our communities, and to society as a whole. As forms of human expression, they help to fully define our lives, the meaning we derive from them, and ultimately, our well-being.
Living Standards	Living Standards considers the level and distribution of income and wealth, with particular emphasis on poverty rates, income volatility, employment, economic security, and work-related issues and outcomes.
Time Use	Time use measures how people experience and spend their time, and how the use of our time our affects well-being. The implicit assumption is the notion of balance, and therefore considers the length of our work week and our work arrangements, our levels of time pressure, and the time we spend in leisure and volunteerism.

Table 1: Domains of the Canadian Index of Wellbeing

Component <i>Sub-component</i>	Indicator
Social Engagement	
<i>Social and civic participation</i>	Percentage of population reporting participation in organized activities
	Percentage of population who provide unpaid help to other living on their own
<i>Social support</i>	Percentage of population with six or more close friends
<i>Community safety</i>	Crime Severity Index
	Percentage of population that feels safe from crime walking alone at night
Social Norms and Values	
<i>Attitudes towards community</i>	Percentage of population that feels most or many people can be trusted
	Percentage of population that has experienced discrimination in the past five years
	Percentage of population reporting a very or somewhat strong sense of belonging to the community

Table 2: Well-being indicators for the Community Vitality domain

Reliability – consistency in the way in which the indicator has been measured in different years (e.g., consistent question wording across survey years);

Feasibility – the extent to which data were available and relatively easily accessible.

The final decision on which indicators to include was based on practical issues [12] such as the consistent and credible availability of data provided by agencies such as Statistics Canada.

3.1 Community-level CIW

Based on the request of several community-level organizations the CIW framework has been modified to apply at this level of granularity [12]. Questions from the national survey that applied at the community level such as “How many close friends do you have?” were adapted appropriately. Second, where national-level questions were not available or appropriate, alternative questions were derived from the literature. Finally, when national questions or questions from the literature were not available, new domain-related questions were created.

Once the survey was created, a sample community population was selected that was representative and large enough to permit analyses of sub-groups and areas. This sample was given a self-administered questionnaire and the sample was selected using the latest Canadian census. It should be noted that three different partners adopted the survey for their community namely: a municipality, a community foundation and a consortium of community partners.

4 Why Collect Community-level Data - Some Examples

Why do we need social indicators that characterize a community? In this section we present two simple examples to illustrate the power of the concept. The first example is from 1854 and the second from the present.

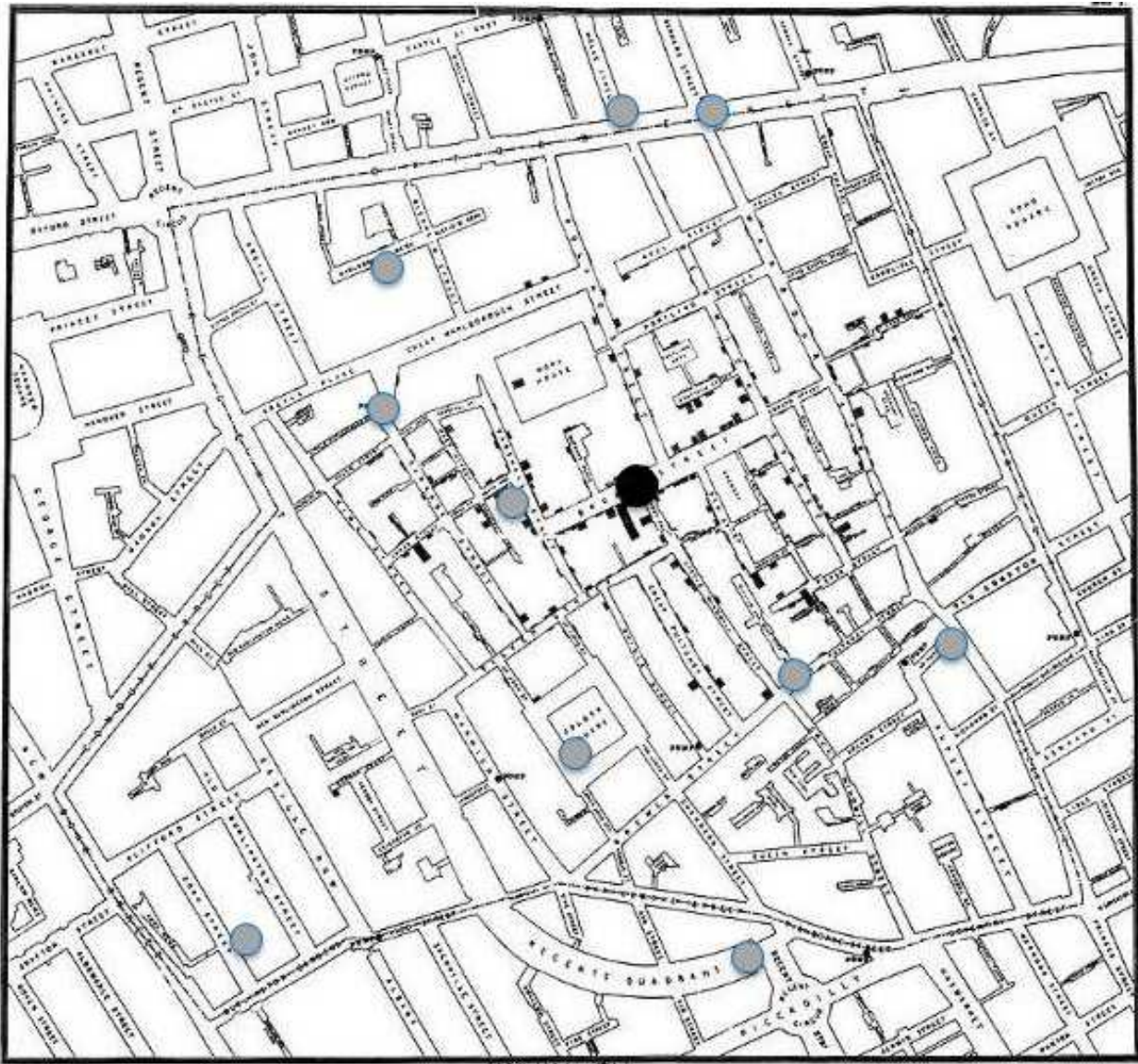


Figure 2: Cholera Map for Soho in 1854

4.1 A Look into the Past - the 1854 Cholera Epidemic

In 1854 there was a cholera epidemic in the Soho district of London because of a lack of water treatment and dumping of raw sewage into the Thames [16]. The apparent source of the infection was found to be the Broad Street pump and the handle was removed thus preventing people from accessing that water.

Although the procedure to isolate the pump was more complex than described here, the principle is still the same. First Dr. John Snow drew a map showing the location of the pumps in Soho as shown in Figure 2 and depicted by the large dots. Then Dr. Snow plotted all the cases of cholera on the map and noticed that many were clustered close to the Broad Street pump shown as a black dot.

A number of cholera cases were closer to other pumps, but when Dr. Snow checked those houses, he found that almost all the people involved used the Broad Street pump because they liked the flavour of the water. The Broad Street pump was disabled by removing its handle and the cholera epidemic disappeared quite quickly.⁴

⁴There is an open question whether the cholera epidemic was already on the wane when the pump was disabled. For our purposes of showing the need for community-level data, it does not matter.

How does this cholera case apply to community-level data such as we intend to collect? First the census in Great Britain operated in 1851 and 1861, so there would be no data available for this epidemic. Even if the census had just occurred, the type of data required would likely not have been available. The number of pumps and number of cases of cholera in Soho might have been known, but not the location, which was critical to solving the problem. However, providing location can be considered an invasion of privacy although where public health is concerned that may not be relevant.

4.2 A Look into the Present - Access to Food

The following two examples illustrate how community-level data can change access to food based on the type of data made available.

A study [15] from the Institute of Urban Studies at the University of Winnipeg illustrates an example related to food security and community-level information. There are several different ways to examine food security.

One can look at accessibility of grocery stores. In other words where do people live relative to the food supply? Can they easily go to a grocery store and obtain food? Of course one can also look at income. Even if grocery stores are close, can the individuals in the neighbourhood of a store afford the food?

It is well documented [3] that eating a nourishing breakfast is important for student achievement in school. Recognizing this fact, education authorities in many jurisdictions offer breakfast programs in schools where average income is below a certain threshold. The belief is that higher income households can afford to provide a nourishing breakfast, whereas those with lower incomes are not likely to have the resources. There are at least two basic assumptions behind the operation of these breakfast programs, which may be false, namely that:

- higher income families can take the time to produce a nourishing breakfast, and
- there are no students from lower income families in a school with a higher income ranking.

In order to determine the location of grocery stores, student access to breakfast, people's income and location, or other similar community relationships far more local or community-level information is needed. This kind of information is not provided by higher-level government surveys such as the census that aggregates information.

4.3 Community-level Indicators are Different

Community-level indicators are certainly different from those provided by current approaches. These examples illustrate many of those differences. Indicators such as those in the CIW may be initially identified through citizen consultation, but they are narrowed down, compiled and validated by experts, whereas community-level indicators are often identified and derived by community-level organizations and individuals. Thus, the difference may be in the choice of indicators and the detail provided.

Once an indicator has been chosen, the user and the supplier of the data must realize that the indicator's validity relies on the credibility of the supplier of the data and a consistent approach to collection and reporting. Of course validity could be monitored through audits of the data suppliers. In creating community-level social indicators, care also needs to be taken on providing access to certain kinds of information. This topic will be discussed more in Section 8.

5 Community Data Identification and Collection

Although the rationale behind collecting and using community data is provided in Section 4, it is still necessary to find a way to identify what data should be collected and then collect that data effectively. There are three approaches that can be used and these are briefly described next.

5.1 Identifying and Using Community Expertise

Community experts - individuals or groups within a community who understand a community's data needs and can compile a view that is appropriate.

External experts - individuals from outside a community who have a general view of community-level data needs. This approach is the one used by the Canadian Index of Well Being with initial input from a survey of citizens as described in Section 3. Here they used the country wide index developed by a panel of experts and modified it to fit their perception of community-level data needs.

Both community and external - this approach represents a combination of the two previous methods.

Ideally one should combine both community and external experts as they each can bring different perspectives to the problem of identifying the community data that is most valuable. External experts may choose different sets of indicator data as they are often interested in comparing communities, whereas local experts are more interested in choosing data to derive value for their specific community. For example, in the cholera epidemic in Section 4 the local people were interested in the location of the pumps in order to combat a local outbreak of disease, while outside experts might be more interested in overall water supply or number of pumps per inhabitant.

However, the ideal situation of combining both community and external experts may be difficult to make operational because of the difficulty of identifying external expertise and associated costs. However, over time it may be possible to distill a standard expert view of communities perhaps based on size of municipal population or similar grouping criteria such as business demographics. For example, the community with which we are currently working has used the Canadian Index of Well Being as described in Section 3 to guide them in their choice of indicator data for their Quality of Life Report [13], which is currently published every four years.

Because of the difficulties just outlined, we have chosen to use community experts only, modified by the experience gained in producing earlier reports, as they now have the in-depth knowledge of community data and community needs, and therefore can identify what data should be collected and assembled.

How are these community experts identified and assembled into a useful working group? An essential component of identifying and acquiring community experts and data is to work with a community-level party that is trusted across the community such as a community foundation or a charity that has community-wide impact. In addition such a party should have extensive community knowledge and connections. One example is the United Way, which is usually recognized as community-based and trusted by members of the community. In addition, the United Way is a collection point for funds and requests for funds, therefore it has strong identifiable community connections. One can make similar comments about community foundations. Having such a group involved is essential to support the notion of community engagement and trust.

5.2 Identifying and Collecting Community Data

Once the high-level community partner has been determined, then a process must be defined and put into operation to identify appropriate community data and decide how that data should be collected. Such a process should be inclusive in that community organizations and even individuals should not feel ignored. There are many local government departments and formal and informal community agencies and groups that could supply data to a community trends platform, some of which are identified in Figure 3. Formal groups could be the social service agencies, government departments, organized charities, educational authorities and public health units, where informal groups might be food banks, sports groups and churches.

A method that is being tested is to establish small focus or stakeholder groups representing the community organizations and individuals that can and want to deliver data. Each group has to identify the data to be supplied based on the questions being posed.

Invitations are issued to attend stakeholder focus meetings where attendees can identify and prioritize the type of data that is required and name its source. This list is compiled and a partial sample of the type

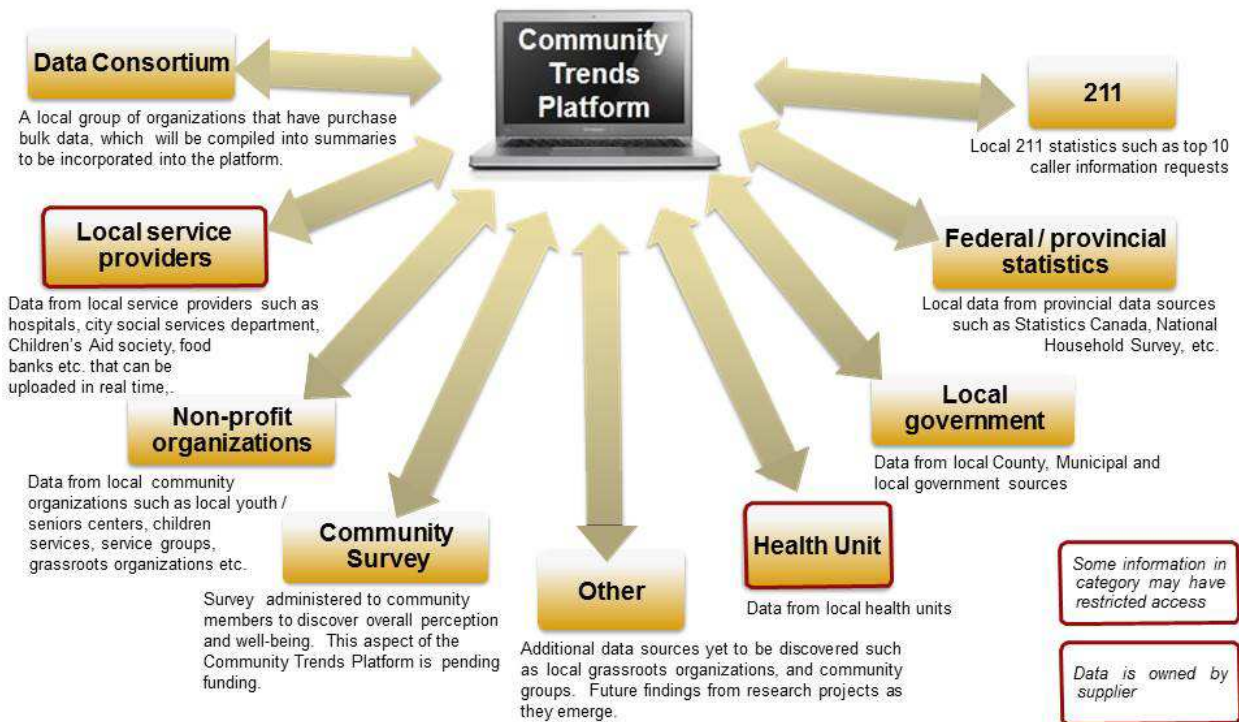


Figure 3: Community Trends Platform Flow Chart

of social indicator data identified is shown in Tables 3 through 7. Once the data is characterized the trusted community-level party can start contacting community organizations to provide the data. Initial data will likely be uploaded in bulk, whereas incremental additions to data will likely be through online forms or other automated processes [2].

Collecting and maintaining data in mixed urban and rural communities can be a problem as the local government and many of the community agencies may have a small staff complement or are run by volunteers with limited time commitment. Hopefully this lack of human resources can be addressed by a combination of temporary assistance and software design. An experienced data analyst could be hired to assist governments, agencies and NGOs in identifying, organizing and downloading the initial data sets. Of course technology could be used to automate the process of downloading the initial data and keeping data current. This process may require installation of foreign components into a local information system, a process that requires a level of trust, something that should be possible.

5.3 Combining Data to Expose New Relationships

Of course there is the possibility that groups may want to combine data in order to answer cross-disciplinary questions such as the relationship between poverty and health based on location of subsidized housing or income distribution. In addition, new questions will arise as experience is gained. Thus, the community-level party must maintain a stakeholder structure that supports modifications and additions to the original scheme of data generation. Further the underlying information technology implementation must support the addition of different data types as the community system expands. Every group that provides data should also sign a data sharing and display agreement to indicate that they know the potential privacy violations that can occur when sharing data. In fact, there should be a mechanism in place whereby they are not able to share data without some form of data-sharing agreement and related mutual permissions including what can be presented and to whom. A proposal to manage data sharing is provided in Section 8.

Domain	Indicators
Affordability Benchmarks	Income-based rental benchmark
	Market-based rental benchmark
	Income-based ownership benchmark
	Market-based ownership benchmark
	Percentage of new residential construction that is affordable to low and moderate income households
Core Housing Need (affordability, adequacy and suitability)	Shelter cost-to-income ratio (percentage of households paying more than 30% of income on housing)
	Percentage of households that require major repair
	Percentage of household living in overcrowded housing according to the National Occupancy Standard
	Percentage of households in core housing need (by tenure)
Density	Average # of people per hectare
	Average # of people and jobs per hectare
	Ratio of people to jobs
	Ratio of land in residential use
	Average # of residential dwellings per hectare
	Average # of people per household
	Percentage of new residential units constructed in the “built-up” areas
Eviction Prevention	Usage of Rent Bank programs
	Usage of Emergency Energy programs
Homelessness	# and type of emergency shelter beds (resources)
	Number and profile of sheltered and unsheltered homeless
	ALOS in shelters
Household size and composition	Percentage of one person households
	Percentage of lone parent households

Table 3: Housing and Homelessness - Part I

Domain	Indicators
Supply - Private Market Ownership	Number and mix of occupied dwellings (single-detached dwellings, semi-detached dwellings, row and town houses and apartments)
	Number and mix of new residential dwellings that have been completed each year
	Number and type of residential dwellings that have been demolished each year
	Number of rental units that have been converted to condominiums each year
	Average resale prices (by structure type)
	Average (or median) value of dwellings
Supply - Private Market Rental	Vacancy rates (by unit size)
	Average rental rates by unit size
	Average amount spent on rent + utilities per month per renter household
	Number of new accessory apartments registered each year
Subsidized housing	# of subsidized housing units (permanent v. temporary)
	# of supportive housing units (permanent v. temporary)
	Percentage of population that lives in subsidized housing
	Percentage change in the number of active households on the centralized waiting list (by unit size, household type and geography)
	Average waiting time for social housing (by unit size, household type and geography)
	Ratio between housed households and new applicants (by unit size, household type and geography)
	Trends in the number of rent supplement (housing allowance) units
	Building condition audits
Tenure	Ratio of owners to renters
	Age of primary household maintainers by tenure

Table 4: Housing and Homelessness - Part II

Domain	Indicator
Income	After-tax median income of economic families
	Ratio of top to bottom quintile of economic families, after tax
	% of persons in low income (all persons, children, seniors)
	Gender wage gap
	% of population who are working poor (not earning a living wage)
	% of companies paying a living wage
	Wage rates by industry and occupation
Financial Assistance	OW rates
	ODSP rates
	EI rates
Labour Force	Unemployment rate (%)
	Long-term unemployment rate
	Youth unemployment rate
	Full-time employment rate
	Part-time employment rate
	Self-employment rate
	% working multiple jobs
	Employment rates (# of jobs) by industry and occupation
Economic development	# of new businesses
	# of business closures
	Retail sales
Labour Mobility	# of people leaving area for work (by age groups)

Table 5: Income, Work and Economic Security

6 Analysis of Community Data

Once the community indicator data is identified it becomes necessary to decide how the data should be analyzed. The forms of desired output determine the analysis and types of tools required. The next subsection describes some types of analysis that are being done or might be done with a view to providing clues as to the tools that should be provided to the data analyst or other users of the system.

6.1 Geography and Other Data Patterns

“Social indicators lend themselves to geographic disaggregation and analysis by successively smaller areal units: nation, region, state, subregion, district, or county. Cities, similarly, are divided into ecological areas through use of divisions such as census tracts, blocks, neighbourhoods and school districts. Such detail makes possible comparison of trends of ecological units, and opens the analytical possibility of interrelating social with physical and biological indicators. This yields insights into the influences of population, social organization, technology, and environmental effects upon a public problem or concern.” [5]

As can be seen from the quote, geography plays an important role in social indicators. Do different income levels live in certain areas of a community? What about level and distribution of educational attainment? Do certain health patterns when combined with other indicators emerge based on local geography? These are only many such questions that arise as one looks at social indicator data at a community-level based on geography.

As indicated in [5], social indicators tend to be examined in isolation. It is not often that they are combined in order to determine interactions. However, as also noted in [5], “poverty, unemployment, poor health, poor housing, crime and poor educational standards are not independent variables. Such indicators of deficiency, are not always associated nor linked in ways that early ecological studies led us to believe.

Domain	Indicator
Arts and Culture	participation in arts and culture events (not an organization)
	attendance at cultural events
	employment in cultural industries
	Municipal spending on all aspects of arts and culture as a percentage of total municipal expenditures
	Household expenditures in past year on all aspects of culture and recreation as a percentage of total household expenditures
Physical Activity	Leisure-time physical activity; moderately active or active (%)
	% of school-aged children walking to school
	Household spending on sports and recreation as a percentage of total household expenditures
	% of population using gyms, walking clubs, running clubs, etc. on a regular basis
	# of free physical activity programs
	# of children/youth receiving subsidies for sports and recreation programs
	Municipal spending on recreation and parks as a percentage of total municipal expenditures
Usage of local recreation and cultural facilities	
Library	% of population that are active patrons
	Library use (as measured by circulation/types of loans)
Tourism	Amount of time visitors spend here
	# of visitors to festivals
	# of tourists
	Proportion of residents planning a vacation in coming year
	Average number of days on holidays in previous year
Volunteering	Volunteer hours associated with minor sports activities
	Volunteer hours associated with cultural activities
Faith	% of population attending church/religious events on a regular basis
	Charitable giving rate for faith groups

Table 6: Leisure and Culture Indicators

Domain	Indicator
Transportation (Getting Around)	Median commuting duration
	Mode of transportation to work
	Mode of transportation to school
	Amount of time spent taking kids to extracurricular activities
	Amount of time spent getting to community activities, medical appointments, etc.
Time Spent Working for Pay	Mean hours worked in pervious week
	% of residents working over 50 hours per week
Time Spent on Unpaid Work	Hours spent looking after children, without pay
	Hours spent doing unpaid housework
	Hours spent looking after seniors or adults with disabilities
	Hours spent on formal volunteer activities (for a non-profit organization)
Time Spent on Social and Leisure Activities	Average amount of time respondents spent sleeping
	Average amount of time spent on screen time (social media)
	Average amount of time spent on social activities
Life Satisfaction	% of population who are satisfied or very satisfied with life

Table 7: Time Use Indicators

Crime rates are not highest in poorest urban areas. Unemployment is not highest in areas of lowest educational attainment. Age-specific death rates are not highest in the areas of poorest housing. The largest suburban agglomerations show the greatest concentrations of poverty, but they are not the poorest, and their public fiscal effort may well be high. The highest divorce rates do not occur where there are most children in one-family households. Yet this quote about relationships among social indicators indicates many possibilities that need to be examined in order to dispel our prejudices and understand how society operates in a community.”

There are many possible combinations. What is the geography of a community in relation to education, food security, health and income levels? How do income levels relate to food security or volunteering? Are poor people less healthy than those with a higher income? There are just so many possibilities to explore.

In addition the relationships among social indicators are not what they seem. If these statements are true in a large urban setting, what are they like at a more refined community level? What biases can be dispelled about a community? These are questions that need to be addressed.

In other words, community-level data is collected in silos and often not effectively used within those sectors because it is frequently difficult to share. If the data is not used when it is collected in specific sectors, one can imagine that sharing across silos rarely ever occurs because of the difficulty of sharing and the possibility of revealing data that might be a significant invasion of privacy. However this form of sharing is extremely powerful and can be of great value when trying to assess a community’s well-being and plan through the implementation of future laws and policies. Such sharing is possible as long as the conditions related to who can share the data and who can see the results are stated and followed in a data-sharing agreement.

6.2 Types of Analysis and Output

From the discussion and quotes provided earlier in this Section, it is clear that there are a number of tools required to analyze the community indicator data. These tools include:

- the ability to access multiple datasets from different sources to display a single combined dataset;
- reports in the form of lists;

- tools such as graphs and charts to show data relationships;
- statistical tools to manipulate data; and
- maps to show geographic relationships of multiple data layers and to support geographic queries.

Many uses of community indicators involve combining datasets provided by groups offering different services such as public housing, education, food banks, public health and police services. Combining datasets may reveal relationships that are not apparent otherwise. However, combining such datasets and displaying the results must be carefully controlled so as not to violate personal privacy. How such data is shared is described in Section 8. In order to share and display data each dataset must have some form of access control so only those allowed to manipulate such data may use, examine the data, and present the results.

Data needs to be queried and displayed in many different forms. Data can be presented in lists or charts and graphs or as multiple layers on maps. Data can be manipulated using various statistical techniques to produce values and graphs such as averages, medians, standard deviations and co-relations. Maps can also be used to query datasets. For example, show all the individuals in a certain area who use a food bank.

6.3 Combining Community Level Data to Provide Insights

In the previous section there is a discussion of using different tools such as maps to combine datasets to reveal novel community relationships. What questions might be asked that would offer new insights? Data can often be combined and sometimes mapped to reveal insights into a community. Conclusions are often drawn from correlations of two or more types of data, not from a causal relationship and so one has to be very careful. This section explores a number of such questions to start to explore the concepts and indicate the breadth of possibilities.

The reader should notice that each application will likely require the sharing of community-level data among two or more organizations. Such sharing often involves confidentiality both in who can share the data and in determining who can see the results. Often such sharing requires the completion of a data sharing agreement. Section 7 explores how this data sharing might be implemented.

6.3.1 Parking and Loss of Business

A small city raises its parking rates for street and lot parking in the core of the city. Once these changes go into effect, the businesses in the city core notice that customer traffic is decreasing and over time the revenue and hence the bottom line of most business is dropping. The local core business development association that tracks all businesses in the core raises an alarm, as it notices this problem affects almost all businesses in the core based on a survey.

This action of changing rates appears to be encouraging shoppers to travel to malls on the outskirts where parking is free, thereby reducing traffic in the core. This problem is brought to the attention of city council that has to decide how to deal with it. Eventually they decide to remove all parking charges from the city core. However, they along with the business association have to develop a strategy to lure customers back to the core.

This simple example illustrates the insights that can be gained by combining data from more than one source to come to some conclusions and take action. Note that it was assumed that the increase in parking fees in the core was the cause for the drop in business. The cause could be other things that were not mentioned in the problem statement such as massive road construction to replace infrastructure or to install a rapid transit line or new bus routes.

6.3.2 Youth Crime and Recreation Facilities

It seems to be a commonly held belief that youth crime and lack of recreational facilities are related, that is, there is more youth crime where there are little or no recreational facilities. How might a relationship be examined?

The police department could supply a data set showing each incidence of a crime committed by someone under the age of 18, the location of the crime and perhaps the home address of the perpetrator. The city government could provide a data set of all recreational facilities directed at youth. These two data sets could be plotted on a city map where one can see if there is any apparent relationship between the scene of the crime and the location of recreational facilities or the address(es) of the perpetrator(s) and the facilities. Based on these observations, the city authorities might investigate further.

For example, the city might carry out an experiment by installing new facilities in a neighbourhood where crime is particularly bad and see if this has any impact on the youth crime rate. Another possibility would be to make it easier for youth to access existing facilities by providing some form of subsidized transportation. Such an approach can be viewed as a natural experiment [19] and is a way to determine if the two concepts not only appear to be related, but actually cause one another.

6.3.3 Use of 211 Social Services

Governments are often looking for ways to help people find human services. 211 is a telephone helpline and website that provides information on and referrals to community, social, health-related and government services. The vision for 211 is to be the primary source of information and gateway to human services for individuals and planners.

All the calls to 211 are logged with type of call and its geo-coordinates recorded. It would be interesting to know what type of people make 211 calls. Are most of the people who use 211 from low income situations?

One way to examine such a question initially for a community would be to map income against origin of 211 calls. The location of public housing is a reasonable way to measure income as public housing supported by government is available based on income. Although public housing does not cover all people living on a low income, it is a reasonable proxy.

Thus, a map of location of public housing versus location of 211 calls and the percentage of calls or types of calls from public housing might offer a clue as to problems facing lower income people. Thus, for government planners and other decision-makers, 211 when mapped against other information may provide rich data about caller/user needs that will help inform their investment and policy decisions regarding social, health and government services.

7 Governance

A governance model for gathering and using community-level data is evolving. Based on the experience gained so far, a few principles are clear.

- Privacy of the individual and security of the data should be paramount;
- Gathering community-level data should be governed by the community and data suppliers;
- Suppliers of data own the data and can remove that data from availability at any time;
- Organizations that wish to share data must respect the privacy of the individual; and
- The consumers of the data must have confidence in the results.

7.1 Privacy and Security

Community-level data are very resolute and can often be used to identify individuals. If such data is used to pinpoint that someone plays on the local football team because participation in sports is shown by neighbourhood, the individual is not likely to be upset. However, if data about salaries is used and someone is identified, then a serious privacy breach has likely occurred.

Thus, the suppliers and users of data, particularly when datasets are combined should be very sensitive as to how the data are to be used. Specifically, they should be clear about the audience for the results and whether the results should be accessible to the public as discussed in Section 7.3.

Security is also an issue. If the data can be accessed by unauthorized parties, then privacy is likely to be violated. The platform needs to be designed to prevent unapproved access to data, particularly multiple datasets that when combined can often reveal personal details.

7.2 Community Governance

Gathering community-level data should be governed by the community and organizations and individuals that supply data. A steering committee should be established that is representative of the community and oversees the data suppliers and acts on their behalf.

Violations of privacy should be reported to the steering committee who should decide on the appropriate action. One consequence is that suppliers of data own the data and have the right to remove that data from availability at any time.

7.3 Sharing data

Organizations that wish to share data to gain new insights into the community must sign data sharing and presentation agreements that respect the privacy of individuals and thus determine who can view the results.

7.4 Reliability of Data

Data and the results from the data must be truly representative of the community and must be deemed to be reliable so that consumers of the data may have confidence in the results. Therefore, we should ensure that the data and data suppliers are using appropriate processes for gathering and distributing data. There should be an audit function that can examine data, data sources and associated collection and distribution processes.

8 System

Up to this point the paper has described why community-level data and derived indicators are desirable. This section describes how the base data might be catalogued, collected and made accessible, once identified.

Some of the data that is collected such as sports team schedules, number of teams or team composition (male or female) is probably not sensitive and publishing this data would not constitute an invasion of privacy. However if one started to publish public health information, use of food banks, public housing residency or police information it is quite possible that privacy violations could occur. Such occurrences are very likely in a small community or a city neighbourhood where it is possible to determine identity by the granularity of the information or location.

How is it possible to make information accessible to appropriate agencies and support sharing among agencies without violating privacy?

A community-based system can have a two-level structure. Data can be:

- Public where the possibility of invasion of privacy is non-existent even when data sets are combined,
or

- Published but protected where parties who want access to the data can only have access if they agree to a restrictive data sharing and display covenant.

Though not meant to be exhaustive, we describe how such a data and display covenant might operate. There are a number of steps.

1. A data sharing and display agreement is signed among all parties who will be supplying or using data setting out the conditions under which the data becomes available and can be displayed.⁵ Hopefully a generic agreement can be developed.
2. The data is uploaded to a cloud platform and secured as described in the next two steps.
3. Access to the data can be provided through a two-factor or two-step verification process [17] where users must supply two different components to identify themselves.
4. In order to provide an extra degree of security the data can be encrypted where sharing of keys is limited.

If two or more government agencies or NGOs wish to share data then the data sharing and display agreement is signed, the data is uploaded if not already available and two-step credentials and cryptographic keys are exchanged.

Besides a set of tools for manipulation and analysis of the data, the system must keep the data current. As soon as data is published by an agency that data should appear in the cloud platform.⁶

There are a number of steps that must be taken to ensure that the data in the cloud platform are current. Scripts⁷ must be written that copy the original database to the cloud platform. These scripts must query the original database on a suitable schedule and determine if any new data has been added. If this is the case then the new data is copied. Another approach would allow updates to the original database to be copied to the cloud platform as they are being entered [2].

9 Sustainability

Once the research project called myPerthHuron.ca is complete it is intended that this social indicators platform will continue to operate in Perth and Huron counties and hopefully also be made available in other jurisdictions in Ontario and beyond. In order to ensure that the social indicators platform continues to operate and possibly expand after the research funding ends, a sustainability plan is necessary.

9.1 Factors Supported by Sustainability

The sustainability plan must consider several factors in order to ensure adequate functioning of the platform including:

- Working with one or more local partners to identify and connect with community organizations and local governments that can and should contribute data to the platform.
- Ensuring that the local partners have adequate resources to sustain the partnership.
- Enabling continuing data generation from the organizations and local governments identified as data suppliers.

⁵It will be very difficult if not impossible to decide in advance what data can be combined and displayed and the composition of the audience for the data. One way to solve this problem is to limit viewing initially to the approved members of the agencies that are combining the data. Broader distribution of viewing privileges will have to be decided by an approval mechanism similar to an ethics committee in a university or hospital.

⁶Of course if the data is not stored directly in the cloud platform there should be an announcement in the cloud that a dataset has been updated.

⁷Generation of these scripts can and should be automated as manual creation of such scripts is prone to error.

- Ensuring that the data is owned and controlled by the groups that generate the data.
- Ensuring that the data from the data suppliers is kept up to date in a timely fashion while trying to automate that process wherever possible.
- Assisting with data gathering if needed.
- Operating an audit service to ensure that supplied data meets adequate standards.
- Providing a data access and analysis service for all governments and organizations that require such a facility, while specifically supporting organizations or groups that do not have access to such a service. Many community organizations are based on part-time volunteers and are not able to afford this service on a permanent basis, but could use some support particularly when seeking external funding.
- Keeping the software for the social indicators platform maintained and operational at all times.

For each community being served, this social indicators platform will require personnel focusing on identifying and nurturing suppliers of community data, data analysis and data auditing, as well as the funds for software maintenance and a cloud computing environment to support storage, manipulation and presentation of community data. How can revenue be generated to pay for the afore-mentioned services?

9.2 Revenue Generation

Although governments, foundations and other grants are a source of revenue, they are often not reliable in the long term. Priorities of these groups often change as new needs are identified. Therefore, some other form of sustainable revenue should be considered to fund the long-term operation of the social indicators platform.

The data collected in a community, particularly when aggregated, can have great value for many data-consuming organizations including businesses, consultants, municipal, provincial and federal governments, economic development agencies and foundations. Similarly being able to identify data sources who could form focus groups to address specific issues can also have significant value.

The data-consuming organizations could pay for data and group access, and analysis services to generate reports about the community and its activities. For example, a business such as a large retailer, manufacturer or service bureau may want to know about spending and traffic patterns or availability of skilled labour or housing before entering a community. In contrast a government will want to know about the impact of new services and policies and an economic development agency will want to use the data to achieve its objective of retaining and attracting new business. Fees for these services may depend on how much the organization needing these data services is already directly engaged with the community and supplying community-level data.

Thus, we propose establishing a revenue-generating structure, which will be an interacting set of organizations to provide the services just described as well as other services related to community data as they are identified. Revenue generated by these organizations will be used to support the social indicators platform for each community. The generated revenue will be used to support the services described in the list in Section 9.1 as well as providing other revenue to encourage community growth. The revenue-generating structure for the social indicators will consist of three components:

1. a service organization;
2. the complete group of organizations and individuals (the “Data Community”) that supply data to the community social indicators platform; and
3. a steering committee acting as representatives of the Data Community.

9.3 Principles of Operation

The three components of the revenue-generating structure will work together using a set of actions based on principles that work toward maximizing community participation, benefit and privacy.

The service organization will:

- maintain and modify the software for the social indicators platform including the data infrastructure;
- provide access to the data by parties both internal and external to the community providing the data based on data-sharing agreements;
- provide data auditing services for the community; and
- provide data analysis services for the community and parties external to the community.

Each organization or individual that is a member of the Data Community will:

- collect and provide data for the social indicators platform following accepted standards based on their own internal mandate;
- make their data available for analysis often through data sharing agreements; and
- work with other members of the Data Community to combine data for analysis with the support of data-sharing agreements.

The steering committee represents the Data Community and will be elected by the members of the Data Community. The number of members of the steering committee and the terms of the members is determined by each member of the Data Community and the members are selected from the Data Community. Each member of the Data Community has one vote in the election of members to the steering committee. The steering committee will:

- set standards for sharing community data both internally and externally;
- determine with the advice of the Data Community what data can be shared internally and externally;
- develop data sharing agreements among Data Community members internally and with the service organization externally; and
- negotiate any payment that might be made between members of the Data Community and the service organization for external use of data.

Once established in a community the steering committee will operate under a set of principles some of which follow:

- The steering committee will govern what individuals and groups can join the Data Community and what data and analysis can be made public. Most of the other rules governing composition and operation of the steering committee have yet to be determined.
- Any group or individual belonging to the Data Community and wishing to supply community data to the social indicators platform must notify the steering committee of its intentions. The steering committee can stop such participation but must actively do so.
- There will be no cost to any individual or group that belongs to the Data Community and is adding community-level data to the social indicators platform.

9.4 Principles of Revenue Sharing

The service organization generates revenue by providing access to data and data analysis services to organizations both inside and outside the Data Community. Once the basic expenses of operating the service organization are covered then the Data Community should share in the revenue as the Data Community provides the data. Revenue sharing can be achieved by providing services to members of the Data Community and funding to the steering committee that can be used to operate the steering committee and distributed to members of the Data Community. Distribution could be in the form of grants or payments for services.

10 Conclusions

This paper provides a rationale for timely collection of community-level data that underlies social indicators and outlines an information system to collect, manage and display the social indicators. This approach allows communities to monitor community well-being when it is happening and to track the effects of laws and social policies. Further the paper outlines a financial model that aims to provide sustainability for the social indicators process.

This method of gathering community data contrasts with current practice, which typically uses data collected from government agencies or private companies that perform surveys such as the census or collect community market data. This data typically is from the past and presents an “old” and incomplete picture of the community.

11 Acknowledgements

The authors wish to thank the United Way of Perth Huron and the Ontario Local Poverty Reduction Fund for their generous financial support of the project described in this paper. Thanks are also due to the members of the Community Trends Steering Committee of Social Research and Planning Council of Perth Huron for their oversight of and input to the project.

References

- [1] Integrated water resources management.
- [2] P. Alencar, T. Oliveira, D. Cowan, and D. Mulholland. Towards Monitored Data Consistency and Business Processing Based on Declarative Software Agents. In A. Garcia, C. Lucena, and et al., editors, *Software Engineering for Large-Scale Multi-Agent Systems Research Issues and Practical Applications*, volume Lecture Notes in Computer Science (LNCS), vol. 2603, pages 267–284. Springer, 2003.
- [3] Australian Red Cross. Why is breakfast important for students? [http://www.redcross.org.au/files/Fact_sheet_4_\(2\).pdf](http://www.redcross.org.au/files/Fact_sheet_4_(2).pdf), 2011.
- [4] L. Carvalho. Smart cities from scratch? A socio-technical perspective. *Cambridge Journal of Regions, Economy and Society*, 8:43–60, 2015.
- [5] A. L. Ferriss. The uses of social indicators. *Social Forces*, 66:610–617, 1988.
- [6] R. Khatoun and S. Zeadally. Smart Cities: Concepts, Architectures, Research Opportunities. *Communications of the ACM*, 59(08):46–57, August 2016.
- [7] K. C. Land, A. C. Michalos, and M. J. Sirgy. *Handbook of Social Indicators and Quality of Life Research*, chapter Prologue: The Development and Evolution of Research on Social Indicators and Quality of Life (QOL), pages 1–22. Springer Science+Business Media B.V., 2012.
- [8] G. Mone. The New Smart Cities. *Communications of the ACM*, 58(7):20–21, July 2015.
- [9] A. Pentland. *Social Physics: How Social Networks Can Make Us Smarter*. Penguin Publishing Group, January 2015.
- [10] N. Silver. *The Signal And The Noise: Why So Many Predictions Fail—but Some Don’t*. Penguin Publishing Group, February 2015.
- [11] B. Smale and M. Hilbrecht. Canadian Index of Wellbeing. In A. Michalos, editor, *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht, Netherlands, 2014.

- [12] B. Smale and M. Hilbrecht. From National to Local: Measuring Well-being at the Community Level. In R. Phillips and C. Wond, editors, *The Handbook of Community Wellbeing*. Springer, Dordrecht, Netherlands, 2016.
- [13] Social Research & Planning Council of the United Way of Perth Huron. Quality of Life. <http://perthhuron.unitedway.ca/wp-content/uploads/2014/09/SRPC-QOL-Report-FINAL.pdf>, 2014.
- [14] A. Toffler. Toffler Quotes. http://www.brainyquote.com/quotes/authors/a/alvin_toffler.html.
- [15] K. Wiebe and J. Distasio. Confronting the Illusion: Developing a Method to Identify Food Mirages and Food Deserts in Winnipeg. <http://winnspace.uwinnipeg.ca/bitstream/handle/10680/1205/2016%20Confronting%20the%20Illusion%20In-Brief.pdf?sequence=6&isAllowed=y>, June 2016.
- [16] Wikipedia. 1854 Broad Street cholera outbreak. https://en.wikipedia.org/wiki/1854__Broad_Street_cholera_outbreak.
- [17] Wikipedia. Two-factor authentication. https://en.wikipedia.org/wiki/Two-factor_authentication.
- [18] Wikipedia. Unintended Consequences. https://en.wikipedia.org/wiki/Unintended_consequences.
- [19] Wikipedia. Natural experiments. https://en.wikipedia.org/wiki/Natural_experiment, 2017.