Homelessness and open city data

Sonia Hugh [a]
Mark S. Fox [b]

[a] Centre for Social Services Engineering, University of Toronto
s.hugh@griffith.edu.au
Sonia Hugh is a Research Analyst in the Centre for Social Services Engineering at the University of Toronto. She obtained her MSc in Environmental Studies from Australian National University, Australia. Her work primarily involves analysing global and regional open datasets with GIS multi-purpose applications.

[b] Centre for Social Services Engineering, University of Toronto
msf@eil.utoronto.ca
Dr. Fox is a Distinguished Professor of Urban Systems Engineering, Professor of Industrial Engineering and Computer Science, and founding director of the Centre for Social Services Engineering at the University of Toronto. He is a Fellow of Association for the Advancement of Artificial Intelligence.

ABSTRACT

Homelessness is a global issue that continues to affect the developed and developing world. The absence of comprehensive data collection and measurement of homeless has led to a general lack of understanding of the global homeless. Improving our knowledge of homelessness requires information that reflects the reality of homelessness and housing exclusion. In this paper, we review homeless data that is openly available from 14 global cities. With the use of ISO37120 and open city data we analyzed the extent to which homeless data is standardized, consistent and comparable and the validity of the ISO37120 indicators in the context of this homeless data. The availability, accessibility and formatting of the open city data is directly compared using ISO37120 shelter themed city indicators (ISO37120 2014) and the City Data Openness Measure (CIDOM). We found that the majority of the homeless data could be placed into to two general categories: Homeless profiles and homeless services. There is no standard data model between cities, and periodicity of data collection varied. When directly compared using CIDOM, Toronto came out as the city with the most comprehensive dataset for each of the ISO37120 shelter indicators. ISO37120 shelter indicators do provide a standardized means to measure the homeless and unsheltered population of a city but city open data lacks the robustness and utility to produce indicators. Underlying data is not fully published raising questions of indicator value validity. In the global context, open city homeless data is currently not standardized, consistent or comparable.

KEYWORDS

Open City Data; ISO 37120; Shelters; Homeless; Data Standards

HIGHLIGHTS

- Questions whether the validity of city indicator data can be determined without the open publishing of the data used to derive them
- Suggests the CIDOM metrics as a basis for determining the openness of city data
- Reviews the availability of shelter/homeless data across 14 cities with significant open data portals
- Discovers that cities publish little shelter/homeless data on the portals

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

City Indicator Data Openness Measure (CIDOM)

Continuum of Care (CoCs)

Department of Housing and Urban Development (HUD)

European Typology of Homelessness and Housing Exclusion (ETHOS)

INTRODUCTION 1

Homelessness is a social issue that affects both developing and developed nations. Current data is insufficient to monitor and evaluate the trends and extent of homelessness around the globe (Busch-Geertsema et al. 2016). For the homeless data that does exist, information tends to be segregated into discrete locals, regions or nations that have their own frameworks, vocabularies and methodologies (Busch-Geertsema et al. 2016; Richter & Botha 2012). This inhibits global collaboration and understanding of the homeless experience by hindering information exchange and sharing of best practices (McCarney 2015). More cities are moving towards an open and transparent government (open government), making data and information publicly available (open city data), including homeless data sets that could be used to fill in the gaps of global homeless knowledge. Open city data provides a rich resource of information that can be used to promote accountability, engagement and innovation. It has the ability to be improved, standardized and continuously updated creating the potential for a source of robust and reliable homeless data.

Reliable data is a basic need of policy makers to effectively address the problem of homelessness (Springer 2000). The international standard ISO 37120 provides a standardized set of indicators with definitions and methodologies that allow for global comparison across cities (McCarney 2015; Deng et al. 2017). The indicators of the international standard were developed in order to help cities: measure performance management of city services; learn from one another by allowing comparison of performance measures; and share best practices (ISO37120 2014). The ISO 37120 shelter theme indicators measure the level of a city's homeless and under-sheltered populations, providing a standardized account of homeless numbers. In an attempt to create a reliable and robust homeless dataset, open city data can be used as source inputs for the ISO 37120 shelter themed indicators. However, the issue of validity of the indicator value arises. By validity, we ask is the indicator true/correct or false/incorrect? Although the ISO 37120 includes certification and third-party verification, the validity of an indicator's value cannot simply be taken at face value. A single indicator value is misleading as the reader has no clue to the provenance of the underlying data. The validity of an indicator is dependent on the publishing of the underlying data used to derive the indicator value (Fox, 2017).

With the lack of essential homeless data, it is necessary to extend all levels of homeless data, from local to global. We examine open city data to gain an understanding of the current state of global homeless data and investigate the validity of shelter indicators of the ISO 37120. We review the extent to which homeless data is standardized, consistent and comparable:

Standardized – Data sets that consist of common representation (terminologies, vocabularies, coding schemes) that allows for semantic interoperability. A standard vocabulary that ensures accurate interpretation and repetition of data. Data that is described systematically in unambiguous language to make the data machine-readable;
Consistent – How the data is collected and the periodicity of the data collection allowing for longitudinal analysis of homelessness. Data collected conforms to the definition of the indicator in which it is used;
Comparable – Data that is used to derive city indicators (i.e. ISO 37120 shelter indicators) and examined to note similarities and differences (transversal analysis);

We use the ISO 37120 shelter theme indicators, open city data and City Indicator Data Openness Measure (CIDOM) to compare cities for their data completeness and explore the indicator validity. Looking at the extent to which the open data city data sets and city indicators can be used in the longitudinal and transversal analysis of global homeless and unsheltered populations.

BACKGROUND 2

The impact of having good data on policy is exemplified by a study of income across Toronto's neighborhoods. Hulchanski (2010) discovered that in 1970 low income neighborhoods were located in the centre of the city (figure 1) but by 2005 low income neighborhoods moved into the inner suburbs (figure 2).
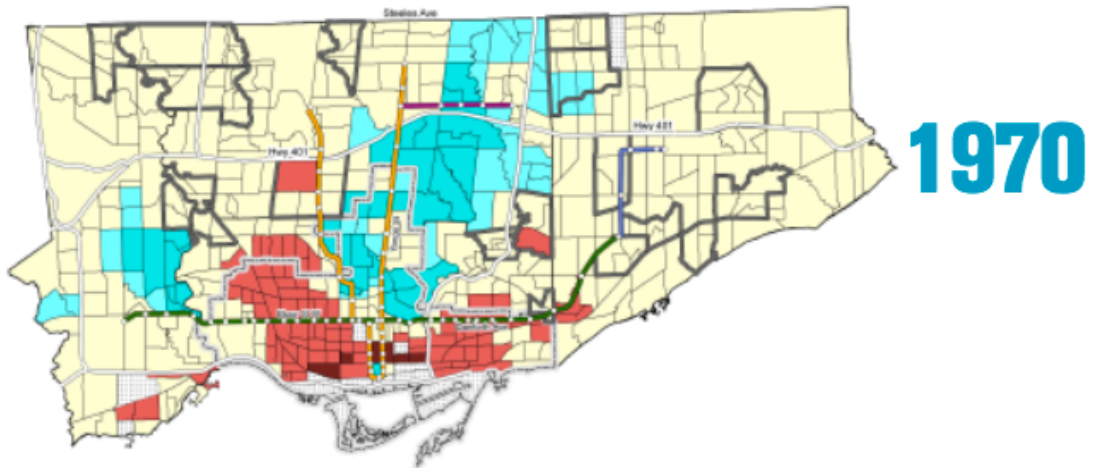
Census Tract Average Individual Income Relative to the Toronto CMA Average of $30,800* (estimated to 2001 census boundaries)

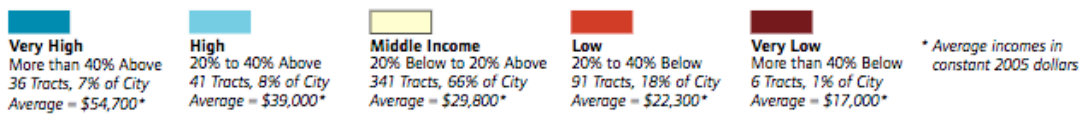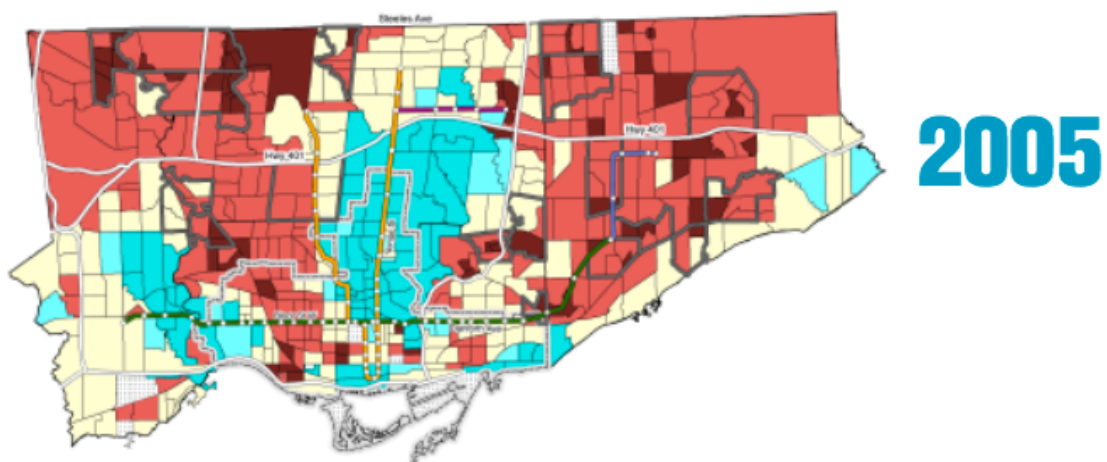| Very High More than 40% Above 36 Tracts, 7% of City Average = $54,700* | High 20% to 40% Above 41 Tracts, 8% of City Average = $39,000* | Middle Income 20% Below to 20% Above 341 Tracts, 66% of City Average = $29,800* | Low 20% to 40% Below 91 Tracts, 18% of City Average = $22,300* | Very Low More than 40% Below 6 Tracts, 1% of City Average = $17,000* | * Average incomes in constant 2005 dollars |

Figure 1: From Hulchanski (2010)

Census Tract Average Individual Income Relative to the Toronto CMA Average of $40,704 (estimated to 2001 census boundaries)

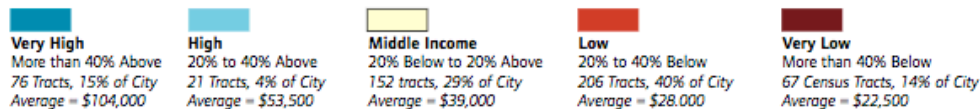| Very High More than 40% Above 76 Tracts, 15% of City Average = $104,000 | High 20% to 40% Above 21 Tracts, 4% of City Average = $53,500 | Middle Income 20% Below to 20% Above 152 tracts, 29% of City Average = $39,000 | Low 20% to 40% Below 206 Tracts, 40% of City Average = $28,000 | Very Low More than 40% Below 67 Census Tracts, 14% of City Average = $22,500 |

Figure 2: From Hulchanski (2010)

A direct consequence of this analysis was the movement by the United Way of Greater Toronto of many of their services from the inner city to the inner suburbs. Hulchanski's analysis relied upon census data. The question is whether relevant data exists for the homeless.

In order to understand global homelessness, we must conceptualize what it means to be homeless. There is no accepted unified global definition of homelessness. Definitions vary locally, nationally and globally, and are usually influenced by political, economic, climatic and social factors (Springer 2000; UN-Habitat 2000; Busch-Geertsema et al. 2016). A definition is of particular importance as it determines who will be recognized as homeless and dictates the prioritization of homeless policy (Springer 2000; Peressini et al. 1996).

Along with a homeless definition, reliable data about the homeless is needed by policy makers to make informed decisions. There is a general lack of global homeless statistics. Numbers on the homeless are largely drawn from developed nations in North America and Europe. There is a

patchwork of national homeless statistics available, but it is not currently possible to calculate a reliable estimate or derivatives of global homelessness (Springer 2000; Gaetz, S., Gulliver, T., & Richter 2013; Busch-Geertsema et al. 2016).

Most data are obtained through surveys, point-in-time counts and homeless management information systems in shelters and local authorities. Homeless information gathering and dissemination varies across the globe, there is no standard methodology or periodicity of data collection. Ideally, data collected on a consistent periodic basis is needed to perform longitudinal analysis and generate trend statistics (Busch-Geertsema et al. 2016). Data is fundamental in evaluating the state of homelessness by highlighting the whole system: inputs, services used, outputs and outcomes.   The data must reflect the reality of homelessness and housing exclusion, including the experiences and processes leading to becoming homeless (Gaetz et al. 2016). This will drive policies that are more sustainable over time and less emergency oriented, proactive rather than reactive (Springer 2000).

With more cities moving towards policy-making based on data (Fox 2013), open city data has potential to influence future homeless policy. It makes an ideal data source candidate as it is freely available, can be updated continuously and has the capacity to be standardized. But it also has its drawbacks. Cities publish the data in a variety of formats (i.e. PDFs, spreadsheets, XML), some more accessible than others. The data can also have little utility and are often not comparable as data models are not standardized with no semantic interoperability (Fox & Pettit 2015; Fox 2013).

Lack of data and data inconsistency makes it difficult to compare, homeless statistics, policies and programs between cities. ISO 37120 defines an international standard for city indicators that allows cities to measure performance and compare with other cities (McCarney 2015). ISO 37120 defines three shelter indicators, consisting of 1 core and 2 supporting indicators, measuring the homeless and unsheltered populations. The ISO 37120 shelter indicators are defined as:

15.1 Percentage of city population living in slums (Core),
"The percentage of city population living in slums shall be calculated as the number of people living in slums (numerator) divided by the city population (denominator). The result shall then be multiplied by 100 and expressed as a percentage. The number of people living in slums shall be calculated as the number of slum households multiplied by current average household size."

15.2 Number of homeless per 100 000 population,
"The number of homeless per 100 000 population shall be calculated as the total number of homeless people (numerator) divided by one 100 000th of the city's total population (denominator). The result shall be expressed as the number of homeless per 100 000 population."

15.3 Percentage of households that exist without register legal titles.
"The percentage of households that exist without registered legal titles shall be calculated as the number of households that exist without registered legal titles (numerators) divided by the total number of households (denominator)." (ISO 37120 2014)

Along with open city data, ISO 37120 provides the opportunity to create reliable homeless data that is standardized, consistent and comparable. The World Council on City Data (dataforcities.org) offers certification and third-party verification, but the processes are hidden, which raises the issue of indicator validity. Underlying data used to derive indicators are currently not required to be published; homeless indicators are only as good as the openness of the supporting data.

The integrity and validity of a standard city indicator relies on the supporting data being publicly available (Fox & Pettit 2015). The City Data Openness Measure (CIDOM) assesses the completeness of open city data in the context of measuring city indicators (Fox & Pettit 2015). It looks at the extent to which cities openly publish indicator data by quantifying the amount available and assessing the format of the supporting data used to derive the city indicator. CIDOM has 3 measures:

CIDOM – 1: Measure of the completeness of the data published for an indicator. The percentage of nodes in the dependency graph that are openly published by the city.

CIDOM – 2: Measure of the 'depth' to which supporting data is completely published. The number of levels of the dependency graph openly published by the city.

CIDOM-3: Determines the dominating format used to openly publish the data. The average of the format type for each node in the dependency graph openly published by the city (Fox & Pettit 2015). A dependency graph is a diagram of the supporting data that is used in the computation of the indicator. An example of a dependency graph can be seen in Figure 3 for the 15.1 shelter indicator.
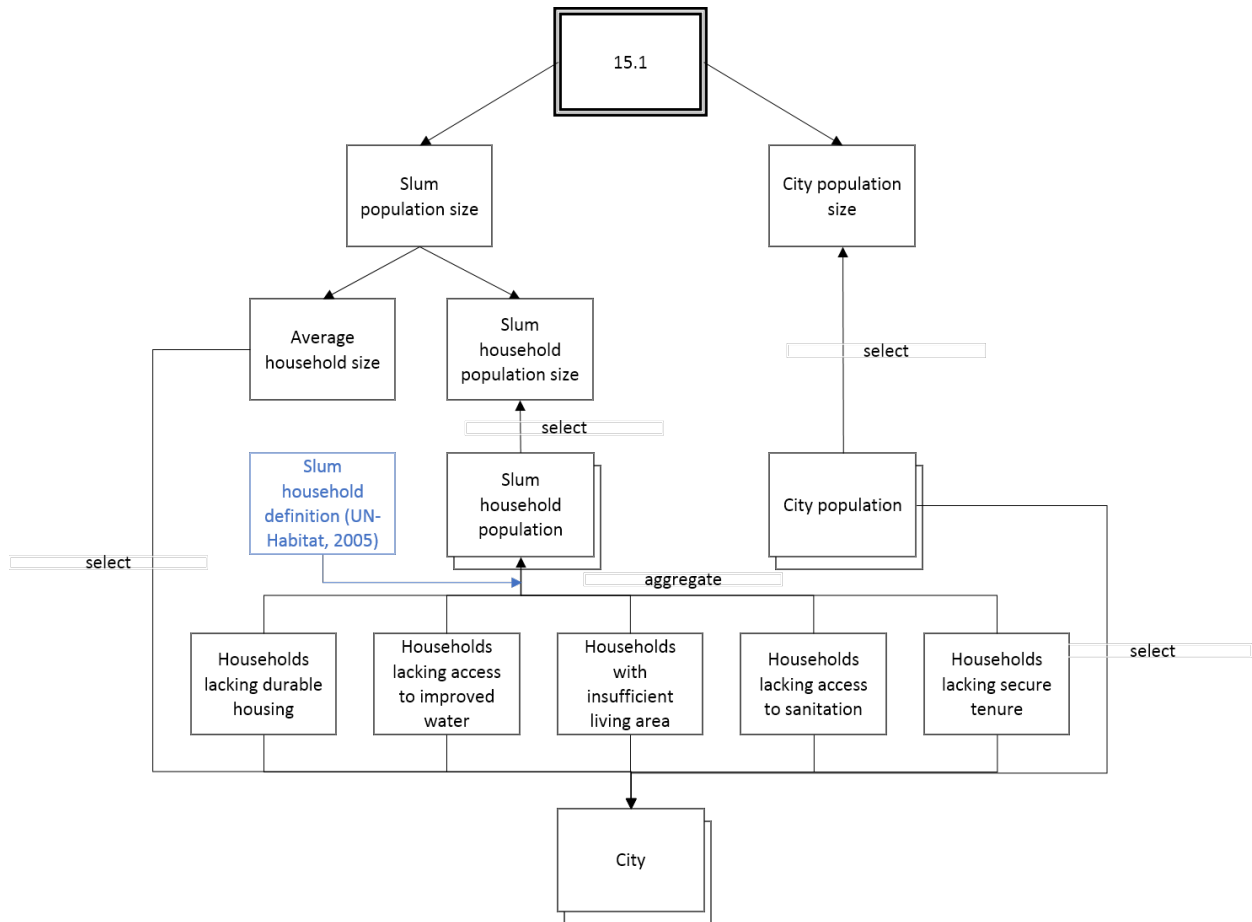

Figure 3: Dependency graph of ISO37120 Shelter Themed Indicator 15.1 Percentage of city population living in slums. Blue boxes represent ISO37120 definitions while the black boxes represent the actual data. Dependency graph adapted from Wang and Fox (2015).

Using CIDOM to compare the openness and completeness of city data in the context of the ISO37120 Shelter indicators, we get an indirect measurement of the validity, quality and reliability of the indicator data.

METHODOLOGY 3

We reviewed openly available homeless data for 14 cities chosen for their high degree of openness of their city data: Calgary, Toronto, New York, Chicago, San Francisco, Miami, London, Paris, Rome, Barcelona, Beijing, Shanghai, Tokyo and Singapore (Table 1). By openly available we mean, all homeless data, publicly available on the city's open data website and/or official city website. This includes linked homeless data from the official city website. We acknowledge that cities, like Calgary, have outside organizations that work with the homeless and homeless data, but were not included in the review as they were outside of our domain of official city websites. This also excludes data from the Department of Housing and Urban Development (HUD), including Continuum of Care (CoCs)

data, unless specifically linked or referred to in the official city websites. The official city websites were searched using the keywords found in table 2. The keywords were all related to themes found in the ISO 37120 shelter indicators definitions.

City	City website	Open data website
Calgary	www.calgary.ca	data.calgary.ca
Toronto	www.toronto.ca	www.toronto.ca/city-government/data-research-maps/open-data/
New York	www.nyc.gov	opendata.cityofnewyork.us
Chicago	www.cityofchicago.org	data.cityofchicago.org/
San Francisco	sfgov.org	datasf.org/opendata/
Miami	www.miamigov.com	data.miamigov.com/
London	www.cityoflondon.gov.uk	data.london.gov.uk
Paris	www.paris.fr	opendata.paris.fr/
Rome	www.comune.roma.it	dati.comune.roma.it/
Barcelona	www.bcn.cat	www.opendata-ajuntament.barcelona.cat/en/
Beijing	www.egeijing.gov.cn	www.bjdata.gov.cn
Shanghai	www.shangai.cov.cn	www.datashanghai.gov.cn
Tokyo	http://www.metro.tokyo.jp/	
Singapore	www.gov.sg	www.data.gov.sg

Table 1: List of city and city open data websites investigated

It is acknowledged that the keyword search is not exhaustive and that there is potential to miss homeless data entries. The websites were searched between September and December 2016. Any homeless data found on each city's website was noted and aggregated into categories. For cities with websites in languages other than English, sites were translated and translated versions of the keywords were used. The review aimed to establish the type of data available on the subject of homelessness for each city and report any standards and consistency of data.

Keywords		
homeless/ness	population	slum
PiT count	shelter	clean water
census	overcrowding	housing
households	secure tenure	rough sleeping

Table 2: Keyword search for city homeless data

The homeless city data was then compared using the CIDOM. CIDOM assesses the completeness of open city data and is applied in the context of the ISO 37120 shelter theme indicators. The ISO 37120 shelter theme indicators are used as a measure of the homeless and unsheltered population of a city. Using the definition of the shelter theme indicators, three dependency graphs were created (Figures 3-5). The dependency graphs were used as a roadmap to determine the openness of the

indicators when calculating the three measures of CIDOM. The indicator's definition and dependency graph are based on the shelter ontology provided by Wang & Fox (2015). CIDOM calculates an indirect measure of comparability, reliability and validity. The model allows data to be interpreted as standardized and whether the integrity of the data can be measured using supporting data.
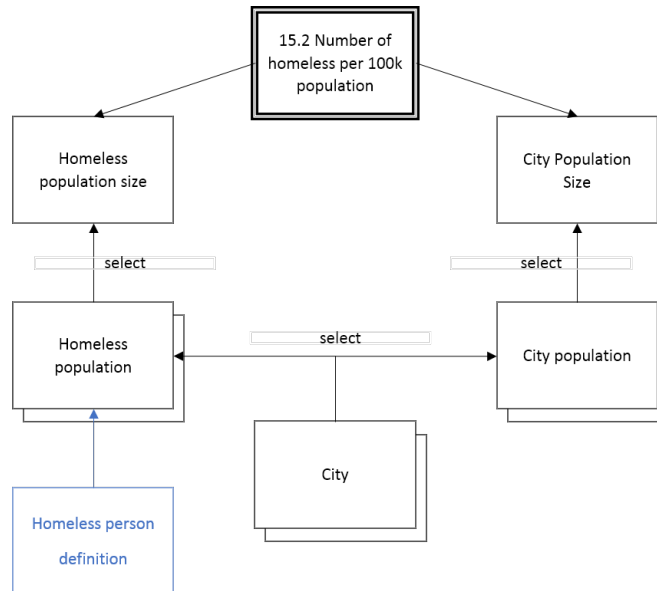

Figure 4: Dependency graph of ISO37120 Shelter Themed Indicator 15.2 Number of homeless 100k population.
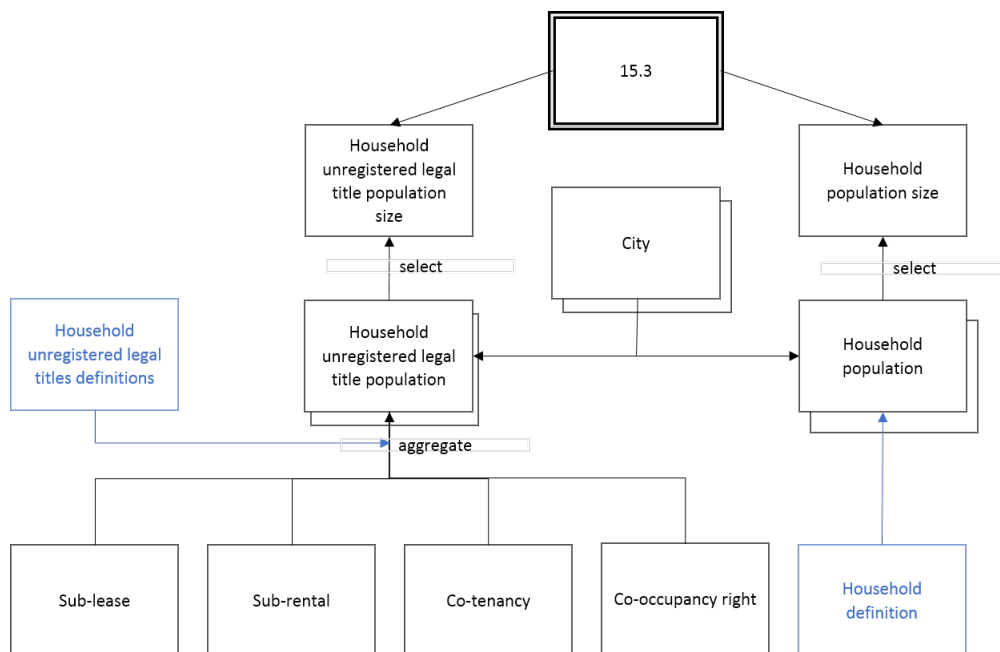

Figure 5: Dependency graph of ISO37120 Shelter Themed Indicator 15.3 Percentage of households that exist without registered legal titles.

RESULTS AND DISCUSSION 4

Standardization 4.1

Homeless definition 4.1.1

The definition of 'homeless' is of the utmost importance as it determines how homeless individuals are enumerated and measured. This in turn influences policies that ultimately dictate who receives services and support. There is no globally accepted definition of homelessness and cities and countries vary in their definition. The North American cities examined tended to go with literal definitions of homelessness, which is usually made up of two groups: unsheltered and sheltered individuals (Busch-Geertsema et al. 2016). The 'relative homeless' (Cooper 1995) or 'hidden homeless' are not considered (all homeless definitions referred to can be found in table 5 and 6 in the appendix). All American cities performed homeless surveys according to the US Department of Housing and Urban Development's (HUD) definitions of unsheltered and sheltered homeless, also for chronic homeless for data after 2015. The city of San Francisco extends HUD definitions of homelessness to include inadequate housing (Busch-Geertsema et al. 2016) and those found in correctional, health and treatment facilities (ASR (Applied Survey Research) 2015).   Similarly, Toronto's definition includes correctional, health and treatment facilities, but excludes the hidden homeless (Toronto 2013). London's homeless definition consists of the statutory homeless and rough sleepers (unsheltered), while the inadequately housed and the hidden homeless are not measured. An individual is considered 'statutory homeless' if they lack a secure place in which they are entitled to live or not reasonably be able to stay in the current accommodation as determined by the local authority, adding an element of subjectivity. Barcelona based their definition on the European Typology of Homelessness and Housing Exclusion (ETHOS), which includes those with insecure and inadequate housing (Sales et al. 2015; FEANTSA 2007).

The data used to inform homeless policies and programs are based on individuals defined as homeless. The cities examined have similar definitions of homeless, however, there are many subpopulations that are not considered or are missing from city data. The homeless problem is idiosyncratic to each city; thus, it makes sense that a single definition could not be applied uniformly across all cities. Busch-Geertsema *et al.* (2016) suggests the use of a framework that encompasses a wider range of homeless categories/definitions. The framework would be used as a point of reference for cross-continental discussions and comparisons, providing a means of transparency by clarifying homeless populations included (or not) in the city's definition. There is no current consensus on a global homeless framework. While ISO 37120 offers a starting point for a standard and comparable global homeless dataset, the standard is very coarse. The ISO 37120 shelter indicators use a homeless definition (Table 5 in appendix), 'those without any physical shelter", that loosely describes the literal homeless, but is much narrower in scope. This definition has the potential to leave many subpopulations out of the derivation of the indicator because it primarily focuses on rough sleepers. A reworking of the definition to be clearer and inclusive of other sheltered subpopulations might lead to more reliable and representative measure of homelessness.

Homeless data 4.1.2

All of the cities examined had an official city website and open city data website, except for Tokyo. Of the 14 cities, 8 had some form of quantitative homeless data. Most data values were found in tables, text (HTML) or PDFs, no spatial data was available for any of the cities other than locations of shelters. Over 500 types of homeless data entries were found (but not limited to) on these city websites. The data can be broken down into two broad categories: homeless profiles and homeless services (Table 3). Homeless profiles consist of estimates, trends and characteristics of those experiencing homelessness. While the homeless service category focuses on the homeless services offered and utilized by the homeless. Table 3 breaks down the homeless data types into broad categories and sub-categories. The data sets also vary in data range for each city. For example, San Francisco had approximately 171 quantitative data entries while Miami had only 27 entries openly available. In the United States, data is readily available on the HUD website but is often not linked to by the city websites.

Category	Sub-categories	
Homeless Profiles		Total

		Indoor
		Outdoor (unsheltered)
		Rooflessness
		Homelessness
		Insecure housing
		Inadequate housing
		Youth
		Chronic homeless
		Veterans
		Families with children
		Statutory homelessness
Homeless Services	Services Needs	Housing and homeless services
		Health and treatment services
		Non-housing specific services
		Youth
		Housing assistance
		Wait list for housing
	Services Used	Housing and homeless services
		Health and treatment services
		Non-housing specific services
		Government assistance
		Reasons for not receiving government Assistance
		Adults
		Youth
		Housing assistance
		Service outcomes
		Homeless prevention enrollments
		Housing placement
		Length of time in temp accommodation
		Housing inventory
		Daily shelter census
		Prevention
		Relief
		City homeless/housing plan

Table 3: Homeless data types

Eight cities had homeless data (Toronto, New York, Chicago, San Francisco, Miami, London, Barcelona and Tokyo) and a common homeless data category of total homeless. There were no other homeless data categories that matched between the cities. This follows trends of other open city data, where the data has no common representation or semantics (Fox 2013; Fox & Pettit 2015; Nalchigar & Fox 2014; Fox 2017).

Homeless profiles contain major homeless subcategories that further focus on key populations. Most cities begin by estimating the total homeless population followed by counts of the sheltered and unsheltered populations. This is then further broken down into key homeless sub-populations detailing demographics, histories and current situations of homeless individuals. Sub-populations are very similar in definition but differ in the scope of who is considered homeless. For example, Barcelona's roofless definition includes those sleeping in the rough and in night shelters but does not include women's shelters, while Toronto's definition of general homeless includes rough sleepers and shelters including women's shelters. This is one of the reasons why there are so many types of homeless data and why it would be difficult to unify the data in the current state.

Homeless services consist of two major categories: service needs and services used. This includes services like housing, health services and governmental assistance. The service needs category details the nature of services that are met and unmet and the services used describes what and how

services are utilized. The services category again varies depending on country and city because each have their own unique response to the homeless problem.

Consistency 4.2

Collection of data on a consistent basis is critical in order to generate reliable trend statistics. Trend data highlights shifts and responses to current policy, informing future policy (Busch-Geertsema et al. 2016). In this study, we viewed consistency from a temporal context of data collection, how often data is collected; and from a methodological context, how is the data collected?

The extent and period of the data vary depending on city and category of data. In most cases, data is produced on an annual scale, but is not consistently measured. Some cities, like New York and Toronto, have daily shelter census but perform PIT (Point-In-Time) counts in yearly intervals. London produces quarterly statutory homeless statistics, but annual rough sleeping counts. American cities belonging to the Continuum of Care (CoCs) are required to perform PIT counts on a yearly basis (HUD 2014). Homeless open data produced by cities is still in the initial stages, the earliest record of for open data in this study is from 2001 in Tokyo.

Methods to enumerate the sheltered homeless population usually consist of survey or sampling regimes on one given night of the year. Seven of the 8 cities with homeless open city data mention point-in-time (PiT counts) as their mode of enumeration, however, PiT methodologies can vary from city to city. PiT counts have been criticized for their lack of consistency and validity. In the United States, varying PiT methodologies have been found to lead to unreliable results (Schneider et al. 2016). Aligning homeless counts from various cities is difficult due to varying details in the enumeration methodologies. Clarity and transparency of count methods will aid in comparability of subsets of the overall homeless population (Smith 2015). Other methods of homeless estimates include homeless management information systems that collect data at shelters or government facilities. Qualitative data, like interviews of homeless individuals, were found in Barcelona, Calgary and Chicago. Like the definition of homelessness, there is no globally accepted method for measuring homelessness. A systemic approach, adjusted for context, would be required for reliable and comparable homeless estimates (Busch-Geertsema et al. 2016).

Comparability and Validity 4.3

We use CIDOM as an indirect measure of validity and comparability of city datasets. The quality and reliability of the indicator data can be measured by looking at the totality of data published for an indicator, the completeness of supporting data and the dominating format used to publish. The calculated CIDOM values can be found in Table 4.

CIDOM	Indicator	Calgary	Toronto	New York	Chicago	San Francisco	Miami	London	Paris	Rome	Barcelona	Beijing	Shanghai	Tokyo	Singapore
1	15.1	25	100	25	17	8	17	25	17	0	33	0	0	25	8
	15.2	40	100	100	100	40	100	80	20	0	100	0	0	100	0
	15.3	0	44	0	0	0	0	0	0	0	11	0	0	0	0
2	15.1	0	2	0	0	0	0	0	0	0	0	0	0	0	0
	15.2	0	2	0	0	0	0	0	0	0	0	0	0	0	0
	15.3	0	2	0	0	0	0	0	0	0	0	0	0	0	0
3	15.1	1	1	1	1	1	1	1	1	0	1	0	0	1	1
	15.2	1	1	1	1	1	1	1	1	0	1	0	0	1	0
	15.3	0	1	0	0	0	0	0	0	0	1	0	0	0	0

Table 4: CIDOM results. CIDOM-1 is measured in %. CIDOM-2 and CIDOM-3 are measured in levels.

Toronto, London, Barcelona and Shanghai have implemented the ISO 37120 standards. However, Toronto is the only city to have shelter indicator values openly available on the city website. It was the most comprehensive of the city data sets reviewed.

CIDOM 1 - Totality of the data published for an indicator. 4.3.1
The ISO 37120 shelter indicators act as reference points providing standardized values for open city datasets, allowing for comparisons between cities. The supporting data used to calculate the indicators are often not openly available or accessible (Fox & Pettit 2015), making the value of the shelter indicator an abstract number with no currency. CIDOM-1 is a measure that looks into the totality of published supporting data for indicators. Toronto is the only city that consistently openly published supporting data for an indicator, publishing data for most of the nodes in the dependency graphs. All CIDOM calculations can be found in Table 4. Toronto is one of two cities to have data for indicator 15.3, the other being Barcelona. Calgary, Paris and Singapore show up in the CIDOM analysis because their city websites contained data on household and city populations, fundamental values in the calculation of indicators 15.1 and 15.2. A majority of the cities have totally published data sets for indicator 15.2, but lack data for indicators 15.1 and 15.3. Indicator 15.2 is a direct measure of homelessness and has the potential to provide a standard and reliable homeless estimate for global homeless numbers. Only half of cities reviewed could openly source indicator 15.2.

CIDOM 2 – Number of levels of supporting data openly published 4.3.2
CIDOM-1 looks at the data from one dimension quantifying how much data is available. CIDOM-2 reviews the integrity of the indicator value and its supporting data. Knowing the extent to which the data is openly published is just as important as knowing how much is openly available. The metadata of supporting data is crucial to the validity of the indicator value. Toronto is the only city to have a level higher than zero (level 2) because the indicator value is openly published. For indicator 15.2, 5 other cities have all of the openly available data needed to calculate the indicator, but simply have not performed the ISO 37120 calculation or openly published the indicator value.

CIDOM 3 – Dominant format published 4.3.3
The data was published in a mixture of tables and reports in pdf format, csv (Excel) and html format. The most pervasive format was PDF. In terms of publishing formats, PDF is the least accessible. Homeless values are difficult to extract reducing the opportunity for semantic interoperability.

Toronto is the only city to calculate the ISO 37120 shelter theme indicators and produce underlying data. It could be argued that indicator 15.2 is the most significant (in the context of homelessness) of the three ISO 37120 indicators as it focuses on the literal homeless, while the 15.1 and 15.3 focus on relative homelessness. Shelter indicator 15.2 could easily be calculated for many cities producing a globally comparable value for homeless numbers and most of the supporting data was found in half of the cities reviewed. A strength in using the city indicators is also one of its weaknesses, standardized definitions and methodologies. For example, Toronto has a different definition of households living below living standards. Canada's occupancy standard (Toronto 2013) states overcrowding at >2 people per bedroom. Indicator 15.1 defines sufficient living spaces as not exceeding 3 people per room. The slight difference in definition allows Toronto to over report what is considered to be overcrowding by the ISO 37120, skewing the overall value of Indicator 15.1. The indicator's definition is based on slum households, defined by the UN HABITAT (2006) in a developing nation context but incorrectly used on developing nations. This showcases the need for openly publishing supporting data, it allows for regional contexts to be transparent and known, aiding in meaningful comparisons of the city data. In addition, publishing underlying data supports the validity of the indicator value.

CONCLUSION 5

In this paper, we reviewed the current state of openly available homeless data by evaluating how standardized, consistent and comparable the data was across cities and examined the validity of ISO 37120 indicators in the context of open homeless data. Improving our knowledge of the complicated dynamics of homelessness requires information that reflects the reality of the homeless. If the ultimate goal is to eradicate homelessness, current homeless data is very limited and informs common mitigation strategies like increased shelter and affordable housing. The drivers of homelessness are multifaceted and capturing the data required to explain these issues will be a major challenge for future policy. In order to achieve a better understanding of the homeless problem, it is necessary to extend and improve our current local and global database, which includes foundational

statistics like homeless counts. We found that there is still a lot of work to be done with open city data to realize a vision of reliable homeless estimates, as data from different nations are neither standardized, consistent or comparable.

Fundamental barriers like differences in homeless definitions, methodologies, and periodicity of data capture, make city data difficult to compare. A global definition/ framework for homelessness will aid in the foundation of a reliable global homeless data set. Global entities like the Institute of Global Homelessness or the United Nation's HABITAT or Human Rights program are all good candidates to spearhead this initiative as they have already begun dialogue on homeless definitions and measures (UN-Habitat 2000 2006; UN-Human Rights 2015; Busch-Geertsema et al. 2015). Setting a frame of reference to identify who is homeless, when to perform data captures and permitting room for regional context provides a transparent base to foster meaningful comparisons and information exchange. For example, the U.S Department of Housing and Urban development has stringent definitions for homelessness. Methods of enumeration and data capture are standardized allowing for city to city comparison. While the HUD system is not flexible, cities like San Francisco still use the pre-defined standards but also add their own regional interpretation (i.e. expanding on the homeless definitions, measuring other homeless sub-categories like youth) in their city reporting.

Using a coarse global standard like the ISO 37120 shelter indicators, currently allows for the comparison of city data. The supporting data from which the indicator values are derived must be published. As long as the underlying data is transparent and available, the validity of calculated indicator values can be verified. This gives rise to the issue of how to represent the indicators and the underlying data. Development of a semantic approach will help to unify, link and ground the data.   A shelter ontology for homeless city data (Wang & Fox 2015) helps to achieve computational accessibility allowing for longitudinal and transversal analysis, breaking down the fundamental barriers of homelessness data mentioned above.

This work highlights the need for global standards for homeless data. We were able to show that it is possible to calculate globally standard measures of homelessness (ISO 37120, indicator 15.2) using open city data, ISO 37120 shelter indicators and CIDOM. However, wide data gaps still exist limiting the utility of the shelter indicators.   It also showcases the need for openly published supporting data to prove the robustness and integrity of indicator values. In the global context, open city homeless data is currently not standardized, consistent or comparable. This review contributes to a strategic prioritization for improved open data collection, measurement and standardization of homeless data.

Much research remains. To support the standardization of homeless data effort, there are at least two issues researchers need to address:
1. A consensus needs to be developed among homeless researchers in the social sciences as to the various types of data that they need to move their research forward, and
2. Ontologies and standards based on these ontologies need to be developed to precisely, and unambiguously represent the semantics of the data.

The latter issue is being explored not only by our group at the Centre for Social Services Engineering (csse.utoronto.ca), but more general by standards organizations such as the ISO/IEC Joint Technical Committee Working Group 11 on Smart Cities.

ACKNOWLEDGEMENTS

REFERENCES

ASR (Applied Survey Research), 2015. *San Francisco homeless point-in-time count and survey: comprehensive report 2015*, San Francisco, Applied Survey Research. Available at: http://dhsh.sfgov.org/wp-content/uploads/2016/06/2015-San-Francisco-Homeless-Count-Report_0-1.pdf.

Busch-Geertsema, V., Culhane, D. & Fitzpatrick, S., 2015. A global framework for understanding and measuring homelessness. Available at: http://docs.wixstatic.com/ugd/d41ae6_97a693a1aba845058f91e9cf38f7c112.pdf

Busch-Geertsema, V., Culhane, D. & Fitzpatrick, S., 2016. Developing a global framework for conceptualising and measuring homelessness. *Habitat International*, 55, pp.124–132. Available at: http://www.sciencedirect.com/science/article/pii/S0197397515300023.

Cooper, B., 1995. *Shadow people: the reality of homelessness in the 90s*, Sydney City Mission. Available at: https://www.researchgate.net/publication/228268057_Shadow_People.

Deng, D., Liu, S., Wallis, L., Duncan, E., & McManus, P. 2017. Urban Sustainability Indicators: how do Australian city decision makers perceive and use global reporting standards? *Australian Geographer*, 0(0), pp.1–16. Available at: http://dx.doi.org/10.1080/00049182.2016.1277074.

FEANTSA, 2007. *FEANTSA Proposal: A retrospective module on homelessness for household surveys*, Brussels.

Fox, M.S., 2013. *City Data: Big, open and linked*, Working Paper, Enterprise Integration Laboratory, University of Toronto. Available at: http://www.eil.utoronto.ca/wp-content/uploads/smartcities/papers/City-Data-v5.pdf.

Fox, M.S., 2017. The PolisGnosis Project: Enabling the computation analysis of city performance. In *Proceeding of the 2017 Industrial and Systems Engineering Conference*. Available at: http://eil.mie.utoronto.ca/wp-content/uploads/2015/06/IISE-2017-3349.pdf.

Fox, M.S. & Pettit, C.J., 2015. On the completeness of open city data for measuring city indicators. In *2015 IEEE 1st International Smart Cities Conference, ISC2 2015*. Guadalajara.

Gaetz, S., Gulliver, T., & Richter, T., 2013. *The State of Homelessness in Canada 2013*,

Gaetz, S. et al., 2016. *The state of homelessness in Canada 2016*, Toronto. Available at: http://homelesshub.ca/SOHC2016.

HUD (U.S. Department of Housing and Urban Development), 2014. *Point-in-Time count methodology guide*, Available at: https://www.hudexchange.info/resources/documents/PIT-Count-Methodology-Guide.pdf.

Hulchanski, J. D. (2010). *The three cities within Toronto: Income polarization among Toronto's neighbourhoods, 1970-2005*. Toronto: Cities Centre, University of Toronto.

ISO37120, 2014. ISO37120: Sustainable development of communities - Indicators for city services and quality of life.

McCarney, P., 2015. The evolution of global city indicators and ISO37120: The first international standard on city indicators. *Statistical Journal of the IAOS*, 31(1), pp.103–110. Available at: http://search.ebscohost.com.myaccess.library.utoronto.ca/login.aspx?direct=true&db=crh&AN=101648412&site=ehost-live.

Nalchigar, S. & Fox, M.S., 2014. An ontology for open 311 data. In *Workshop on semantic cities: Beyond open data to models, standards and reasoning, AAAI14*. Quebec City: AAAI14.

Peressini, T., McDonald, L. & Hulchanski, D., 1996. *Estimating homelessness: Towards a methodology for counting the homeless in Canada*, Canada Mortgage and Housing Corporation.

Richter, S. & Botha, A., 2012. Happenings / L'Événement - New Research Initiatives: Addressing Global Homelessness. *Canadian Journal of Nursing Research Archive*, 44(4). Available at: http://cjnr.archive.mcgill.ca/article/view/2379.

Sales, A., Uribe, J. & Marco, I., 2015. *2015 diagnosis: The situation of homelessness in Barcelona. Evolution and intervention policies*, Available at: http://www.bcn.cat/barcelonainclusiva/ca/2016/2/sense_sostre2015_ang.pdf.

Schneider, M., Brisson, D. & Burnes, D., 2016. Do We Really Know How Many Are Homeless?: An Analysis of the Point-In-Time Homelessness Count. *Families in Society: The Journal of Contemporary Social Services*, 97(4), pp.321–329. Available at: http://dx.doi.org/10.1606/1044-3894.2016.97.39.

Smith, A., 2015. Can We Compare Homelessness Across the Atlantic? A Comparative Study of Methods for Measuring Homelessness in North America and Europe1. *European Journal of Homelessness _ Volume*, 9(2).

Springer, S., 2000. Homelessness: A proposal for a global definition and classification. *Habitat International*, 24(4), pp.475–484.

Toronto, C. of, 2013. *2013 Street assessment results*, Available at: http://www.toronto.ca/legdocs/mmis/2013/cd/bgrd/backgroundfile-61365.pdf.

Toronto, C. of, 2013. Toronto's 2013 results under ISO37120: Indicators of city service deliver and quality of life. Available at: https://www.toronto.ca/wp-content/uploads/2017/11/989c-Final-Summary-of-Torontos-WCCD-ISO-37120-Results-6-AODA-.pdf.

Un-Habitat, 2006. *The State of World's Cities 2006/2007*, UN-Habitat. Available at: https://unhabitat.org/books/state-of-the-worlds-cities-20062007/.

UN-Habitat, 2000. *Strategies to combat homelessness*, UN-Habitat.

UN-Human Rights, 2015. Report of the Special Rapporteur on adequate housing as a component of the right to an adequate standard of living, and on the right to non-discrimination in this context. Available at: http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/31/54

Wang, Y. & Fox, M.S., 2015. A shelter ontology for global city indicators (ISO 37120)

APPENDIX

Table 5: Homeless definitions

Type of homelessness	Definition	Source
Homelessness	Absolute homelessness refers to those without any physical shelter, for example, those living outside, in parks, in doorways, in parked vehicles, or parking garages, as well as those in emergency shelters or in transition houses for women fleeing abuse.	ISO37120
(Toronto)	Any individual sleeping outdoors on the night of the survey, as well as those staying in emergency shelters, in Violence Against Women (VAW) shelters, individuals in health or treatment facilities with no permanent address, as well as those in correctional facilities who are registered in a Toronto court as having no fixed address or a shelter address. Excludes hidden homeless.	(Toronto 2013)
	Individuals and families who are: Living in a supervised publicly or privately operated shelter designated to provide temporary living arrangement; or with a primary nighttime residence that is a public	(HUD 2014)

	or private place not designed for or ordinarily used as a regular sleeping accommodation for human beings, including a car, park, abandoned building, bus or train station, airport, or camping ground.	
(San Francisco)	The definition of homelessness in San Francisco expands HUD's definition to include individuals who are "doubled-up" in the homes of family or friends, staying in jails, hospitals, and rehabilitation facilities, families living in Single Room Occupancy (SRO) units, and in substandard or inadequate living conditions including overcrowded spaces.	(ASR 2015)
Unsheltered	Individual or family with a primary nighttime residence that is a public or private place not designed for or ordinarily used as a regular sleeping accommodation for human beings, including a car, park, abandoned building, bus or train station, airport, or camping ground.	(HUD 2014)
Sheltered	An individual or family living in a supervised publicly or privately operated shelter designated to provide temporary living arrangement (including congregate shelters, transitional housing, and hotels and motels paid for by charitable organizations or by federal, state, or local government programs for low-income individuals).	(HUD 2014)
Chronic	(1) A ''homeless individual with a disability,'' as defined in section 401(9) of the McKinney-Vento Homeless Assistance Act (42 U.S.C. 11360(9)), who: (i) Lives in a place not meant for human habitation, a safe haven, or in an emergency shelter; and (ii) Has been homeless and living as described in paragraph (1)(i) of this definition continuously for at least 12 months or on at least 4 separate occasions in the last 3 years, as long as the combined occasions equal at least 12 months and each break in homelessness separating the occasions included at least 7 consecutive nights of not living as described in paragraph (1)(i). Stays in institutional care facilities for fewer than 90 days will not constitute as a break in homelessness, but rather such stays are included in the 12-month total, as long as the individual was living or residing in a place not meant for human habitation, a safe haven, or an emergency shelter immediately before entering the institutional care facility; (2) An individual who has been residing in an institutional care facility, including a jail, substance abuse or mental health treatment facility, hospital, or other similar facility, for fewer than 90 days and met all of the criteria in paragraph (1) of this definition, before entering that facility; or (3) A family with an adult head of household (or if there is no adult in the family, a minor head of household) who meets all of the criteria in paragraph (1) or (2) of this definition, including a family whose composition has fluctuated while the head of household has been homeless.	Homeless emergency assistance and rapid transition to housing: Defining "Chronically Homeless", 80 Fed.Reg. 75791 (December 4, 2015). *Federal Register: The daily journal of the United States.* Web. 4 December 2015. url: https://www.gpo.gov/fdsys/pkg/FR-2015-12-04/pdf/2015-30473.pdf
Absolute	People without an acceptable roof over their heads, living on the streets under bridges and deserted buildings.	(Cooper 1995)
First degree relative	People moving between various forms of temporary or medium term shelter such as refuges, boarding houses hostels or friends.	(Cooper 1995)
Second degree relative	People constrained to live permanently in single rooms in private boarding houses	(Cooper 1995)
Severely inadequate and/or insecure accommodation	3 (a) People sharing with friends and relatives on a temporary basis 3 (b) People living under threat of violence 3 (c) People living in cheap hotels, bed and breakfasts and similar 3 (d) People squatting in conventional housing 3 (e) People living in conventional housing that is unfit for human habitation 3 (f) People living in trailers, caravans and tents 3 (g) People living in extremely overcrowded conditions 3(h) People living in non-conventional buildings and temporary structures, including those living in slums/informal settlements	(Busch-Geertsema et al. 2016)

Hidden	Individuals living with others in conventional housing but on an emergency basis.	(Busch-Geertsema et al. 2016)
Statutory	Households which meet specific criteria of priority need set out in legislation, and to whom a homelessness duty has been accepted by a local authority. Such households are rarely homeless in the literal sense of being without a roof over their heads, but are more likely to be threatened with the loss of, or are unable to continue with, their current accommodation.	(Department of communities and local government UK 2013)
Rough sleepers	People sleeping, about to bed down (sitting on/in or standing next to their bedding) or actually bedded down in the open air (such as on the streets, in tents, doorways, parks, bus shelters or encampments) People in buildings or other places not designed for habitation (such as stairwells, barns, sheds, car parks, cars, derelict boats, stations, or 'bashes'). The definition does not include people in hostels or shelters, people in campsites or other sites used for recreational purposes or organised protest, squatters or travellers.	(Department of communities and local government (UK) 2013)
Literal	People without any accommodation, and those living in temporary or emergency accommodation specifically provided for homeless people.	(Busch-Geertsema et al. 2016)

Table 6: ETHOS homeless categories (Amore et al. 2011; Sales et al. 2015)

Homeless	Definition
Roofless	People living rough or in a public space. People sleeping in a night shelter and/or forced to spend the day in a public space.
Houseless	People living in hostels or in accommodation for the homeless. Temporary accommodation. Women's shelter accommodation. People living in temporary accommodation for immigrants or asylum seekers. People who live in housing institutions or penal institutions, prospect of being dismissed in a deadline without shelter housing available. People who live in a continued support accommodation for homeless people.
Insecure housing	People who live in insecure tenancy housing. Without paying rent. People who live under threat of eviction. People who live under threat of family or partner's violence.
Inadequate housing	People who live in temporary or nonconventional structures. People who live in inappropriate housing according to legislation. People who live in overcrowded housing