Article

Enhancing Disaster Management: Development of a Spatial Database of Day Care Centers in the USA

Nagendra Singh *, Mark Tuttle and Budhendra Bhaduri

Geographic Information Science and Technology Group, Oak Ridge National Laboratory, One Bethel Valley Road, P.O. Box 2008 MS-6017, Oak Ridge, TN 37831, USA;
E-Mails: tuttlema@ornl.gov (M.T.); bhaduribl@ornl.gov (B.B.)

* Author to whom correspondence should be addressed; E-Mail: singhn@ornl.gov;
Tel.: +1-865-241-9596; Fax: +1-865-241-6261.

Abstract: Children under the age of five constitute around 7% of the total U.S. population, and represent a segment of the population that is totally dependent on others for day-to-day activities. A significant proportion of this population spends time in some form of day care arrangement while their parents are away from home. Accounting for those children during emergencies is of high priority, which requires a broad understanding of the locations of such day care centers. As concentrations of at risk population, the spatial location of day care centers is critical for any type of emergency preparedness and response (EPR). However, until recently, the U.S. emergency preparedness and response community did not have access to a comprehensive spatial database of day care centers at the national scale. This paper describes an approach for the development of the first comprehensive spatial database of day care center locations throughout the U.S. utilizing a variety of data harvesting techniques to integrate information from widely disparate data sources followed by geolocating for spatial precision. In the context of disaster management, such spatially refined demographic databases hold tremendous potential for improving high-resolution population distribution and dynamics models and databases.

Keywords: day care; child care; spatial database; population distribution

1. Introduction

Children are one of the greatest priorities of our society and are dependent on parents as they grow up in their home. However individual families, particularly with working parents, typically need assistance to look after their small children mainly while they are away from home for work. As the proportion of mothers entering the labor force increases, so does the proportion of children in day care. Based on sample data from the U.S. Census Bureau Survey of Income and Program Participation (SIPP) of 2011 [1], around 61% (12.5 million) of kids below the age of five are in in some type of day care arrangement while around 50% (29.5 million) of kids aged 5–14 spend time in before- or after-school day care. Therefore, on a given work day, approximately 32 million kids (*i.e.* 11% of the U.S. population) are in the care of someone other than their parents and in a place outside their home. A small percentage (approx. 3%) are cared for in their homes. Most of the kids who are in day care range in age from six weeks to less than five years.

Child care centers represent a unique section of population that is critical for disaster preparedness and response due to the complete reliance of this population segment on others during an emergency. While individual states in the U.S. monitor and regulate day care centers, there is a lack of spatial data that can be readily used during disasters and for use in simulation exercises. Therefore, there is a critical need for a spatial database of day care centers for disaster planning and response. At the same, time they are a crucial component of the population and need to be accounted for in population distribution and dynamics.

2. Background

Based on the classification by the SIPP, day care can be broadly grouped into two major types: relative (where day care service provided by relatives of the child which could include father, mother, siblings, or grandparents) or non-relative (that includes day care centers, head starts, family based day care or service provided by individuals who are non-relatives). According to the 2010 U.S. census, of the total number of children in need of day care in the U.S., around 47% are under the supervision of relatives, 32% are under supervision of non-relatives, and 21% receive care from a combination both relatives and non-relatives. Non-relative care is comprised of organized day care services (day care centers), which constitute around two-thirds of total child care, while the rest one-third of day care services are provided by individuals, usually at their residence (home based day care). Of the 12.5 million kids below the age of five, approximately 4.8 million kids are housed in various types of child care centers around the country (Table 1). Of the remaining 7.7 million kids, approximately 2.3 million are cared for by non-relatives in residential facilities, where the number of kids can range from 1 to 12 and which can increase for school aged kids. The average amount of time that preschool aged children with employed mothers spend in a day care is 36 hours per week, while the average for those with unemployed mothers is 21 hours per week.

Figure 1 shows the classification of various types of day care arrangements. The total types of day care centers as defined by individual states is over 100, however a majority of these can be broadly divided into two major types: day care centers and group/family based day care. Day care centers are typically housed in

non-residential facilities, which could be independent buildings or part of another commercial building, while family and group based day care are usually located at the residence of the provider.

Table 1. Summary statistics of day care centers in the U.S.

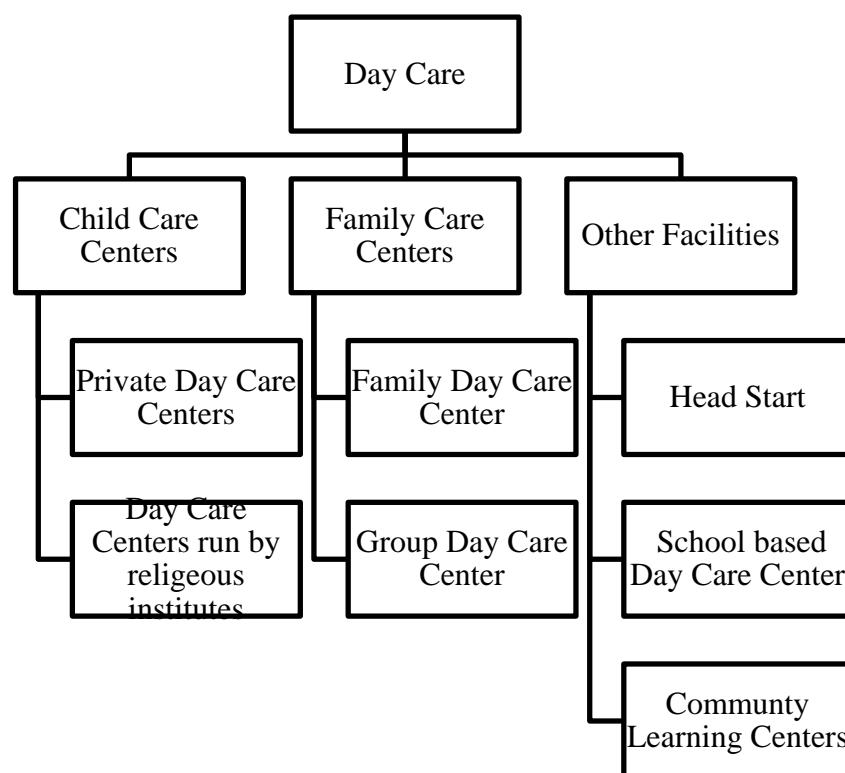| Total Number of Day Care Centers	120K
Total Number of Family Day Care	240K
Number of Pre-School Kids needing day care	12.5 million
% of Pre-school kids who require day care	60%
Average Time for Pre-School kids in Day Care Weekly	36 Hours

Figure 1. Schematic classification of day care types.

Day care centers represent a critical segment of the population, and the requirements and rules related to day care centers in the U.S. varies greatly across states with every state having its own regulating standards. Though most states have some requirements [1,2] for operating day care centers, most of the requirements pertain to the day-to-day operation of the day care center (e.g. ratio of kids to staff, sleeping requirements, space requirements) [3]. However, data regarding the spatial locations of day care centers is scarce. The spatial locations of day care centers is a critical factor in case of emergency response due to natural or manmade hazards, as their occupants will require special procedures and considerations due to their young age and their inability to comprehend orders or understand the situation. Thus it is critical to accurately map locations of all day care facilities for effective rescue and response during emergencies. Though most states maintain some form of record to monitor and track day care centers, this information is not readily available or is available in formats which is not suitable for emergency preparedness purposes.

The objective of this research was to create a geospatial dataset of day care centers (*i.e.* day care facilities), which are in non-residential buildings and care for a large number of kids (from tens to a few hundred). Group and family based day cares were not included, as some states do not provide data about such facilities due to privacy concerns, and they are mostly housed in residential facilities.

The methods employed to create an accurate geospatial database of day care centers for use in emergency response and planning, and for improving population distribution and dynamics, is described in this paper. The novelty of the approach is the integration and mining of widely disparate data sources (mostly open source) including portable data formats (pdf), scanned pages, word documents, and various tabular formats to create a spatial database. The utility of such a dataset is exemplified by identifying high-density clusters of day care centers across the lower 48 states in the USA as hotspots of population at risk as well exposure to environmental pollution. Other factors to consider for emergency response and preparedness and recommendations for further analysis in using the dataset to plan for disaster risk management are also discussed.

In another important context, day care centers are a crucial component of the daily population distribution and need to be accounted for in population distribution and dynamics models. Though statistics from the U.S. Census Bureau remains the primary source of information for population there have been efforts to spatially and temporally refine population distribution [4–6] for better representation of population at risk during disasters and emergencies and for planning purposes. The LandScan USA population distribution and dynamics model [7] developed at Oak Ridge National Laboratory (ORNL) is one such example, which has been used in variety of studies [8–10] to estimate population at risk. The LandScan USA model provides a distribution of the U.S. population at a spatial resolution of three arc-seconds (~90 m cells) for both nighttime and daytime scenarios. There is continuous effort to refine and enhance the LandScan USA model by incorporating new datasets and techniques [11,12]. The creation of the day care center spatial database contributes to the list of datasets [7], which are used to refine and enhance the LandScan USA population model. In addition, the day care dataset also contributes to the Homeland Security Infrastructure Program (HSIP). HSIP is a collection of infrastructure related spatial data, which is used by various communities for emergency response and recovery [13].

3. Methodology

Day care centers in different states are regulated by various departments (Table 2) which can be broadly grouped into five main departments including Children & Family Services, Early Education and Care, Family Services, Health, Human Services, and Social Services. These departments disseminate data online in various formats and with various sets of attributes. Data for the states of Colorado, Delaware, Hawaii, Idaho, Illinois, Kansas, Maryland, Montana, New Hampshire, New Mexico, North Dakota, Oregon, Pennsylvania, Puerto Rice, South Dakota, and Wyoming were not publically available and were obtained directly from the respective state agencies. Data from other states were obtained in various formats that range from pdf, Microsoft Word document and tabular formats or through a web application interface.

The attribution of the dataset also varies considerably across states with some states only providing the name and location of day care centers while others include attribution such as capacity, phone number, age group, compliance reports, hours, manager name and ratings. One of the attributes, critical

for emergency preparedness and planning, is the capacity of the day care center which provides an idea about how many kids are expected to be present in a day care at a given point of time. However, while having a capacity number is useful, it does not necessarily mean that the day care center will be full to its capacity at a given time.

A combination of data extraction tools which include text mining and web-scraping were used to extract the relevant data into tabular format. Over 220,000 records were acquired from the various source documents. Though the goal was to extract only standalone day care centers, many states provide data which includes home and group based day care facilities and do not provide attribution which can be used to classify the day care centers into various sub-types as shown in Figure 1. The total domain for all the day care types was 75, but there was significant overlap between the definitions. All types that indicated a home or residence based day care, such as family home, group home, family child care, and residential child care, were removed from the database. The remaining records (approximately 120,000) were retained for inclusion in the database. These records were regrouped into four distinct types of day care (Center based, School based, Head Start, Religious facility) to remove over-lapping definitions and to group the day care centers into meaningful groups (Table 3).

Gaps from missing attribute data were filled using open web searches until the minimum set of attribution for every day care center was complete, which included the name and complete address of each day care center. The tabular dataset was then geocoded using a parcel-level geocoder to create an initial spatial dataset, addresses for records that could not be geocoded were updated using open source search. Once all the day care centers were geocoded, individual points were moved on the top of a building, using high-resolution imagery and open-source resources thus making sure that points were on-entity.

Table 2. Departments responsible for regulating day care in different states.

State Govt. Department	States
Children & Family Services	LA, NJ, NM, NV, NY, RI, VT, WI
Early Education and Care	DE, GA, MA, MD, OR, SC, WA
Family Services	IN, OH, TX, WY
Health & Human Services	AK, AZ, CT, FL, ID, KS, KY, ME, MO, MS, MT, NC, NE, NH, UT, WV
Human Services	AL, AR, CO, HI, IA, IL, MI, MN, ND, OK, PA, TN
Social Services	CA, SD, VA

Table 3. Classification of day care types.

Type	Description	Examples
Center Based	Day care provided at an independent center	Child Care Center, Day Care Center, Licensed Provider, Licensed Center, Preschool, Early Care and Education, Approved Provider
School Based	Day care provided at a school	School-Age Centers, School-Age Child Care, Child Care Public School
Head Start	Head Start Location	Head Start
Religious Facility	Day care provided at a religious establishment	Registered Child Care Ministry, Religious Facility, Religious Exempt

4. Results

The effort led to the creation of the first ever national level child care center spatial database for the U.S. Though data about child care centers is available from many states, it is not available in a format where it can be used for geospatial analysis. The total number of child care centers per state is shown in Figure 2. It can be observed that the distribution of day care centers closely matches the distribution of the population. As expected, heavily populated states have a higher concentration of day care centers. California and Florida have over 10,000 day care centers while West Virginia, Utah, Montana, and Nevada have less than 300 day care centers. However, normalizing the data by the total number of children that are five years old and younger significantly changes the distribution pattern (Figure 2). States like California and New York, with the largest number of day care centers in the country, have some of the lowest ratios of day care centers to number of children five years and younger indicating that there might be severe shortage of proper day care facilities in these states. The states of Tennessee, Arkansas, and Florida have the highest ratios of day care centers to total number of children aged five and younger. Further investigation of these observations was beyond the scope of this analysis. However, a plausible explanation for the high ratio could be greater number of families with both parents working due to lack of high paying jobs. Similarly, the lower ratios could result from a lesser number of working parents due to availability of high paying jobs or a greater availability of family/residence based private day care centers that are not captured in the statistics.
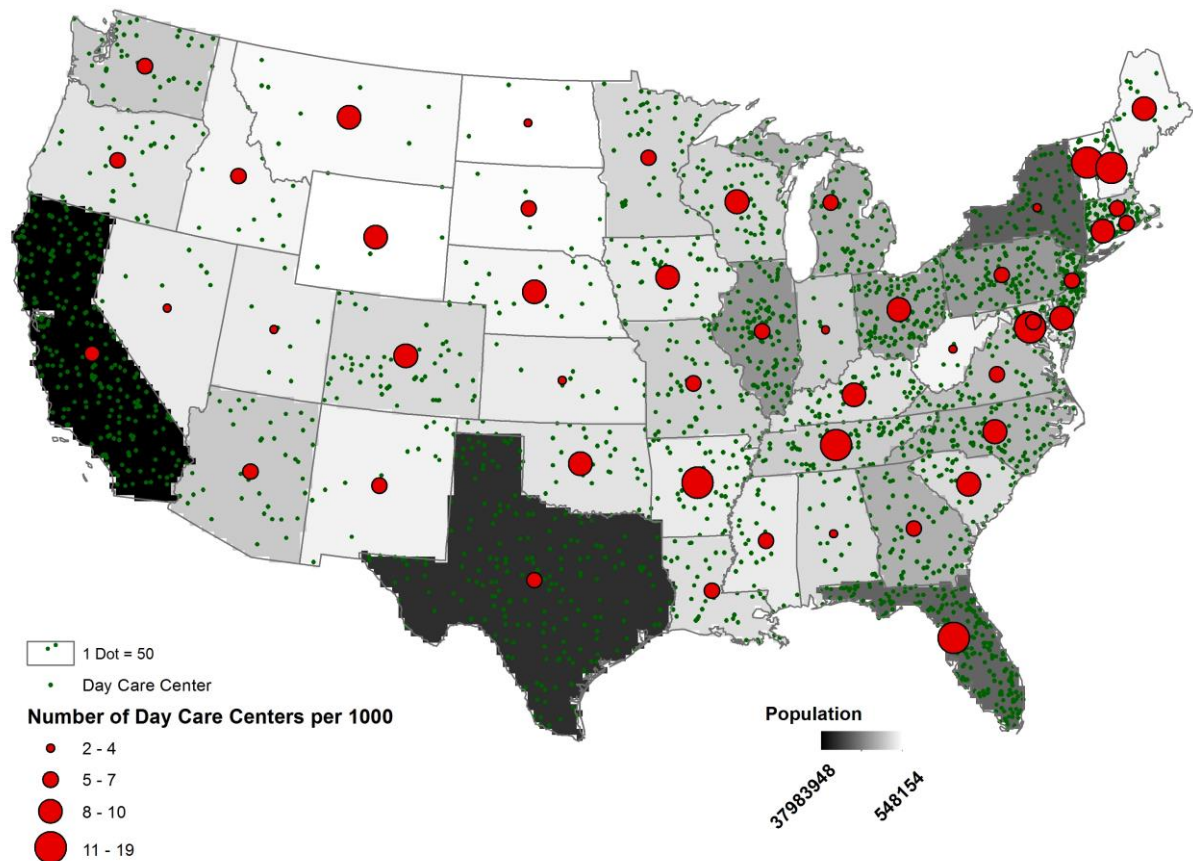

Figure 2. Number and proportion of day care centers by state as compared to population distribution.

In addition to the location of day care centers, another factor to take into consideration for disaster risk management is the total capacity of the day care centers, which indicates the total number of children that could potentially be present in a day care facility at a given point of time. Texas leads the states in the total capacity, its day care centers can accommodate a total of approximately 860,000 kids, while child care centers in California can accommodate around 700,000. States that have a deficit capacity of child care center seats have a higher concentration of family and group based day care.

To further understand the concentration of child care centers in the U.S. and hence identify hot-spots of population at risk, a point density analysis was performed on the spatial dataset. Figure 3 shows the locations and distribution of such hotspots derived from point density mapping of day care centers. It can be seen that these hotspots are coincident with some of the major urban centers in the U.S. The metropolitan regions of San Francisco, Los Angeles, and New York City have the highest density of child care centers followed by Detroit, Baltimore, Chicago, and Houston. The figure shows that the density of day care centers closely follows the pattern of urbanized areas as defined by the U.S. Census Bureau [14]. These urban areas are concentrations of high-density population within the U.S. per the 2010 census.
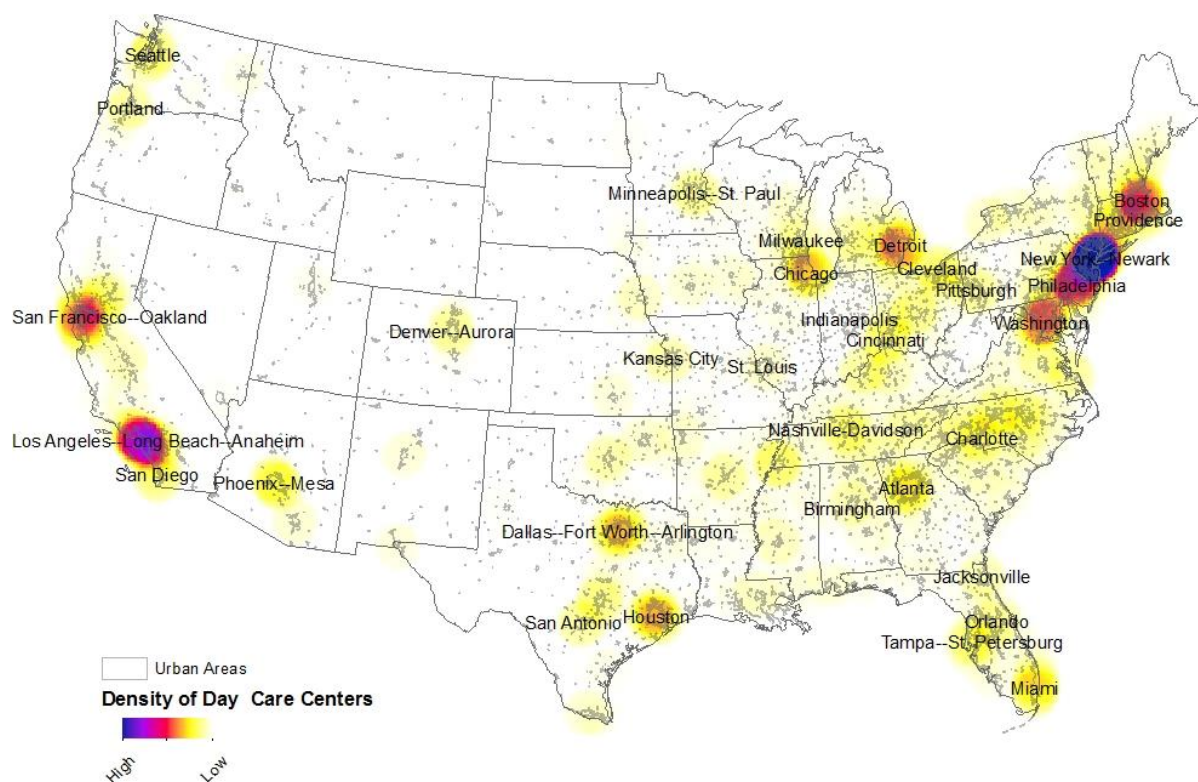

Figure 3. Distribution of day care center hotspots around urban areas as defined by the U.S. Census Bureau.
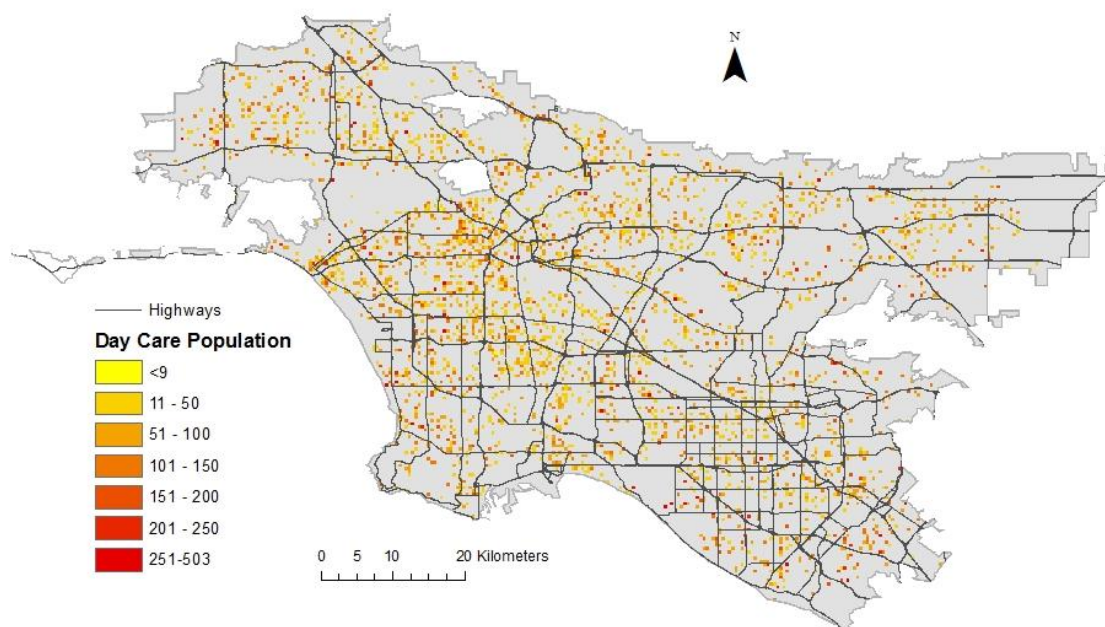
Figure 4. Population distribution of kids in Day Care around highways in the Los Angeles-Long Beach-Anaheim urbanized area in the state of California.

The fact that most of the day care centers are concentrated in dense urban areas expose infants and toddlers to a host of risks associated with densely populated areas such as congested roads, high traffic volumes, air and sound pollution. One of the major concerns [15] is the exposure of kids to toxic fumes especially from trucks. A buffer analysis around interstates reveals that around 4500 child care centers with a total capacity of around 300,000 kids fall within 1000 feet or less of all the interstate highways while around 1800 child care centers with a total capacity of around 123,400 fall just within 500 feet of interstates. The states of California and New York have the largest number of child care centers both within a 1000 and 500 foot buffer of major interstates. Houston *et al.* [16] reported that in California, over 200,000 kids in child care were within 650 feet of major roadways potentially exposing them to harmful pollutants with around 57,000 kids exposed to more than 50,000 vehicles per day. Figure 4 depicts the how the day care dataset can be easily used to map the population distribution of kids in day care centers around highways in the Los Angeles metropolitan and thus its utility for various disaster management and risk analysis studies.

5. Discussions

A spatial dataset of the day care centers for the entire U.S. has been created by combining data from a variety of sources. The task was challenging as available data from over fifty state sources were in disparate formats, structure, and attribution. Even though day care centers represent a concentration of at-risk population, spatial data for day care centers are lacking for most of the states. This is a great hindrance for disaster preparedness and response, which relies on accurate spatial locations for emergency response and management. The day care dataset prepared in this work can be used for risk assessments from both natural and manmade hazards and an effective response can be planned accordingly. It can also be used for evacuation modeling as infants and toddlers represent a special

segment of the population who are dependent on others during emergencies. The dataset can be used to study urban dynamics given the high concentration of day care centers within urbanized areas (Figure 3). The dataset provides a key input to the LandScan USA population distribution and dynamics model which is extensively used for identifying population at risk during emergency preparedness and response. Day care centers represent population ranging from tens to few hundred and thus can have significant impact on population distribution and dynamics. In this study a few examples have illustrated how this dataset can be used to generate hotspots of at risk population in the U.S. In addition, the data can be used to perform risk analysis around the child care centers. This could include assessing risks to potential hazards like proximity to chemical facilities, flood plains, HAZMAT routes (Figure 4).

6. Conclusions

Though most of the data acquired in this study are openly available and distributed through the internet, the format and structure of the available datasets reduces their usefulness for any type of further analysis. For example, nine states distributed data in pdf documents, which were different in structure and content, making it challenging for the common user to convert the data to a geospatial format and extract relevant attributes for further analysis. Some states distributed data through a web interface, which required viewing of several pages before all the relevant information could be obtained for ingestion into a spatial database. In addition, the attributions of datasets varied from just two (name and address of the day care) to over 20. These highlight some of the common challenges associated with freely disseminated data and the need for development and adoption of common standards for data reporting and distribution so that the data is beneficial and used by the community. When a lot of time and effort has to be spent in cleaning and pre-processing data and whole purpose of distributing data openly is defeated.

The available spatial dataset can be readily used for a wide variety of geospatial analysis including disaster management as exemplified in this study. In addition, combined with other population attributes like income, race, and age groups, various socio-economic insights of the population using day care across the country can be gained. In addition, the spatial location of day care centers can be used for monitoring and controlling the spread of diseases like flu and various viruses, which spread through smaller kids or to which young kids are more susceptible to. Other potential use could be monitoring and measuring the amount of environmental pollution kids are exposed to so as to enforce stricter regulations around areas with a high concentration of day care centers. Since the data have been prepared at a national scale, analysis can be done at various scales to compare and contrast national, regional, and local patterns. Future work will involve updating the dataset and mapping smaller group and home based day care as some of those are not well regulated by states but represent a significant number (2.3 million) of kids. Further, an individual residence can house up to 12 kids in some states creating random clusters of population at risk. The challenge with updating and maintaining such a database can be very complicated due the lack of uniformity amongst the source datasets. This is further complicated as it has been observed that sources along with the format, attribution, and structure of the dataset can change over time thus further necessitating the adoption of a uniform standard by all the concerned data producers.

Acknowledgments

Author Contributions

Nagendra Singh and Mark Tuttle performed the data assimilation. Nagendra Singh performed the analysis and created the database. Nagendra Singh, Mark Tuttle and Budhendra Bhaduri conceived and wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Laughlin, L. *Who's Minding the Kids? Child Care Arrangements: Spring 2011*; US Census Bureau: Washington, DC, USA, 2013; pp. 70–135.
2. National Association for Regulatory Affairs. The 50 State Child Care Licensing Study. Available online: http://www.naralicensing.org/Resources/Documents/2011-2013_CCLS.pdf (accessed on 9 July 2015).
3. Fischer, S.; Jana, M. Child Care Licensing Study, 2011. Available online: http://doi.org/10.3886/ICPSR34550.v1 (accessed 15 March 2015)
4. Dobson, J.E.; Bright, E.A.; Coleman, P.R.; Durfee, R.C.; Worley, B.A. LandScan: A global population database for estimating populations at risk. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 849–857.
5. Gallego, F.J. A population density grid of the European Union. *Popul. Environ.* **2010**, *31*, 460–473.
6. Gridded Population of the World (GPW). Available online: http://sedac.ciesin.org/plue/gpw (accessed on 30 April 2015).
7. Bhaduri, B.; Bright, E.; Coleman, P.; Urban, M.L. LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal* **2007**, *69*, 103–117.
8. Rowley, R.J.; Kostelnick, J.C.; Braaten, D.; Li, X.; Meisel, J. Risk of rising sea level to population and land area. *Eos Trans. Am. Geophys. Union* **2007**, *88*, 105–107.
9. Garb, J.L.; Cromley, R.G.; Wait, R.B. Estimating populations at risk for disaster preparedness and response. *J. Homel. Secur. Emerg. Manag.* **2007**, *4*, doi: 10.2202/1547-7355.1280.

10. Elvidge, C.D.; Sutton, P.C.; Ghosh, T.; Tuttle, B.T.; Baugh, K.E.; Bhaduri, B.; Bright, E. A global poverty map derived from satellite data. *Comp. Geosci.* **2009**, *35*, 1652–1660.

11. Graesser, J.; Cheriyadat, A.; Vatsavai, R.R.; Chandola, V.; Long, J.; Bright, E. Image based characterization of formal and informal neighborhoods in an urban landscape. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1164–1176.

12. Jochem, W.C.; Sims, K.; Bright, E.A.; Urban, M.L.; Rose, A.N.; Coleman, P.R.; Bhaduri, B.L. Estimating traveler populations at airport and cruise terminals for population distribution and dynamics. *Nat. hazards* **2013**, *68*, 1325–1342.

13. Jackson, S.; Vaughan, C.; Homeland Security Infrastructure Program 2010. Available online: http://proceedings.esri.com/library/userconf/feduc10/papers/user/hsip_gold_2010_status_esri_fed ucv2_jackson.pdf (accessed on 16 March 2015)

14. U.S Census Bureau. 2010 Census Urban and Rural Classification and Urban Area Criteria. Available online: https://www.census.gov/geo/reference/ua/urban-rural-2010.html (accessed on 9 July 2015).

15. Somers, T.S.; Harvey, M.L.; Rusnak, S.M. Making child care centers SAFER: A non-regulatory approach to improving child care center siting. *Public Health Rep.* **2011**, *126*, 34–40.

16. Houston, D.; Ong, P.; Wu, J.; Winer, A. Proximity of licensed child care facilities to near-roadway vehicle pollution. *Am. J. Public Health* **2006**, *96*. 1611–1617.