Geospatial Health 6(2), 2012, pp. 157-159

Letter to the Editor

Geocoding-protected health information using online services may compromise patient privacy - Comments on "Evaluation of the positional difference between two common geocoding methods" by Duncan et al.

Dear Editor,

I was very excited to read the paper by Duncan et al. (2011), which described the locational accuracy and ease of geocoding address information using online geocoding services, published in *Geospatial Health*. The authors produced a very thorough analysis highlighting the practical utility of Batchgeo, and they promote it as a free and powerful resource for geocoding addresses. Unfortunately, they failed to recognize that the use of online geocoding services such as Batchgeo and ArcGIS Online World Geocoding Service (which were used in their study) may have inadvertently disclosed protected health information to an external organisation.

Location of residence is identifiable information. Passing address information to Batchgeo or other online geocoding services jeopardizes patient privacy because the information may be logged and stored in their data servers. Batchgeo's map data security and privacy policy states that they may record information such as your web request, Internet Protocol (IP) address, and the date and time of the transaction (2011). Static IP addresses are unique and identifiable, and many websites such as Batchgeo use cookies to track users for personalized marketing purposes. Therefore, it is possible for an online geocoding service provider to identify the organisation submitting the geocode request and attribute the list of addresses it processes. Furthermore, this information may be augmented with the date of the geocode request and the dataset in question may become reverse identifiable.

Recognizing the sensitivity of health and finance data, and the requirement to keep the information private and secured, ESRI (host of ArcGIS Online World Geocoding Service) recommends that geocoding of these data be performed using locally-stored reference street address datasets behind a secure firewall (2011). Online geocoding services should not be used for geocoding-protected health information because patient privacy may be compromised and the organisation may be in violation of privacy legislation.

The intent of this letter is not to embarrass the authors of the study, nor is it to vilify online geocoding services (or to suggest that they may have malicious intent to use the information that was inadvertently disclosed to them). Instead, public health practitioners and researchers should be aware that online geocoding services should not be used on sensitive and protected health datasets for the reasons stated above.

References

Batchgeo. "Map data security and privacy policy." Webpage visited December 13, 2009. http://batchgeo.com/features/security/ (accessed on 13 December 2011).

Duncan DT, Castro MC, Blossom JC, Bennett GG, Gortmaker SL, 2011. Evaluation of the positional difference between two common geocoding methods. Geospat Health 5, 265-273.

ESRI, 2009. New geocoding solutions provide many options. ArcNews Online. Spring 2009 http://www.esri.com/news/arcnews/spring09articles/new-geocoding.html (accessed on 13 December 2011).

Sunny Mak

Public Health Analytics, British Columbia Centre for Disease Control
655 West 12th Avenue, Vancouver, BC, V5Z 4R4, Canada
Tel. +1 604 707-2575; Fax +1 604 707-2516
E-mail: sunny.mak@bccdc.ca

Response to Geocoding-protected health information using online services may compromise patient privacy - Comments on "Evaluation of the positional difference between two common geocoding methods" by Duncan et al.

Dear Editor,

We thank Mak for his positive comments about our article previously published in *Geospatial Health* (Duncan et al., 2011), and for bringing attention to the important issue of potentially compromising patient privacy (including protected health information) when geocoding such data using online services. We acknowledge that we did not discuss the issue of confidentiality when geocoding data using online services, and we appreciate the opportunity to briefly reflect on this topic. Assurance in protecting participants' confidentiality is of the utmost importance, and special care should be given to geospatial datasets with individual-level sensitive health information. Accidental sharing of this information may result in job discrimination, and social stigma, to name a few. In this case, use of online geocoding services can inadvertently disclose individual location (and perhaps other) information to an external organisation, since addresses are loaded onto an external server. Even if data storage on the server is temporary and anonymous there is still reason for concern due to breached privacy. Therefore, the nature of online geocoding services may not be suitable for projects with individual-level sensitive/confidential data.

Yet, it is worth mentioning that security procedures are determined by the characteristics of the research. In some projects, including the one discussed in our article, study participants are not patients. Specific security procedures can be specified in a data management plan, approved by an institution's human subjects committee, and described in the process of obtaining informed consent from study participants. In some cases confidentiality can be protected by using a large enough geographic level. Indeed, there could be a certain level of spatial aggregation suitable for online geocoding. For instance, when geocoding is used to find geographical coordinates for zip codes in a dataset using online geocoding services may not compromise study participants' locations. The US Census Bureau considers a census block as the smallest spatial unit at which confidentiality of census respondents can be preserved, but the scale and nature of the data should determine if online geocoding services are suitable or not. We cannot assume that a census block is an adequate unit to protect confidentiality of health data, since one needs to consider several issues that could potentially facilitate the identification of subjects in the block (e.g. the rarity of the health event, and the selectivity of the disease by age group, race and gender). Common sense and care from the researcher/practitioner is absolutely crucial in deciding which services to utilise for geocoding.

We note that several methods to ensure confidentiality in geocoding have been discussed previously, including when using online geocoding services (such as submitting randomised bundles of erroneous as well as real data) (Gittler, 2008; Goldberg, 2008). Also, it is worth noting that disclosure and confidentiality agreements can and should be agreed upon between the submitter of the data and the service provider. We urge public health researchers and practitioners to adapt existing policies, guidelines and protocols for geocoding, develop new ones (as needed), and effectively use them in order to ensure confidentiality. This issue was discussed at the First International Geospatial Geocoding Conference (http://geocodingconference.com/) in December 2011. Given the increasing prominence of online services in the future, we hope that new recommendations will also address circumstances when one should (or should not) use online geocoding.

References

Duncan DT, Castro MC, Blossom JC, Bennett GG, Gortmaker SL, 2011. Evaluation of the positional difference between two common geocoding methods. Geospat Health 5, 265-273.

Gittler J, 2008. Cancer Registry Data and Geocoding: Privacy, Confidentiality and Security Issues. In: Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice. Ruston G, Armstrong

MP, Gittler J, Greene BR, Pavlik CE, West MM, Zimmerman DL (Eds.). CRC Press, Boca Raton, FL, 2008.

Goldberg DW, 2008. Ensuring Privacy and Confidentiality. In: A Geocoding Best Practices Guide. Springfield, IL, North American Association of Central Cancer Registries, Inc. Boca Raton, FL, 2008.

Dustin T Duncan, Marcia C Castro and Jeffrey C Blossom

Harvard School of Public Health
677 Huntington Avenue, Kresge 7th Floor
Boston MA 02115, USA
Tel. +1 617 384-8732; Fax +1 617 384-8730
E-mail: dduncan@hsph.harvard.edu