

Assessment of Geographic Information Systems and Data Confidentiality Guidelines in STD Programs

JENNIFER M. BISSETTE, BS^a
JEFFREY A. STOVER, MPH^{a,b}
LORI M. NEWMAN, MD^c
PHILIP CHRISTOPHER DELCHER,
MS^{a,d}
KYLE T. BERNSTEIN, PhD, ScM^e
LINDSEY MATTHEWS, MPH^a

SYNOPSIS

Objective. Advancements in technology, such as geographic information systems (GIS), expand sexually transmitted disease (STD) program capacity for data analysis and visualization, and introduce additional confidentiality considerations. We developed a survey to examine GIS use among STD programs and to better understand existing data confidentiality practices.

Methods. A Web-based survey of eight to 22 questions, depending on program-specific GIS capacity, was e-mailed to all STD program directors through the National Coalition of STD Directors in November 2004. Survey responses were accepted until April 15, 2005.

Results. Eighty-five percent of the 65 currently funded STD programs responded to the survey. Of those, 58% used GIS and 54% used geocoding. STD programs that did not use GIS (42%) identified lack of training and insufficient staff as primary barriers. Mapping, spatial analyses, and targeting program interventions were the main reasons for geocoding data. Nineteen of the 25 programs that responded to questions related to statistical disclosure rules employed a numerator rule, and 56% of those used a variation of the "Rule of 5." Of the 28 programs that responded to questions pertaining to confidentiality guidelines, 82% addressed confidentiality of GIS data informally.

Conclusions. Survey findings showed the increasing use of GIS and highlighted the struggles STD programs face in employing GIS and protecting confidentiality. Guidance related to data confidentiality and additional access to GIS software and training could assist programs in optimizing use of spatial data.

^aOffice of Epidemiology, Division of Disease Prevention, Virginia Department of Health, Richmond, VA

^bDepartment of Epidemiology and Community Health, School of Medicine, Virginia Commonwealth University, Richmond, VA

^cDivision of STD Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA

^dVirginia Health Information, Richmond, VA

^eSan Francisco Department of Public Health, San Francisco, CA

Address correspondence to: Jennifer M. Bissette, BS, Office of Epidemiology, Division of Disease Prevention, Virginia Department of Health, 109 Governor St., Richmond, VA 23219; tel. 804-864-8004; fax 804-864-7970; e-mail <Jennifer.Bissette@vdh.virginia.gov>.

©2009 Association of Schools of Public Health

Since the 1990s, interest in geographic information systems (GIS) has increased within the public health community, and GIS has become increasingly more common within the epidemiologist's toolbox. By definition, GIS can be regarded as an automated system for the capture, storage, retrieval, analysis, and display of spatial data.¹ "Spatial data" refers to any data that can be mapped.² For example, a communicable disease and correlated health event information, including geographic location and other health-related attributes such as socioeconomic status, can be considered spatial data.³ A growing number of sexually transmitted disease (STD) programs are realizing GIS enhances surveillance activities by providing the ability to manage spatial data, identify and map patterns of disease, and analyze spatial relationships such as health-related disparities and resources.

Closely integrated with GIS software is geocoding technology, which is the process of turning descriptive location-specific data (e.g., a postal address) into an absolute geographic reference,⁴ also referred to as latitudinal and longitudinal points. Prior to widespread use of geocoding, zip codes were the smallest geographic area of detail commonly displayed on maps. With geocoded coordinates, the exact map location (within a certain range of error) of a place, such as a patient's residence, can now be determined. Geocoding can also serve to improve and standardize addresses and ensure appropriate jurisdictional assignment of morbidity.⁵

GIS software capabilities have evolved rapidly, enhancing public health analysis, visualization, and spatial reporting mechanisms. Advancements in GIS have also created a new dilemma for a long-standing public health practice—the use of statistical disclosure limitation methods to protect patient confidentiality. Ensuring patient confidentiality while using spatially referenced data to its full public health potential can be challenging. In many instances, point-level data may be the most representative display of information; however, point-level data also represent an increased risk of inadvertent disclosure. Most statutes that govern data confidentiality and/or sharing practices were developed before GIS became widely used,⁶ which serves as an indicator that GIS technology has outpaced STD data confidentiality standards of practice.

Outcome Assessment through Systems of Integrated Surveillance (OASIS) was a demonstration project funded by the Centers for Disease Control and Prevention (CDC) from 1998 through 2005 to promote innovative surveillance techniques. Many of the 17 STD project areas (STD programs funded by CDC at the state and/or city level) funded through OASIS implemented the use of GIS and geocoding technology

through OASIS initiatives. The use of GIS and geocoding technology enabled some OASIS participants to begin mapping to exact geocoded coordinates (or points) and provided an impetus for enhanced spatial analysis capacity.

OASIS participants were confronted with issues related to data confidentiality and disclosure limitation techniques as their use of GIS technology expanded. In July 2004, an informal GIS workgroup was created inclusive of the following OASIS sites: Baltimore, California, Massachusetts, Michigan, New York State, San Francisco, Virginia, and Washington State. CDC's Division of STD Prevention also participated in the workgroup. The specific aims of the workgroup were to (1) discuss and address STD issues related to GIS/geocoding, (2) determine the extent of GIS technology and related confidentiality standards in use in STD programs nationally, and (3) develop recommendations for GIS-related best practices that could be shared with other STD programs.

This article provides a descriptive analysis of a survey conducted with STD programs to assess the use of GIS technology and to gain a better understanding of variance related to data confidentiality practices.

METHODS

STD program staff from the Virginia Department of Health (VDH), in collaboration with the OASIS GIS workgroup, developed and distributed a survey to all STD programs nationally to assess use, variability, and capacity to conduct GIS-related activities. The survey instrument was Web-based and designed to store data upon input, allowing STD programs to temporarily suspend survey completion without loss of information. Survey questions were categorized into six sections (Program summary, GIS, Geocoding, Mapping, Statistical disclosure rules, and Confidentiality guidelines) and pertained to software applications used, technology employed, types and methods of GIS data distribution, data confidentiality standards, and existence of written confidentiality guidelines.

Skip logic was used to guide respondents through the survey based on their program's degree of GIS capacity. As a result, the number of questions elicited from a given program varied from eight to 22. Based on a pilot survey conducted internally by VDH staff, the average time for survey completion was approximately two minutes for STD programs not using GIS and seven to eight minutes for programs using GIS. All respondents were directed to program-specific questions as well as GIS and geocoding questions. Programs that did not use GIS or geocoding software were directed

to follow-up questions. All programs that used GIS were directed to the mapping section; however, only programs that geocoded data were asked about point-level mapping. STD programs were considered more restrictive if maps were only provided to state and local health departments and e-mail was not used as a mode of distribution. An example of a numerator rule, whereby data are not released if the frequency of the count of STD cases is less than a predetermined size or threshold, preceded the questions related to statistical disclosure rules. Only programs that responded to the geocoding question(s) were directed to the questions related to statistical disclosure rules. All programs that indicated use of GIS were directed to questions pertaining to confidentiality guidelines.

The National Coalition of STD Directors, on behalf of the OASIS GIS workgroup, distributed the survey via e-mail in November 2004. The e-mail, sent to all STD program directors, discussed the rationale for the survey and provided an embedded link to the Web-based instrument. Reminder e-mails were distributed in December 2004 to encourage survey participation. From January to March 2005, members of the GIS workgroup made final attempts by phone to contact the remaining programs that had not responded. These efforts resulted in the completion of 21 additional surveys. The survey was closed on April 15, 2005.

RESULTS

Fifty-five (85%) of the 65 federally funded STD programs responded to the survey. Six (11%) programs did not answer all survey questions, and two (4%) answered only the first four questions. The latter two sites were excluded from analyses.

Of the 53 programs in the analysis, 31 (58%) indicated the use of GIS. The length of time GIS had been in use varied; however, of those programs using GIS, 14 (45%) had used it for more than four years. The majority ($n=18$, 58%) of programs using GIS stated that STD program staff performed GIS functions. ArcGIS⁷ software was used by 22 (71%) of the STD programs that indicated use of GIS (Table 1). Nine (29%) STD programs indicated the use of two or more software programs for GIS activities; MapPoint[®],⁸ Street Atlas USA,⁹ or QAS¹⁰ were used only in conjunction with other GIS software programs. Programs that did not use ArcGIS were three times as likely to use more than one software program for GIS activities compared with programs that used ArcGIS.

Twenty-two (42%) of the STD programs that responded to the survey did not use GIS. Lack of training and insufficient staff were the primary reasons

for not using GIS. Eight STD programs identified budgetary constraints as a barrier to GIS use. Twelve STD programs indicated GIS would be employed if the technology and training were made available. Five STD programs indicated the need for more information on GIS capacity and its advantages (Table 1).

All STD programs that geocoded data used GIS applications; however, three (10%) of the 31 programs that used GIS did not geocode. The majority (54%) of the 28 STD programs that geocoded used StreetMap¹¹ for this activity. The primary uses for geocoding data were mapping and spatial analyses. The majority of respondents (71%) also indicated geocoding technology was used to target intervention programs.

All 31 programs that used GIS were directed to aggregate-level mapping questions regardless of geocoding ability (Table 1). For survey purposes, aggregate-level mapping referred only to disease counts. Of these programs, 90% indicated the ability to map aggregate data. Only programs that used both GIS and geocoding ($n=28$) were directed to point-level mapping questions. Of these 28 programs, 75% mapped point data. Programs that mapped data were asked to indicate data dissemination practices to assess confidentiality issues (Table 2).

Nineteen programs mapped both point- and aggregate-level data. Of these, 14 (74%) were more restrictive regarding map recipients of point-level maps when compared with aggregate-level maps, or were equally restrictive of both. In terms of the method for distributing point-level maps, 11 of the 19 STD programs (58%) were more restrictive of point-level maps when compared with aggregate-level maps or were equally restrictive.

Among the 31 STD programs that used GIS, 28 (90%) responded to the questions related to confidentiality guidelines. Twenty-two of the programs did not have written confidentiality guidelines inclusive of GIS and geocoding activities; however, 15 of the 22 programs (68%) indicated that confidentiality of GIS data was addressed informally.

Twenty-five programs responded to the statistical disclosure rule questions, with 76% indicating use of some form of a numerator rule. Twenty percent of programs indicated use of an "other" rule, and one program indicated no rule was used. The most common rule used was the "Rule of 5," with 56% of the programs using a variant of this rule. The programs using a "Rule of 5" defined the rule four distinct ways (0-4, 1-4, 0-5, or 1-5). Other numerator rules included 0, 3, 6, and 10. The 24 programs that reported using a confidentiality rule were asked to indicate the geographic granularity to which the rule was applied. Sixty-three percent

Table 1. Summary results from the GIS and Data Confidentiality Guidelines in STD Programs Survey, 2004–2005

	Number	Percent
Total number of STD programs	65	100
Survey responses received	55	85
Survey responses not received	10	15
GIS		
GIS not used by STD programs	22	42
Reason(s) GIS not used by STD programs ^a		
Budget constraints	8	38
Lack of training	11	52
No program need	4	19
Not a priority	7	33
Insufficient staff	11	52
Other	3	14
Use of GIS if technology and training were available		
Yes	12	57
No	3	14
Other	6	29
GIS used by STD programs	31	58
GIS software used ^a		
ArcGIS (ArcView 3x and higher) ^b	22	71
Epi Info™ EpiMap ^c	5	16
MapInfo ^d	6	19
SAS/GIS ^e	6	19
Other	5	16
Geocoding		
Geocoding not used by STD programs	24	46
Geocoding used by STD programs	28	54
Geocoding software used ^a		
Geostan ^f	6	21
External vendor	2	7
MapInfo ^d	4	14
MapMarker ^g	4	14
Matchmaker ^{gh}	1	4
StreetMap ⁱ	15	54
Other	4	14
Uses of geocoding by STD programs ^a		
Mapping	26	93
Study of socioeconomic/educational relationships with diseases	8	29
Spatial analyses of disease distribution	25	89
Address corrections	6	21
Target program interventions	20	71
Other	1	4
Mapping		
Point data used	21	75
Aggregate data used	28	90

continued on p. 62

of STD programs applied their respective numerator rules at the zip code level. Numerator rules were also applied at more refined geographic areas, including census tracts (42%) and census block groups (29%). These latter two areas are statistical subdivisions of counties used by the U.S. Census Bureau for data reporting. Census tracts are smaller than zip codes

but comprise a larger geographic area and population than block groups.

DISCUSSION

The results of this survey indicated a growing trend of GIS use among STD programs related to surveillance

Table 1 (continued). Summary results from the GIS and Data Confidentiality Guidelines in STD Programs Survey, 2004–2005

	Number	Percent
Statistical disclosure (numerator) rules		
Numerator rule used	25	80
Rule of 10	1	4
Rule of 6	1	4
Rule of 5	14	56
$0 \leq X \leq 4$	1	4
$1 \leq X \leq 4$	5	20
$0 \leq X \leq 5$	4	16
$1 \leq X \leq 5$	4	16
Rule of 3	2	8
$1 \leq X \leq 2$	2	8
Rule of 0	1	4
Confidentiality guidelines ⁱ		
STD programs with written guidelines/procedures	5	18
STD programs with unwritten guidelines	16	70

*STD programs could choose multiple responses.

^bEnvironmental Systems Research Institute, Inc. ArcGIS: a complete integrated system [cited 2009 Jul 3]. Available from: URL: <http://www.esri.com/software/arcgis/index.html>

^cCenters for Disease Control and Prevention (US). Epi Info™ [cited 2009 Jul 3]. Available from: URL: <http://www.cdc.gov/epiinfo>

^dPitney Bowes Business Insight [formerly Group 1 Software & MapInfo]. MapInfo Professional [2009 Jul 3]. Available from: URL: <http://www.pbinsight.com/products/location-intelligence/applications/mapping-analytical/mapinfo-professional>

^eSAS Institute, Inc. SAS®/GIS [cited 2009 Jul 3]. Available from: URL: <http://www.sas.com/products/gis>

^fGroup 1 Software, a Pitney Bowes Company. GeoStan address correction and geocoding solution [cited 2009 Jul 3]. Available from: URL: http://www.pbinsight.com/files/resource-library/resource-files/GeoStan_Data_sheet.pdf

^gPitney Bowes Business Insight [formerly Group 1 Software & MapInfo]. Product documentation: MapMarker and MapMarkerPlus US [cited 2009 Jul 3]. Available from: URL: <http://www.pbinsight.com/support/product-documentation/m/details/mapmarker-mapmarker-plus-us>

^hTele Atlas. Matchmaker® SDK Professional [cited 2009 Jul 3]. Available from: URL: http://216.107.234.202/stellent/groups/public/documents/content/ta_ct015536.pdf

ⁱEnvironmental Systems Research Institute, Inc. StreetMap Premium [cited 2009 Jul 3]. Available from: URL: <http://www.esri.com/data/streetmap/index.html>

^jInclusive of GIS/geocoding data

GIS = geographic information systems

STD = sexually transmitted disease

efforts. However, various barriers exist that limit the use of GIS for enhancing surveillance activities.

The cost of purchasing and maintaining a GIS application and/or appropriately trained staff can be substantial. During times of fiscal uncertainty, the costs associated with GIS may be seen as an unnecessary expense and a means of cost savings. However, the ability to conduct spatial analysis—including ecologic distributions of health events, modeling of geographic risk-factor distributions, and evaluating spatial patterns of service utilization³—presents added value for surveillance efforts. The availability of GIS software licenses, similar to CDC's SAS® licensing agreements,¹² as well as grantor-provided training would help minimize the financial burden of GIS, improve process standardization, and provide for parallel access to GIS software for all STD programs.

The implementation of appropriate controls for confidentiality and protection of data is essential to maintain the trust and support of the public.¹³ Most STD programs restricted the use of point-level GIS data and/or excluded small-scale data from reports, presentations, and data releases in an effort to protect personally identifiable information. However, nearly 43% of respondents that create point-level maps indicated dissemination of such maps via e-mail.

The use of numerator rules for the protection of data confidentiality varied significantly among STD programs. Such variance in data dissemination protocols adds an array of complexity when data are analyzed at regional or national levels. Previous research investigating the complexity and variation of STD statistical disclosure practices documented more than 15 distinct data release procedures within a single national STD dataset.¹⁴

Table 2. Spatial data dissemination practices: results from the GIS and Data Confidentiality Guidelines in STD Programs Survey, 2004–2005

	Point-level maps (n=21)		Aggregate-level maps (n=27 ^a)	
	Number ^b	Percent	Number ^b	Percent
Map recipients				
Local health departments	16	76	23	85
State health departments	13	62	15	56
Community-based organizations	4	19	18	67
Legislators	3	14	11	41
Universities/other schools	0	0	9	33
Police department	0	0	1	4
AIDS service organizations	0	0	10	37
Students	0	0	10	37
Other	4	19	12	44
Method of map distribution				
Regular mail	5	24	15	56
E-mail	9	43	23	85
Secure-access Web server	4	19	11	41
Courier	1	5	3	11
Person-to-person	14	67	18	67
Other	2	10	4	15

^aOne STD program indicated the ability to map aggregate data but did not respond to subsequent questions.

^bSTD programs were able to choose multiple responses.

GIS = geographic information systems

STD = sexually transmitted disease

AIDS = acquired immunodeficiency syndrome

The powerful tools of GIS create new avenues for inadvertent patient-level disclosure of information. Formalized guidance documentation will assist site-specific STD programs with confidentiality issues, but will likely not address cross-program analyses related to variance in statistical disclosure protocols. Maintaining a high level of patient confidentiality is imperative as good data stewards and for maintaining public confidence and perception. This will require balancing data safeguards.

Limitations

Some limitations were associated with this survey. First, the survey was conducted more than five years ago. Therefore, use of GIS and confidentiality protocols may have changed since administration of the survey. Second, the findings may have an inherent bias in that the survey data exclude some programs. Third, knowledge of GIS terminology among survey respondents is unknown with respect to questions such as spatial analysis, for example. The final limitation relates to the ability of respondents to stop the survey and return to it at a later date. This feature was included in the survey design to encourage completion, although some respondents may have failed to return to the survey.

CONCLUSIONS

This survey demonstrated a trend toward increased use of GIS within STD surveillance programs. Based on this trend, we can infer that programs are identifying the added value of GIS within their surveillance efforts. However, apprehension associated with GIS persists, as STD programs continue to (1) toil with affordability of new tools that enhance surveillance, (2) maintain the balance of using and sharing data while maintaining patient confidentiality, and (3) develop guidelines that include new technological advances. Greater consistency related to GIS and data confidentiality will assist in enhancing surveillance efforts and provide improved continuity of data for public health practice, research, and learning.

REFERENCES

1. Clarke KC, McLafferty SL, Tempalski BJ. On epidemiology and geographic information systems: a review and discussion of future directions. *Emerg Infect Dis* 1996;2:85-92.
2. Wade T, Sommer S, editors. *A to Z GIS: an illustrated dictionary of geographic information systems*. 2nd ed. Redlands (CA): ESRI Press; 2006.
3. Cromley E, McLafferty S. *GIS and public health*. New York: Guilford Press; 2002.
4. Goldberg D, Wilson J, Knoblock C. *From text to geographic*

- coordinates: the current state of geocoding. *URISA Journal* 2007;19:33-46.
5. Stover JA, Kheirallah KA, Delcher PC, Dolan CB, Johnson L. Improving surveillance of sexually transmitted diseases through geocoded morbidity assignment. *Public Health Rep* 2009;124(Suppl 2):65-71.
 6. State of Oregon, Joint Task Forces from the Administrative Boundaries & Cultural and Demographics Framework Implementation Teams. GIS and confidentiality, OGIC document number XX. Salem (OR): Oregon Administrative Boundaries & Cultural and Demographics Framework Implementation Teams; 2002. Also available from: URL: http://www.oregon.gov/DAS/EISPD/GEO/docs/ogic/GIS_Confid3.pdf [cited 2009 Jul 3].
 7. Environmental Systems Research Institute, Inc. ArcGIS: a complete integrated system [cited 2009 Jul 3]. Available from: URL: <http://www.esri.com/software/arcgis/index.html>
 8. Microsoft Corp. MapPoint® [cited 2009 Jul 3]. Available from: URL: <http://www.microsoft.com/mappoint/en-us/features.aspx>
 9. DeLorme. Street Atlas USA [cited 2009 Jul 3]. Available from: URL: http://shop.delorme.com/OA_HTML/DELibeCCtpSctDspRte.jsp?section=10120&minisite=10020
 10. Experian QAS. QAS software [cited 2009 Jul 3]. Available from: URL: <http://www.qas.com>
 11. Environmental Systems Research Institute, Inc. StreetMap Premium [cited 2009 Jul 3]. Available from: URL: <http://www.esri.com/data/streetmap/index.html>
 12. Centers for Disease Control and Prevention (US). Sexually transmitted diseases: SAS® license request instructions [cited 2009 Jul 3]. Available from: URL: <http://www.cdc.gov/std/SAS>
 13. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect* 2004;112:998-1006.
 14. Delcher PC, Edwards KT, Stover JA, Newman LM, Groseclose SL, Rajnik DM. Data suppression strategies used during surveillance data release by sexually transmitted disease prevention programs. *J Public Health Manag Pract* 2008;14:E1-8.