MEASURE Evaluation

Working Paper Series

Overview of Issues Concerning Confidentiality and Spatial Data

MEASURE GIS Working Group
August 2008

WP-08-106

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) through Cooperative Agreement No. GPO-A-00-03-00003-00 and is implemented by the Carolina Population Center at the University of North Carolina in partnership with Constella Futures, John Snow, Inc., Macro International Inc., and Tulane University.

Carolina Population Center
University of North Carolina at Chapel Hill
206 W. Franklin Street
Chapel Hill, NC 27516
Phone: 919-966-7482
Fax: 919-966-2391
measure@unc.edu
www.cpc.unc.edu/measure

Printed on recycled paper

This working paper series is made possible by support from the U.S. Agency for International Development (USAID) under Cooperative Agreement No. GPO-A-00-03-00003-00. The opinions expressed are those of the authors, and do not necessarily reflect the views of USAID or the U.S. government.

The working papers in this series are produced by MEASURE Evaluation in order to speed the dissemination of information from research studies. Most working papers currently are under review or are awaiting journal publication at a later date. Reprints of published papers are substituted for preliminary versions as they become available. The working papers are distributed as received from the authors. Adjustments are made to a standard format with little further editing.

This and previous working papers are available, free of charge, from the MEASURE Evaluation Web site, http://www.cpc.unc.edu/measure.

Overview of Issues Concerning Confidentiality and Spatial Data

MEASURE GIS Working Group

Suggested Citation: Overview of Issues Concerning Confidentiality and Spatial Data; MEASURE GIS Working Group, MEASURE Evaluation, University of North Carolina at Chapel Hill; 2007

Acknowledgements: The MEASURE GIS Working acknowledges the contribution of Ijlal Naqvi in the development of this White Paper. Valuable contributions were made by Andrew Inglis and John Spencer.

This report was made possible by support from the U.S. Agency for International Development (USAID) to the MEASURE Evaluation project under terms of Cooperative Agreement GPO-A-00-03-00003-00. *The authors' views expressed in this publication do not necessarily reflect the views of USAID or the United States Government.*

Overview of Issues Concerning Confidentiality and Spatial Data

The increase in the use of spatial data within the MEASURE project has meant a powerful tool, geographic information systems (GIS), has been added to the arsenal available to those using MEASURE data. However, the use of spatial data means that the unique issues regarding confidentiality and privacy that exist with spatial data must be addressed. This white paper presents the current literature on the topic of confidentiality and spatial data. It is intended to provide guidance to the project on the issue, but not a one-size-fits-all solution. The document begins by providing an overview of the terms that are important to the discussion, and then presents some examples of spatial risks to confidentiality. An overview of approaches that have been proposed for preserving confidentiality is then presented, and finally the implications for MEASURE work and possible strategies are presented.

Definition of Terms

The concepts at the core of this discussion, confidentiality and privacy are vague and often have different meanings in different settings. However, the definitions of privacy and confidentiality from the 1971 President's Commission on Federal Statistics are the standard reference and are used by Duncan et al (1993), Golden et al (2005), and the Federal Committee on Statistical Methodology (2005). In the 1971 President's Commission report, privacy is "the individual's right to decide whether and to what extent he will divulge to the government his thoughts, opinion, feelings, and facts of his personal life" (President's Commission 1971: 197). A violation of confidentiality refers to “the transmittal of personal information by someone other than the identified individual,” while confidentiality restrictions “should always mean that a) Disclosure of data in a manner that would allow public identification of the respondent or would in any way be harmful to him is prohibited and b) Data are immune from legal process” (President’s Commission 1971:222). A key difference is that privacy is solely an individual right whereas confidentiality can also apply to corporations. In using the above definitions it is worth taking into account the cautionary note from the American Statistical Association’s Committee on Privacy and Confidentiality that these terms do not have universally agreed upon definitions[1]. For the purpose of clarity in this paper, the term confidentiality will be used and refer to data and information that might identify a person in a way that might cause harm or otherwise violate agreements made with respondents.

Deductive Disclosure

Risks to confidentiality do not necessarily come from blatant release of sensitive data such as names or addresses. It is possible that identities of individuals or other breaches of confidentiality can occur when enough information is released which makes it possible to piece together various data elements to gain insight or knowledge that should be kept secret. This is known as inferential or deductive disclosure. Inferential (or deductive) disclosure can be defined as when “the released data make it possible to determine the value of some characteristic of an individual more accurately than otherwise would have been possible” (Duncan et al 1993). As disclosure risk can never

[1] [1]American Statistical Association Privacy and Confidentiality Committee web page on “Key Terms and Definitions”
http://www.amstat.org/comm/pc/ASA-P&C-Committee-Terms.htm
Accessed 8/14/06

be reduced to zero without withholding (or not collecting) data, an acceptable risk level must then be determined. VanWey et al (2005) present a geospatial example with a 0.05 risk level for illustrative purposes, but expect that actual values would be lower. However it is not always possible to quantify risk and it is sometimes necessary to estimate risk in a less quantitative manner and rely on the analyst's judgment. This judgment will be informed by familiarity with the data and the populations involved.

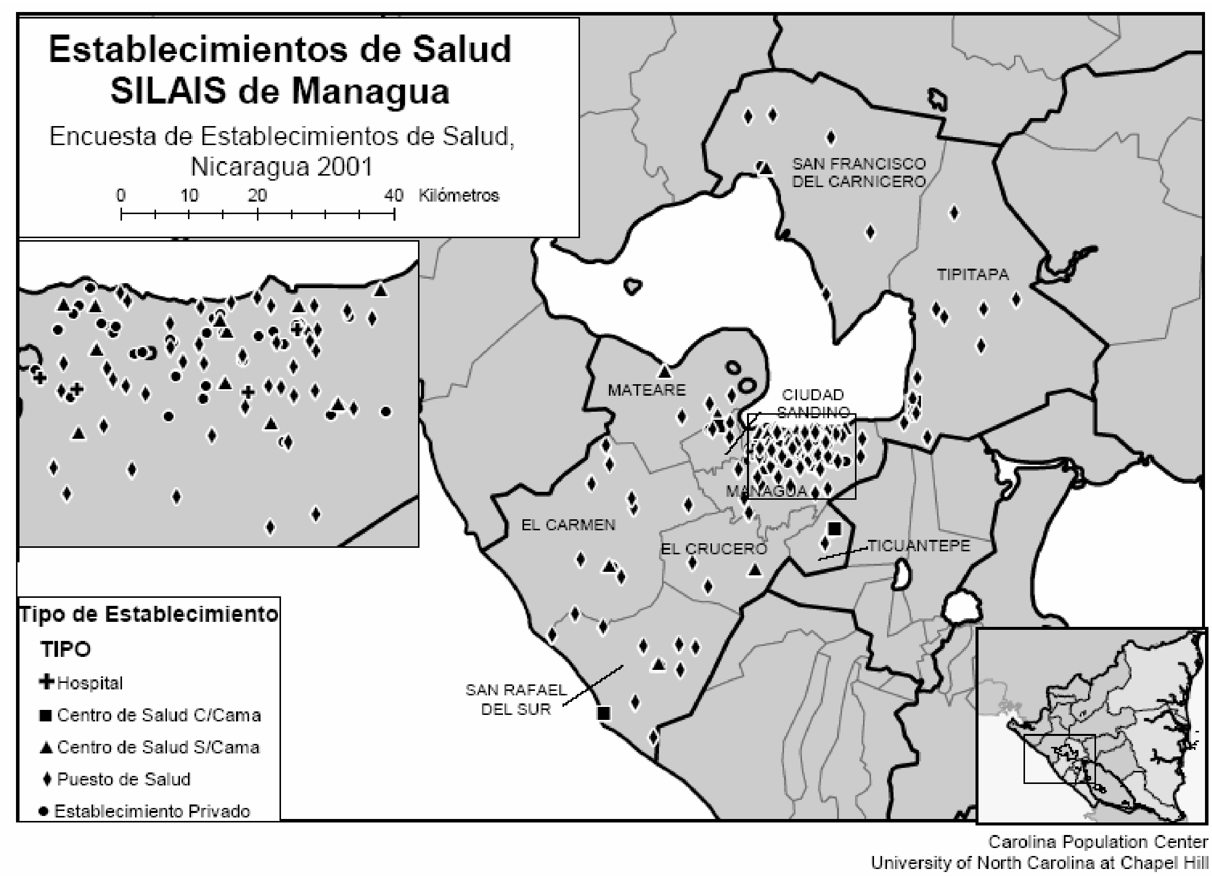
Statistical data can be defined as "A record in a system of records maintained for statistical research or reporting purposes only and not used in hole or in part in making any determination about what an identifiable individual, except as provided by Section 8 [which authorizes certain kinds of data access, including research activities by the bureau of the Census] of Title 13" of the Privacy Act of 1974 (Duncan et al 1993:24, parentheses in original).

MEASURE's Spatial Data

The MEASURE project collects spatial data in a variety of ways, through original data collection such as surveys as well as secondary data collection such as obtaining spatial data from national governments or third parties. Not all of the spatial data maintained by MEASURE pose a risk to violations of confidentiality. But risks of disclosure can arise when disparate spatial data sets are combined. For instance, combining data from a household survey that collects demographic information on respondents can be combined with a dataset from a national statistical agency and could allow people to identify an individual.

MEASURE uses the spatial data it collects in a variety of ways. Each have their own risks with regard to confidentiality. Some collected data is used solely to produce maps that are included in publications or reports. (FIGURE 1). The scale and type of data being displayed will determine the risks to confidentiality breaches. For instance, by itself the scale of the map displayed in figure 1 isn't sufficient to uniquely identify a facility. However, if combined with other data, it could be possible to identify a facility and link it to the survey data.

FIGURE 1 -- Dot map showing health facilities in Nicaragua



MEASURE maintains some of its data and only uses it for in house analysis. Lastly some of its data is released. Each of these uses pose varying degrees of risks. Generally speaking, the broader the release of data, the greater the risk of deductive disclosure.

Often when MEASURE collects primary data, a promise is made to respondents that the information provided will not be used in a way that will uniquely identify the respondents. It is important to honor this promise and any method used to protect confidentiality must adhere to the promises made in the informed consent statement. Even in cases where there is not an informed consent statement, ethical considerations dictate that precautions be taken to protect confidentiality.

Spatial Data and Confidentiality

By definition, a complete spatial coordinate provides uniquely identifying information. A complete coordinate, such as 48.858889° North Latitude and 2. 295833° East Longitude refer to only one location on Earth, in this case the Eiffel Tower in Paris. Likewise complete coordinates collected for MEASURE, provide within a certain margin of error, the location of health facilities, DHS clusters, potential high transmission sites for HIV among others. The data tied to the coordinate can be associated with a specific location on earth.

In addition to coordinates, spatial data can include things such as administrative units such as census enumeration area or wards, addresses, or physical descriptions (e.g.

two miles north of the lake). To varying degrees, this type of spatial data can impact confidentiality of MEASURE data.

The relationship between spatial data and the ethical responsibilities to preserve confidentiality is a complicated one, and there has been much recent discussion of the topic. The Oct 2005 paper by VanWey et al published in the Proceedings of the National Academy of Sciences is perhaps the best recent overview of the topic. In the paper, the authors present four principles that they contend must “guide the collection, analysis, publication, distribution and archiving of spatially explicit micro-, social science data…” (VanWey et al, 2005). These four principles are: protection of confidentiality, social-spatial linkage, data sharing, data preservation and all four have direct relevance to the data collected by the MEASURE project.

To summarize VanWey et al’s four principles: **Protection of Confidentiality** refers to a researcher’s obligation to not release data that might result in harm to individuals. **Social-Spatial Linkage** concerns the necessity that the link between the social and spatial be maintained for use by researchers beyond the original research team. **Data Sharing** is a fundamental concept that underlies most social science work, and spatial location is important to allow others to replicate findings and methodologies and link data. Lastly, the authors assert that **Data Preservation** is important make sure that data collected is available to future researchers.

Reviewing the Literature on Confidentiality

Responses to the problem of protecting confidentiality while allowing access to data can be split into the two general categories of restricting either access or data. In the broadest terms, restricting access allows for more information to be released but imposes costs of time as well as effort which can prevent some researchers from working with the data. Restricted data is cheaper and easier for users to access, but not always straightforward to use. Restricting data can also permit the data provider to also save money by eliminating bureaucratic, time intensive procedures to manage access restrictions, although producing data which protect confidentiality while retaining data integrity does require non-trivial effort. Within both approaches there are considerable variations, and their application can be combined to suit the needs of statistical agencies and data users. This section will describe restricted access and restricted data in turn, going through the variations of each with examples and references for illustration. Not all of the following necessarily apply to data collected for MEASURE, but are presented for illustrative purposes.

Restricted Access

Site Access

The simplest form of restricted access is primarily physical in that the data can only be accessed through travel to the site where it is held. As an example, the US Census Bureau has established six Research Data Centers (RDCs) to make access less onerous for researchers by maintaining locations which are geographically distributed throughout the US. Other US agencies with data centers include the National Center for

Health Statistics and the Agency for Healthcare Research and Quality. RDCs are at sites controlled by the statistical agency and staffed by their employees. Researchers must have their projects approved, enter into a formal contract with the agency, and often pay some fees towards the cost of the center (de Wolf 2003).

Licensed Access

Licensed access allows a researcher to use the data at their home institution. The National Center for Education Statistics allows licensed access for applications which provide a detailed description of the usage of the data, a security plan, and legal agreements that include non-disclosure (Seastrom 2001).

Virtual Enclaves

Although virtual enclaves are technologically becoming increasingly feasible and are recognized as a potentially useful innovation (VanWey et al 2005) there are not currently implemented by statistical agencies as part of restricted access program. In theory, a virtual enclave could save researchers the cost of travel to a RDC while allowing the data owning agency to maintain control over how the data are used.

Intermediate Agents

Intermediate agents are a layer of human or software agents who do the actual processing of research analysis rather than the researcher. To use National Center for Health Statistics data, researchers must submit SAS code which are vetted for restricted commands. The results of these programs are sent back to the researchers via email. To aid the researchers in debugging their programs they are provided with a synthetic data file which mimics some of the basic properties of the research data (de Wolf 2003). Synthetic Data is discussed in more detail under Restricted Data below.

Software Agents for controlling access to research data are not in widespread use. One example discussed by Cromley et al (2004) is the Connecticut Department of Public Health's online system which allows users to perform detailed queries of disaggregate source data and export results as tables or maps. The system is a server side application with limited GIS functionality for which record linking to other databases is not allowed and the user is prevented from seeing any individual level data. Output is restricted to the town level, and cells with less than five results are suppressed.

Karr (2006) discusses more advanced software based solutions using distributed data sources and server based applications which allow many users to share data and analyze it without any of them having access to source data from other users. However the algorithms for performing statistical analyses on distributed data structures are still under development and not yet ready for widespread use.

The applicability of such approaches to MEASURE may be limited because of the burden it would place on those using our data. Given the international audience MEASURE data reaches, requiring people to travel to one location to access the data would place a tremendous burden on many people.

Restricted Data

Limited Data

Suppression of data rows or attributes is perhaps the most straightforward way to protect confidentiality. Personal identifiers such as name, address etc are always removed from a dataset made available for research. In addition, cell values which can identify an individual (i.e. unusual height) may also be suppressed.

Top or bottom coding entails a cut-off point well away from the most extreme results so that outliers are part of a broader group for which either an average other predetermined value is provided. Lane (2005) is critical of top-coding because it can create data artifacts of its own, as she has shown in a studies of income inequality statistics in the US.

Suppressing spatial data without providing a substitute such as aggregate data negates the value in collecting spatial data. The applicability of this approach for MEASURE data is limited.

Geographical Masks

Geographical masks specifically address the unique confidentiality issue of spatial data: spatial data can not only be a unique identifier in the same way as a name, but spatial data can also be an integral part of the research question. Geographical masks are an attempt to walk the fine line between supporting research and preserving confidentiality. Armstrong et al (1999) is the best source for detailed descriptions of geographical masks as well as an evaluation of them.

“An optimal spatial mask is one that ensures that for every location in the feasible geographic space, at least one record might have originated there, and that for every health record there are approximately the same number of other individuals who might be associated with it. An optimal displacement is the minimum displacement that protects confidentiality and that simultaneously permits “valid” statistical analyses to take place.” (Armstrong et al 1999)

Three major types of masks are affine transformations, aggregations, and random perturbations. Affine transformations include rotations, scale transformations and displacements which (to some extent) preserve the relational attributes of the data. Due to the preservation of relational information, there is an increased disclosure risk associated with these masks.

Aggregation masks can either enumerate all data points with a region (essentially raster data) or use points with a region to represent some specified subset of datapoints. Microaggregation means that coordinates for a datapoint are replaced by an average of some number of datapoints similar to them in a manner specified by the researcher. Blurring results in datapoints being placed on an irregular lattice, and is achieved by taking an ordered list of x and y coordinates and replacing each successive group of 10 with its average.

A problem with aggregation is invariably the choice of unit to which aggregate. Health and demographic data are already often released at different aggregation levels, neither of which may be the optimal level for the researcher's analysis (see O'Dwyer 1998). Reconciling data at different aggregation levels necessarily introduces inaccuracies.

The core problem with aggregation of spatial data is that the level of detail required to observe the clusters or patterns of interest (if they exist) is not known in advance. Worse, the location of spatial boundaries of aggregation units can either disguise or artificially introduce patterns. Fundamentally, cell size for aggregation must be small enough to allow clusters to be apparent while sufficiently protecting individual information. For instance, HIV prevalence data reported at the community level across a district will show patterns that will be missing if the same data is aggregated to the district level.

Random perturbations are distinct from affine transformations in that each datapoint is transformed individually, thereby losing any pairwise relationships that might exist. Random perturbations typically make clusters more difficult to find, through clusters that are found are typically present in the original data (i.e. increased risk of type II error). Stinchcomb (2004) discusses precise algorithms for achieving a random dispersal as well as guidance on how to adapt the distribution to respect relevant contextual information, such as a nearby lake or national park that would be an ineligible location. Armstrong et al (1999) generally find that random perturbation is superior to affine transformation for most analytical purposes.

As with aggregation, the cell size within which to perturb data is a critical decision. Kwan et al (2004) found that the US Census blocks were optimal for their study, while Leitner and Curtis (2006) preferred a smaller cell size of between 200m x 200m and 350m x 350m. Clearly this optimal mask size is conditional upon the specifics of the human and physical environment of the area of study. There is some trade-off between a greater cell size for masking and the maintenance of spatial integrity. Too great a cell size and the datapoints are overly smoothed. Too small a cell size and they will be too accurate.

Geographical masks are a potential set of approaches MEASURE could employ. With careful consideration of the methods and techniques employed and the spatial accuracy needs of data users, it is possible that for some data a compromise could be found that would permit sensitive spatial data collected by MEASURE to be released.

The Role of Institutional Review Boards

IRBs are a key player in managing the tradeoff between data access and protecting privacy in the context of human subjects research. As oversight bodies, IRBs are charged with reviewing the ethical consequences (among others) of research proposals put forward by members of that institution for the purpose of ensuring the protection and welfare of the research subjects.

Two concrete measures for strengthening IRBs address their effective capacity. One recommendation is for the IRB of the institution holding the data to review research studies using that data rather than the researcher's local institution, so as to bring the greatest possible relevant expertise to bear (NRC 2000). A common problem with IRBs is the issue of member turnover. To build an institutional memory they should maintain full documentation of data usage (NRC 2000). de Wolf (2003) advocates the establishment of dedicated cross-disciplinary Disclosure Review Boards as per the US Census, but a dedicated DRB is beyond the resources of many institutions.

IRBs are hindered by the lack of a consistent legal framework to indicate appropriate confidentiality standards and impose sanctions. While the absence of such a framework makes the work of IRBs more difficult, legal measures typically struggle to keep pace with technological change anyway (NRC 2000) making it likely that grey areas are somewhat unavoidable.

IRB members must ultimately exercise their judgment[2] in coming to a determination on the beneficience of a study, and they must do so on a case-by-case basis. Lane's (2005) risk-utility framework for optimizing the use of micro-data is useful here as a conceptual tool, but its practical applications are hindered by the lack of meaningful inputs for marginal costs and benefits of data use. More statistically oriented approaches to disclosure limitation (see Feinberg 2005) propose to take on the optimization of risk quantitatively, although in doing so they assume that meaningful inputs on such difficult issues as the appropriate level of risk to accept (see VanWey et al (2005) for a discussion on such risk for spatial data) will be provided.

[2] Hyman (2000) and NRC (2000) both specifically refer to human judgment as an unavoidable component of the IRB process. The exercise of judgment is downplayed in Lane (2005) and Feinberg (2005) whose approaches are more oriented towards decision rules in a quantitative optimization approach.

US Regulatory Framework

The regulatory framework for data access and confidentiality varies from country to country, and within the US it varies in terms of policies and procedures from agency to agency. Golden et al (2005) discuss the major legal foundations for data confidentiality which include HIPAA (1996) and Title V of the E-Government Act (2002). HIPAA supposedly gives patients more control over their personal information, but in practice it is difficult for individuals to enforce their rights (Duncan 2003). Title V of the E-Government Act (also known as CIPSEA) is described as providing strong protection to privacy and confidentiality (Golden et al 2005), directing federal agencies to provide notice to the public before data collection that the data could be used for non-statistical purposes. In addition it directs agencies to take steps towards protecting confidentiality through training, record keeping of data usage, and documentation of compliance. de Wolf (2003) highlights a challenge to this privacy protection framework via OMB circular A-110 which potentially makes research data collected through federally funded grants vulnerable to freedom of information requests. Also see Curry (1997) for a discussion of how US Courts have ruled on the right to privacy in the context of viewing technological change as essentially neutral and autonomous.

The Patriot Act also has a direct impact on questions of privacy and the release of data (Duncan 2003). The release of certain information (e.g. maps of watersheds and nuclear power plants) has been restricted while law enforcement agencies have increased access to some private data (such as internet usage data). It has also had an effect on public perceptions about privacy violations (Lane 2005).

Onsrud et al (1994) is perhaps the seminal piece which draws out the implications of privacy in broader societal terms. Onsrud et al consider privacy to be one of the fundamental underpinnings of democracy and point to the great distinction in regulation of government data versus private data as a major problem. Data gathered and used by private companies is largely unregulated. Goss (1995) takes up the same theme, pointing to a lack of consumer awareness of the extent of data collected on them, the transfer and sale of that data to institutions unconnected with the original source. Curry (1997) suggests that individuals should own their digital representations and be entitled to control and edit them. More generally, as digital representations of individuals become more pervasive and permanent features of our society, Curry argues that the individuals concerned should be have more participation in the creation and use of their digital representations. Confidentiality then can be recast as a question of power, as expressed in NRC (2000): "In a pluralistic society, the translation of a moral right [i.e. privacy] into enforceable policy is a political problem."

Conclusion

The issues surrounding spatial data and confidentiality are complex and potentially controversial. Compounding matters is that technology advances often outpaces discussions and consensus on confidentiality. Work conducted for MEASURE often deals with sensitive data and project activities should consider confidentiality concerns carefully. Determining how best to balance the risk of disclosure while preserving spatial integrity is challenging, but the following principles presented by VanWey can be useful guidelines:

1) Confidentiality should be protected and informed consent promises should be honored.
2) Linking the social and spatial realms can be vital to fully understanding human behavior
3) Sharing data is an important concept that is necessary in order to make data useful

Appendix I

Matrix of Approaches

Approach	Pro	Con
Random Shifting	Simple. Provides moderate level of protection.	Eliminates ability to do distance calculations. If shift isn't sufficiently random or shift radius isn't large enough it is possible to backwards engineer
Restricted access to data – data enclave	Depending on arrangements restrictions can provide control over how data is used and limit risk of it being improperly used.	Managing and administrating can be difficult. Requires additional personnel to manage and staff. Can be burdensome for some to travel to a site
Intermediate agents	Provides control over analysis. Minimizes risk of improper use	Limiting for analysis. Requires additional staffing and management. May be unsatisfactory for some types of analysis.
Suppression of data	Strategic suppression of data ensures sensitive data is protected.	If geographic data is suppressed, not possible to map
Aggregation	Can effectively mask spatial data	Finer patterns lost.
Random Perturbations	If done at a scale sufficient to mask location, it can be easy and effective	True patterns can be lost or masked. Can be backwards engineered by adding other layers, especially if not conducted at a sufficient scale
Synthetic data	No real data is included	Can be difficult to model. Not suitable for all data. If model is incorrect, results may not be correct

Appendix II

Bibliography

Title: HIV Information Confidentiality and Security Workshop, 15th - 17th May, Geneva 2006
Short Title: HIV Information Confidentiality and Security Workshop, 15th - 17th May, Geneva 2006

Author: Abbot, Jo; Chambers, Roberts; Dunn, Christine; Harris, Trevor; De Merode, Emmanuel; Porter, Gina; Townsend, Janet; Weiner, Daniel
Title: Participatory GIS: Opportunity or Oxymoron?
Journal: PLA Notes
Volume: 33
Date: October
Short Title: Participatory GIS: Opportunity or Oxymoron?

Author: Abowd, John ; Lane, Julia
Year: 2003
Title: Economic Consequences of Data Confidentiality
Conference Name: Confidential Data Access for Research Purposes Workshop
URL: http://www7.nationalacademies.org/cnstat/Data_Access_Panel.html

Author: Allison, Maria; Harpham, Trudy
Year: 2002
Title: Southern African perspectives on the geography of health
Journal: Health and Place
Volume: 8
Pages: 223-225

Author: Appelbaum, Paul
Year: 2000
Title: Protecting Privacy while Facilitating Research
Journal: American Journal of Psychiatry
Volume: 157
Issue: 11
Pages: 1725-1726
Date: November

Author: Armstrong, Marc P.
Year: 2002
Title: Global Information Technologies and their Potentially Erosive Effects on Personal Privacy
Journal: Studies in the Social Sciences
Volume: 27
Issue: 1
Pages: 19-28

Author: Armstrong, Marc P.; Ruggles, Amy J.
Year: 2005
Title: Geographic Information Technologies and Personal Privacy
Journal: Cartographica
Volume: 40
Issue: 4
Pages: 63-73

Author: Armstrong, Marc P.; Rushton, Gerard; Zimmerman, Dale L.
Year: 1999
Title: Geographically Masking Health Data to Preserve Confidentiality
Journal: Statistics in Medicine 18
Volume: 18
Pages: 497-525
Short Title: Geographically Masking Health Data to Preserve Confidentiality

Author: Berman, Francine; Brady, Henry
Year: 2005
Title: Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences
Institution: National Science Foundation
Date: May 12

Author: Boulos, Maged N. Kamel
Year: 2004
Title: Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom
Journal: International Journal of Health Geographics
Volume: 3
Issue: 1

Author: Chrisman, Nicholas
Year: 2005
Title: Full Circle: More than Just Social Implications of GIS
Journal: Cartographica
Volume: 40
Issue: 4
Pages: 23-35

Author: Committee, Federal Geographic Data
Year: 1998
Title: FGDC Policy on Access to Public Information and the Protection of Personal Information Privacy in Federal Geospatial Databases
Department: Committee, Federal Geographic Data
Notes: Adopted by the Federal Geographic Data Committee in April 1998.

Author: Control, Center for Disease
Year: 2005
Title: CDC-CSTE Intergovernmental Data Release Guidelines Working Group (DRGWG) Report: CDC-ATSDR Data Release Guidelines and Procedures for Re-release of State-Provided Data

Author: Council, National Research
Year: 2000
Title: Improving Access to and Confidentiality of Research Data: Report of a Workshop
Series Editor: Committee on National Statistics, Christopher Mackie and Norman Bradburn, Eds.; Education., Commission on Behavioral and Social Sciences and
City: Washington, DC
Institution: National Academy Press
URL: http://www.nap.edu/books/0309071801/html/

Author: Council, National Research
Year: 2003
Title: Protecting Participants and Facilitating Social and Behavioral Sciences Research
Series Editor: Panel on Institutional Research Boards, Surveys, and Social Science Research; Constance F. Citro, Daniel R. Ilgen, and Cora B. Marrett eds.; Committee on National Statistics and Board on Behavioral, Cognitive, and Sensory Sciences and Education
City: Washington DC
Publisher: The National Academies Press
URL: http://www.nap.edu/catalog/10638.html#toc

Author: Cromley, Ellen; Cromley, Robert; Ye, Yanlin
Year: 2004
Title: Online Reporting and Mapping of Spatially Aggregated Individual Records Selected by User Queries
Journal: Cartographica
Volume: 39
Issue: 2
Pages: 5-13

Author: Croner, Charles
Year: 2003
Title: Public health, GIS and the Internet
Journal: Annual Review of Public Health
Volume: 24
Pages: 57-82

Author: Curry, Michael
Year: 1997
Title: The Digital Individual and the Private Realm
Journal: Annals, Association of American Geographers
Volume: 87
Issue: 4
Pages: 681-99

Author: Davenhall, William F.
Year: 2002
Title: Protecting Geocoded Health Data
URL: http://www.esri.com/industries/health/business/protecting_geocoded.htm

Author: de Wolf, Virginia
Year: 2003
Title: Issues in Accessing and Sharing Confidential Survey and Social Science Data
Journal: Data Science Journal
Volume: 2
Pages: 66-74
Date: 17 February

Author: Dobson, Jerry
Year: 1998
Title: Is GIS a Privacy Threat?
Magazine: Geoworld
URL: http://www.geoplace.com/gw/1998/0798/798onln.asp

Author: Duncan, George
Year: 2004
Title: Exploring the Tension Between Privacy and the Social Benefits of Governmental Databases
Conference Name: Security, Technology, and Privacy: Shaping a 21st Century Public Information
Policy
Conference Location: Georgetown University Law Center, Washington, DC.
Date: April 24-25, 2003

Author: Duncan, George
Year: 2006
Title: Statistical Confidentiality: Is Synthetic Data the Answer?
Conference Name: Lecture to the Institute for Digital Research and Education
Conference Location: University of California at Los Angeles
Date: February 13

Author: Duncan, George; Jabine, Thomas; de Wolf, Virginia
Year: 1993
Title: Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics
City: Washington, D.C.
Publisher: National Academy Press

Author: Dunn, Christine; Atkins, Peter; Blakemore, Michael; Townsend, Janet
Year: 1999
Title: Teaching Geographical Information Handling Skills for Lower-income Countries
Journal: Transactions in GIS
Volume: 3
Issue: 4
Pages: 319-332

Author: Dunn, Christine; Atkins, Peter; Townsend, Janet
Year: 1997
Title: GIS for Development: a contradiction in terms?
Journal: Area
Volume: 29
Issue: 2
Pages: 151-159

Author: Dunn, Christine; Williams, Craig
Year: 2003
Title: GIS in Participatory Research: Assessing the Impact of Landmines on Communities in North-west Cambodia
Journal: Transactions in GIS
Volume: 7
Issue: 3
Date: June

Author: Elwood, Sarah
Year: 2006
Title: Negotiating Knowledge Production: The Everyday Inclusions, Exclusions, and Contradictions of Participatory GIS Research
Journal: The Professional Geographer
Volume: 58
Issue: 2
Pages: 197-208
Date: May

Author: Fefferman, Nina H; O'Neil, Eileen A; Naumova, Elena N
Year: 2005
Title: Confidentiality and Confidence: Is Data Aggregation a Means to Achieve Both?
Journal: Journal of Public Health Policy
Volume: 26
Issue: 4
Pages: 430-449

Author: Fienberg, S. E.; Slavokvich, A.
Year: 2005
Title: Preserving the Confidentiality of Categorical Statistical Data Bases When Releasing Information for Association Rules
Journal: Data Mining and Knowledge Discovery
Volume: 1
Pages: 155-180

Author: Foley, Ronan
Year: 2002
Title: Assessing the applicability of GIS in a health and social care setting: planning services for informal careers in East Sussex, England
Journal: Social Science and Medicine
Volume: 55
Pages: 79-96

Author: Golden, M. L.; Downs, R. R.; Davis-Packard, K.
Year: 2005
Title: Confidentiality Issues and Policies Related to the Utilization and Dissemination of Geospatial Data for Public Health Applications.
Notes: A Report to: The Public Health Applications of Earth Science Program, National Aeronautics and Space Administration, Science Mission Directorate, Applied Sciences Program. Socioeconomic Data and Applications Center (SEDAC), Center for International Earth Science Information Network (CIESIN), Columbia University.
Link to PDF: http://www.ciesin.columbia.edu/pdf/SEDAC_ConfidentialityReport.pdf

Author: Goss, Jon
Year: 1995
Title: 'We Know Who You Are and We Know Where You Live': The Instrumental Rationality of Geodemographic Systems
Journal: Economic Geography
Volume: 71
Issue: 2
Pages: 171-198

Author: Hyman, Steven
Year: 2000
Title: The Needs for Database Research and for Privacy Collide
Journal: American Journal of Psychiatry
Volume: 157
Issue: 11
Pages: 1723-1724
Date: November

Reference Type: Journal Article
Record Number: 62
Author: Karr, A. F. ; Kohnen, C. N.; Oganian, A. ; Reiter, J. P. ; Sanil, A. P.
Year: 2006
Title: A Framework for Evaluating the Utility of Data Altered to Protect Confidentiality
Journal: The American Statistician
Volume: forthcoming

Author: Koontz, Linda
Year: 2003
Title: Geographic Information Systems Challenges to Effective Data Sharing
Department: Office, US General Accounting
Report Number: U.S. GAO Report 03-874T
Notes: Testimony Before the Subcommittee on Technology, Information Policy, Intergovernmental Relations and the Census, Committee on Government Reform, U.S. House of Representatives.

Author: Kwan, Mei-Po; Casas, Irene; Schmitz, Ben
Year: 2004
Title: Protection of Geoprivacy and Accuracy of Spatial Information: How Effective Are Geographical Masks?
Journal: Cartographica
Volume: 39
Issue: 2
Pages: 15-28

Author: Kwan, Mei-Po; Schuurman, Nadine
Year: 2004
Title: Introduction: Issues of Privacy Protection and Analysis of Public Health Data
Journal: Cartographica
Volume: 39
Issue: 2
Pages: 1-4

Author: Kyem, Peter A. Kwaku
Year: 2000
Title: Embedding GIS Applications into Resource Management and Planning Activities of Local and Indigenous Communities
Journal: Journal of Planning Education and Research
Volume: 20
Issue: 2
Pages: 176-186

Author: Lachow, Irving
Year: 1995
Title: The GPS Dilemma: Balancing Military Risks and Economic Benefit
Journal: International Security
Volume: 20
Issue: 1
Pages: 126-148
Date: Summer

Author: Lane, Julia
Year: 2003
Title: Key Issues in Confidentiality Research: Results of an NSF Workshop
Institution: National Science Foundation

Author: Lane, Julia
Year: 2005
Title: Optimizing the Use of Micro-data: An Overview of the Issues
Conference Name: 2005 American Statistical Association Meeting
Conference Location: Minneapolis, MN

Author: Laymon, Paul
Year: 1999
Title: Spatially Enabling Vital Health Data
Magazine: ArcUser
Short Title: Spatially Enabling Vital Health Data
URL: http://www.esri.com/news/arcuser/0499/vital.html

Author: Leitner, M.; Curtis, A.
Year: 2006
Title: A first step towards a framework for presenting the location of confidential point data on maps - results of an empirical perceptual study
Journal: International Journal of Geographical Information Science
Volume: 20
Issue: 7
Pages: 813-822
Date: August

Author: Martin, Eugene
Year: 2000
Title: Actor-networks and implementation: examples from conservation GIS in Ecuador
Journal: International Journal of Geographical Information Science
Volume: 14
Issue: 8
Pages: 715 - 738

Author: McLafferty, Sara
Year: 2004
Title: The Socialization of GIS
Journal: Cartographica
Volume: 39
Issue: 2
Pages: 51-53

Author: Meador, Mary; Ruggles, Amy J.
Year: 2000
Title: Steps involved in randomizing the coordinates of address-matched locations
Magazine: Public Health GIS News and Information
Volume: 7
Issue Number: 35
Pages: 11-14

Author: Medicine, Institute of
Year: 1994
Title: Health Data in the Information Age: Use, Disclosure, and Privacy
Series Editor: Molla S. Donaldson and Kathleen N. Lohr, Eds.; Services, Committee on Regional Health Data Networks. Division of Health Care
City: Washington, D.C.
Publisher: NATIONAL ACADEMY PRESS
URL: http://darwin.nap.edu/books/0309049954/html/R1.html

Author: Myles, G ; Friday, A; Davies, N
Year: 2003
Title: Preserving Privacy in Environments with Location Based Applications
Journal: Pervasive Computing
Pages: 56-64
Date: January-March

Author: O'Dwyer, Lisel; Burton, Deborah
Year: 1998
Title: Potential meets reality: GIS and public health research in Australia
Journal: Australian and New Zealand Journal of Public Health
Volume: 22
Issue: 7
Pages: 819-823

Author: Ohno-Machado, L; Silveira, PS; Vinterbo, S
Year: 2004
Title: Protecting patient privacy by quantifiable control of disclosures in disseminated databases.
Journal: International Journal of Medical Informatics
Volume: 73
Issue: 7-8
Pages: 599-606

Author: Onsrud, H.J.
Year: 1995
Title: Identifying Unethical Conduct in the Use of GIS
Journal: Cartography and Geographic Information Systems
Volume: 22
Issue: 1
Pages: 90-97

Author: Onsrud, H.J.
Year of Conference: 1997
Title: Ethical Issues in the Use and Development of GIS
Conference Name: GIS/LIS ‘97 Conference

Author: Onsrud, H.J.; Johnson, Jeff; Lopez, Xavier
Year: 1994
Title: Protecting Personal Privacy in Using Geographic Information Systems
Journal: Photogrammetric Engineering and Remote Sensing
Volume: LX(9)
Pages: 1083-1095

Author: Oppong, Joseph
Year of Conference: 1999
Title: Data Problems in GIS and Health
Conference Name: Proceedings of Health and Environment Workshop 4: Health Research Methods and Data
Conference Location: Turku, Finland
Date: 22–25 July

Author: Ratcliffe, Jerry
Year: 2002
Title: Damned if you don't, Damned if you do: Crime Mapping and its Implications in the Real World
Journal: Policing and Society
Volume: 12
Issue: 3
Pages: 211-225

Author: Richards, Thomas; Croner, Charles; Rushton, Gerard; Brown, Carol; Fowler, Littleton
Year: 1999
Title: Geographic Information Systems and Public Health: Mapping the Future
Magazine: U.S. Department of Health and Human Services Public Health Reports
Volume: 114
Pages: 359-373

Author: Schuurman, Nadine
Year: 2000
Title: Trouble in the Heartland: GIS and its critics in the 1990s
Journal: Progress in Human Geography
Volume: 24
Issue: 4
Pages: 569-590

Author: Seastrom, Marilyn
Year: 2001
Title: Licensing
Editor: Doyle, P.; Lane, Julia; Theeuwes, J.J.M. ; Zayatz, L.V.
Book Title: Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies
City: Amsterdam, Netherlands
Publisher: Elsevier Science
Pages: 279-296

Author: Simon, Gregory; Unuetzer, Juergen; Young, Barbara; Pincus, Harold
Year: 2000
Title: Large Medical Databases, Population-Based Research, and Patient Confidentiality
Journal: American Journal of Psychiatry
Volume: 157
Issue: 11
Pages: 1731-1737
Date: November

Author: Statistics, President's Commission on Federal
Year: 1971
Title: Federal Statistics: Report of the President's Commission on Federal Statistics
City: Washington, D.C.

Author: Steel, P.; Sperling, J.
Year: 2001
Title: The Impact of Multiple Geographies and Geographic Detail on Disclosure Risk: Interactions between Census Tract and ZIP Code Tabulation Geography
Department: Bureau, U.S. Census
Publisher: (unpublished monograph)

Author: Stinchcomb, David
Year of Conference: 2004
Title: Procedures for Geomasking to Protect Patient Confidentiality
Conference Name: ESRI International Health GIS Conference
Date: October 19

Author: Tanser, F.C.; le Sueur, D.
Year: 2002
Title: The application of geographical information systems to important public health problems in Africa
Journal: International Journal of Health Geographics
Volume: 1

Author: Ting, D.; Fienberg, S. E. ; Trottini, M.
Year: 2005
Title: ROMM methodology for microdata release
Conference Name: UNECE/Eurostat Work Session on Statistical Data Confidentiality,
Conference Location: Geneva, Switzerland
Date: November 9-11

Author: van Loenen, Bastiaan; Onsrud, Harlan
Year: 2004
Title: Geographic Data for Academic Research: Assessing Access Policies
Journal: Cartography and Geographic Information Science
Volume: 31
Issue: 1
Pages: 3-17

Author: VanWey, L. K.; Rindfuss, R. R.; Gutmann, M. P.; Entwisle, B.; Balk, D. L.
Year: 2005
Title: Confidentiality and spatially explicit data: Concerns and challenges
Journal: Proceedings of the National Academy of Sciences
Volume: 102
Issue: 43
Pages: 15337-15342

Reference Type: Conference Proceedings
Record Number: 37
Author: Weiner, Daniel; Harris, Trevor
Year of Conference: 1999
Title: Community-Integrated GIS for Land Reform in South Africa
Conference Name: GISOC'99
Conference Location: The University of Minnesota
Minneapolis, MN.
Date: June 20-22

Author: Williams, Craig; Dunn, Christine
Year: 2003
Title: GIS in Participatory Research: Assessing the Impact of Landmines on Communities in North-West Cambodia
Journal: Transactions in GIS
Volume: 7
Issue: 3
Pages: 393-410