MEASURE Evaluation

April 2017

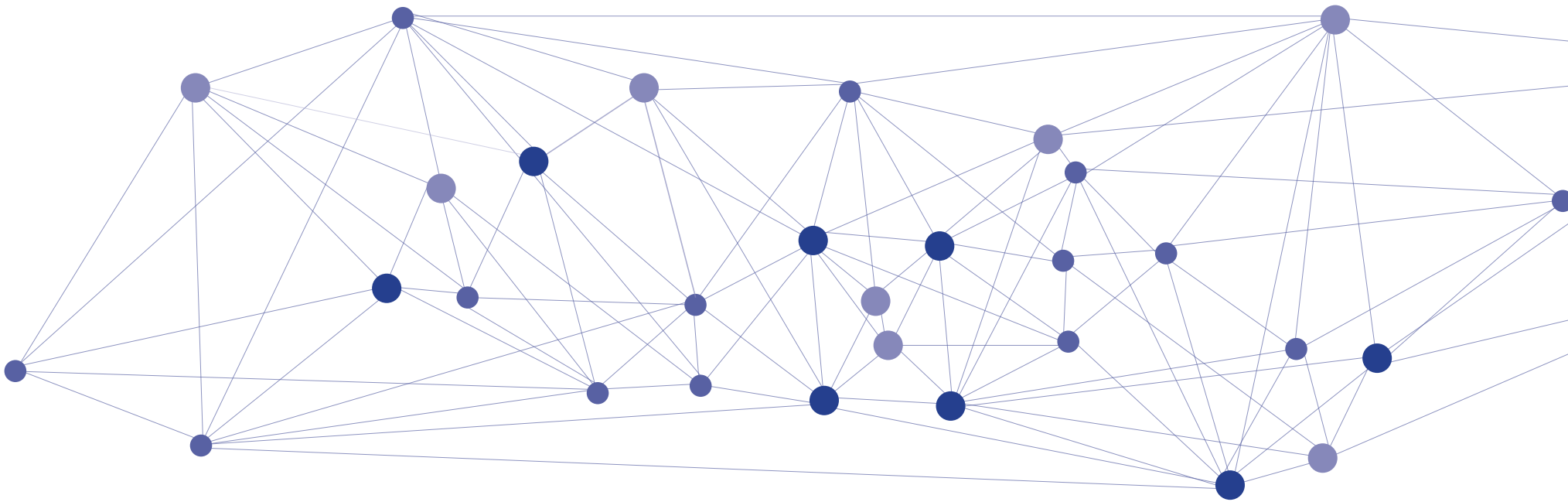
Frequently Asked Questions about Geographic Information Systems

Using Spatial Data Wisely and Ethically

Privacy and Confidentiality

Preface

This is one in a series of FAQs on important topics that are relevant to geographic information systems (GIS) and spatial data. They are intended to provide brief answers to common questions and steer you to sources of more detailed information.



1 People's lives are closely tied to their geography.

Knowing where someone lives, works, or travels can provide insight into where they might get healthcare or what their risk might be for disease.

In short, knowing **where** things are happening can help us understand **why** things are happening.

While spatial data provides great insight, it also comes with great responsibility. Spatial data can serve as a de facto identifier of individuals. Because of this, users of spatial data must consider privacy and confidentiality implications.

What do we mean by "spatial data"?

"Spatial data" refers to anything that includes a geographic identifier. Some examples are:

- Geographic coordinates such as latitude and longitude
- An address
- The name of a district or any other administrative unit

In other words, anything that might link to a location on Earth.

What is the difference between privacy **and** confidentiality?

These terms are often used interchangeably, but they refer to two different concepts.

Confidentiality concerns protecting the identity of individuals.

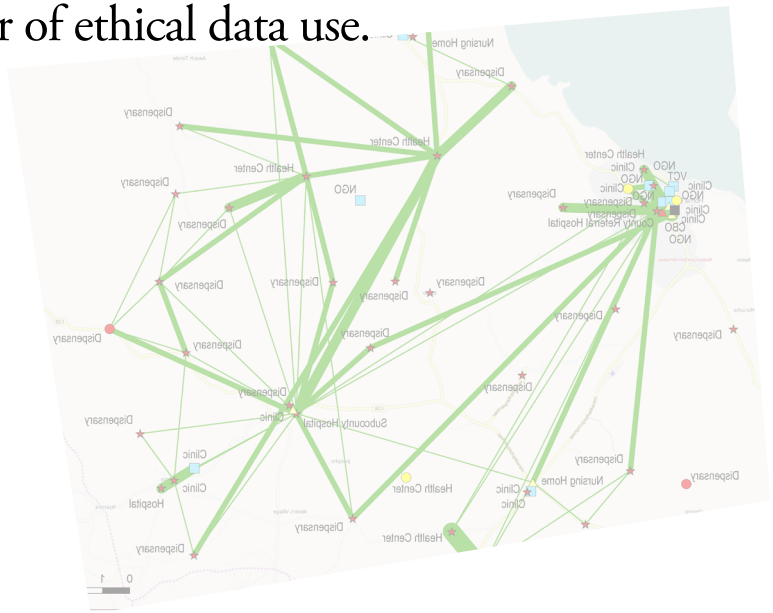
Privacy concerns what people are willing to share or reveal about themselves.

Privacy relates to **people**; confidentiality relates to **data**.

It is the responsibility of those collecting, analyzing, and displaying spatial data to consider confidentiality and privacy in their work. The promises made to people about confidentiality and privacy as part of the informed consent process should be honored.

Spatial data has unique characteristics that can have an impact on privacy and confidentiality. Displaying or sharing data that includes a geographic coordinate can make it possible to identify individuals.

These truths have special bearing on health data, where the stakes are high. People could face harm as a result of disclosures of their health data. Responsible use of this type of data is a fundamental pillar of ethical data use.



IV What risks exist with spatial data?

There are two principle risks: **overt disclosure** and **deductive disclosure**.

Overt disclosure is an actual release of data that breaches confidentiality commitments: for instance, if data is stored on an unsecured server or if someone loses their laptop.

Deductive disclosure is when multiple data elements are combined together in a way that leads to breaches in confidentiality commitments.

Spatial data can greatly reduce the number of data elements needed to make deductive disclosure possible. Releasing highly precise spatial data can change an act of deductive disclosure to an act of overt disclosure.

Spatial data make **overt disclosure** a *possibility* and **deductive disclosure** *easier.*

Confidentiality breach scenario 1:

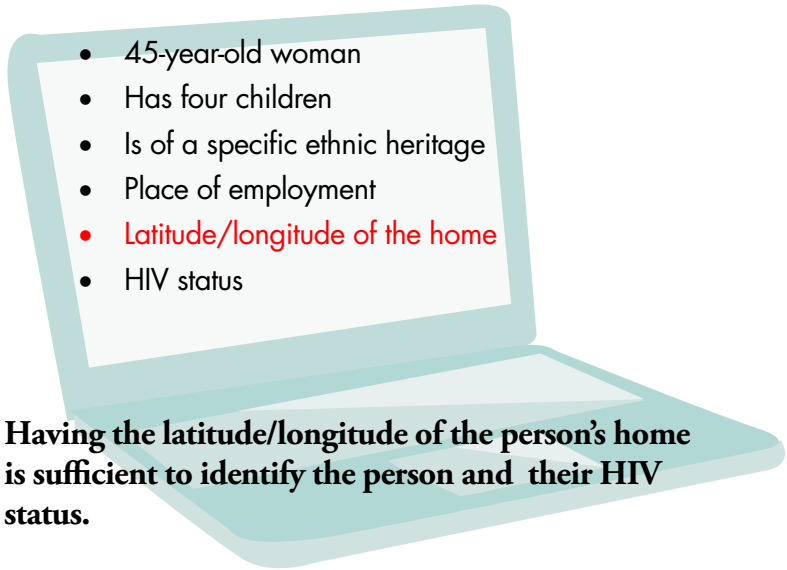
Data from a household HIV survey are stored on a laptop that was stolen. The laptop was not password-protected, so the data is accessible to the person who has the laptop.

Data on characteristics of an individual

Knowing the four characteristics in red could make it possible to identify an individual and then know that person's HIV status.

Confidentiality breach scenario 2, involving spatial data; same stolen laptop:

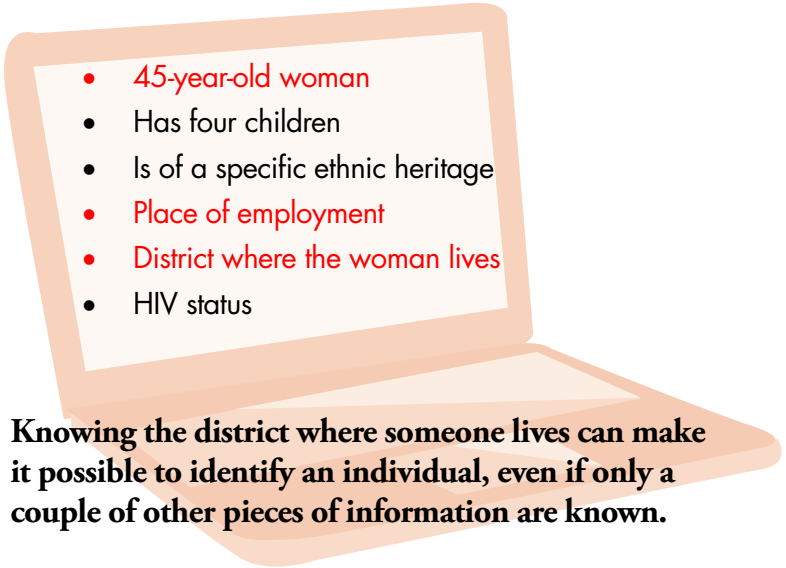
Data on characteristics of an individual



Having the latitude/longitude of the person's home is sufficient to identify the person and their HIV status.

Confidentiality breach scenario 3, also involving spatial data; same stolen laptop:

Data on characteristics of an individual



Knowing the district where someone lives can make it possible to identify an individual, even if only a couple of other pieces of information are known.

V

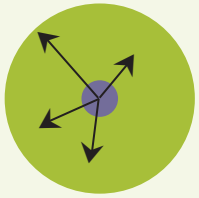
What can be done to protect privacy and keep spatial data confidential?

Several different methods can be employed to protect confidentiality; there is no one-size-fits-all solution. Every approach entails some compromise between minimizing risk of disclosure and spatial accuracy.

Below are some common approaches that people use with spatial data to minimize the risk of disclosing people's identities.

Common ways to keep spatial data confidential

Random Perturbations

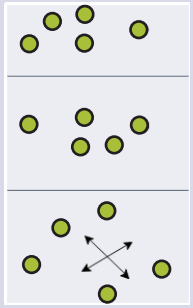


Random shifting of points

Pros: Easy to do

Cons: Lose original location; introduces spatial error

Transformations

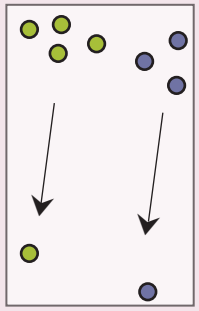


Change scale, rotate, shift a set distance, or some combination

Pros: Easy to do

Cons: Easy to undo; can impact some types of analysis

Aggregate

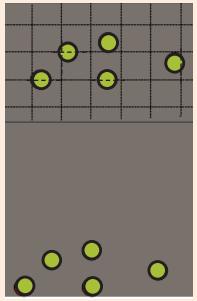


Point locations are aggregated to a higher unit of analysis

Pros: Easy to do

Cons: Requires sufficient data points; finer data variations will be lost

Despatialize

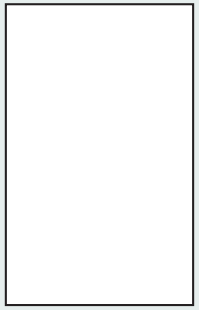


Remove coordinate system from data and use Euclidean space

Pros: Simple; keeps relative position and placement

Cons: Loses contextual data

Nothing

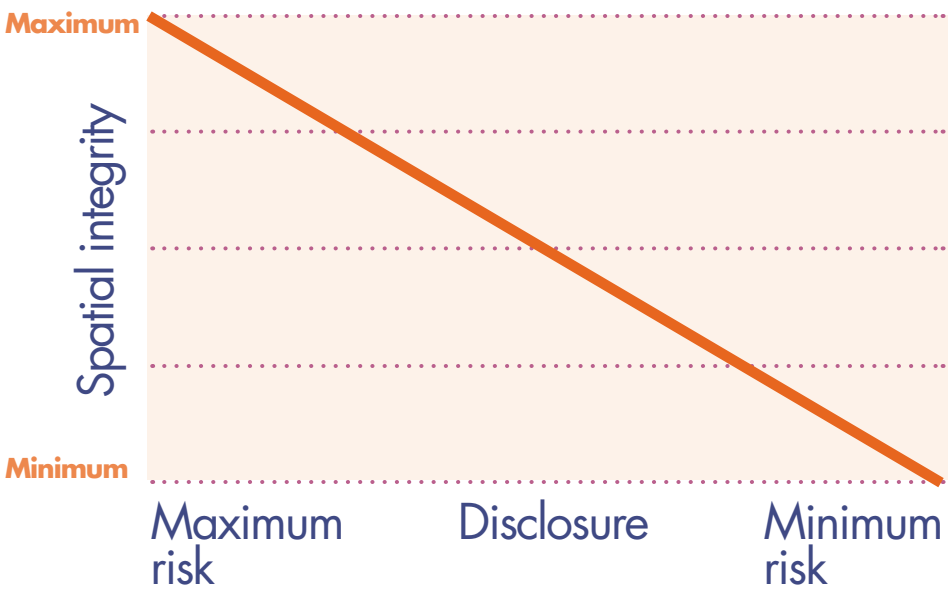


Do not collect or release spatial data. Make data available only on-site in controlled conditions.

Pros: Maintains all of the original spatial data (if collected)

Cons: Complicated; limits data sharing; limits social spatial link

Every approach will represent a **compromise** on either spatial data accuracy or disclosure risk.



V continued

What are the responsibilities of data collectors and data users?

Data collectors and data users have different courses of action before them.

For data collectors:

- If collecting data from individuals, make sure the informed consent agreement is clear on what spatial data is being collected and how it will be used and—just as important—how it *won't* be used.
- Make sure that proper data security measures are in place that minimize the risk of confidentiality breaches. Some data collection teams have protocols that mandate storing spatial data separate from other data.
- Have a clear policy on what spatial data will be released and to whom and the dissemination process.
- Develop data use agreements for users that describe proper and improper uses of the data.

V continued

What are the responsibilities of data collectors and data users?

For data users:

- Differentiate between spatial data products that are used internally and those that are included in publications or presentations. Make sure that those presented do not breach confidentiality requirements or make deductive disclosure easier.
- Produce maps that mask potential individual identities, either by employing one of the approaches described above or by displaying maps that are at a scale coarse enough to protect individual identities

The way data is mapped can protect an individual's privacy. Changing the size of the symbology or adjusting the scale can mask the exact location.

Map showing the locations of hypothetical respondents. Enough error is introduced by the size of the dots and the scale of the map that it would be difficult to know precisely where individuals live.

With a finer scale, respondent locations are more identifiable.

VI Conclusion

Issues around confidentiality and privacy are complex, but they must be considered in work with spatial data. No spatial data should be collected without a clear policy about the confidentiality considerations of the data. No map or other spatial data product should be published or presented without considering whether it overtly discloses sensitive data or makes deductive disclosure possible.

Unfortunately, minimizing the risks of improper disclosure of data can be challenging. No magic button exists on a GIS that can be clicked to solve privacy and confidentiality issues. Different solutions may be required in different situations.

A full examination of the issues and solutions is not possible in this short space. Review the literature, consult with other GIS professionals, or take courses to understand the issues better. Refer to the resources below.

Issues of **confidentiality** and **privacy** *cannot be ignored* in work with spatial data.

VII References

Burgert, C., Colston, J., Roy, T., & Zachary, B. (2013). *Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys.* DHS Spatial Analysis Reports No. *7*. Calverton, Maryland, USA: ICF International.

Haley, D., Matthews, S., Cooper, H., Haardörfer, R., Adimora, A., Wingood, G., & Kramer, M. (2016). Confidentiality considerations for use of social-spatial data on the social determinants of health: Sexual and reproductive health case study. *Social Science and Medicine, 166*, 49–56.

Kwan, M.-P., Casas, I., & Schmitz, B. (2004). Protection of geoprivacy and accuracy of spatial information: How effective are geographical masks? *Cartographica, 39*(2), 15–28.

MEASURE Evaluation. (2008). *Overview of Issues concerning confidentiality and spatial data.* MEASURE Evaluation.

National Research Council. (2007). *Putting people on the map: Protecting confidentiality with linked social-spatial data.* Washington, DC, USA: National Academies Press.

Perez-Heydrich, C., Warren, J., Burgert, C., & Emch, M. (2013). *Guidelines on the use of DHS GPS data.* Calverton, Maryland, USA: ICF International.

Richardson, D., Kwan, M.-P., Alter, G., & McKendry, J. (2015). Replication of scientific research: Addressing geoprivacy, confidentiality and data sharing challenges in geospatial research. *Annals of GIS, 21*(2).

VanWey, L., Rindfuss, R., Gutmann, M., Entwisle, B., & Balk, D. (n.d.). Confidentiality and spatially explicit data: Concerns and challenges. *PNAS, 102*(43).

Zandbergen, P. (2014). Ensuring confidentiality of geocoded health data: Assessing geographic masking strategies for individual-level data. *Advances in Medicine*.

MEASURE Evaluation
University of North Carolina at Chapel Hill
400 Meadowmont Village Circle, 3rd Floor
Chapel Hill, North Carolina 27517
Phone: +1 919-445-9350 | Fax: +1 919-445-9353
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. SR-17-143