Ethics and Best Practices for Mapping Archaeological Sites

Cecilia Smith ⓘ

ABSTRACT

Archaeologists are tasked with balancing a call to open data and the need to maintain confidentiality of sensitive archaeological site locations. Low-resolution mapping and data aggregation are the methods most commonly used to hide site locations; however, we understand little of the effectiveness of these practices. Trends in geomasking, obscuring observed geographic points, to anonymize public health data are suggested as a source of methods for sharing archaeological site data. Archaeologists have available to them a number of geomasking methods that balance open data and site security in different ways. Low-resolution mapping at several scales and random direction with fixed radius, random perturbation donut, and Gaussian donut techniques are tested on a set of archaeological site locations. Random perturbation donuts resulted in the best balance between obscuring archaeological locations and conveying observed spatial patterning. Researchers should carefully consider how they convey archaeological location data, as commonly used low-resolution scales may not provide the desired level of obscurity. Researchers should also be explicit as to how and why their methods of site visualization are chosen.

Keywords: geomask, geospatial data, GIS, site location, visualization, mapping

Los arqueólogos tienen la tarea de equilibrar un llamamiento a las prácticas de datos abiertos y de mantener la confidencialidad de sitios arqueológicos sensibles. La cartografía de baja resolución y la agregación de datos son los métodos más utilizados para ocultar los lugares de los sitios; sin embargo, entendemos poco de la eficacia de estas prácticas. Se sugieren tendencias en el enmascaramiento de la ubicación, el ocultamiento de puntos geográficos observados, para anonimizar los datos de salud pública como fuente de métodos para compartir los datos de los sitios arqueológicos. Los arqueólogos tienen a su disposición una serie de métodos de enmascaramiento de la ubicación que equilibran los datos abiertos y la seguridad del sitio de diferentes maneras. En un conjunto de emplazamientos de sitios arqueológicos se ensayan técnicas de cartografía de baja resolución a varias escalas, dirección aleatoria con radio fijo, rosquillas de perturbación aleatoria y de rosquilla gaussiana. Las rosquillas de perturbación aleatoria dieron como resultado el mejor equilibrio entre el ocultamiento de los sitios arqueológicos y la transmisión de los patrones espaciales observados. Los investigadores deben considerar cuidadosamente cómo transmiten los datos de los emplazamientos arqueológicos, ya que las escalas de baja resolución comúnmente utilizadas podrían no proporcionar el nivel de ocultamiento deseado. Los investigadores también deben ser explícitos en cuanto a cómo y por qué se escogen sus métodos de visualización de los sitios.

Palabras clave: enmascaramiento de la ubicación, datos geoespaciales, SIG, ubicación del sitio, visualización, cartografía

CURRENT LOCATION-SHARING PRACTICES AND STANDARDS

Archaeologists are responsible to a range of governments and institutions for the ethical collection, maintenance, access, and archiving of the data they produce. For example, in the United States, federally funded grant applications require data management plans. At the same time, the push for open access (OA) to scholarship is also securing a foothold in the field (Costa et al. 2013; Huggett 2014; Strupler and Wilkinson 2017). True OA refers to information that is available without restriction on the internet. Evidence of OA in archaeology comes from the rise of journals such as the *Journal of Open Archaeology Data* (https://open-archaeologydata.metajnl.com/); repositories such as the Digital Index of North American Archaeology (DINAA; http://ux.open-context.org/archaeology-site-data/) and the Digital Archaeological Record (https://www.tdar.org/about/), which provide data with locational restrictions; and interest groups such as the Society for American Archaeology's Open Science in Archaeology (https://osf.io/2dfhz/).

A balance must be struck in meeting the security and legal requirements of sensitive archaeological information while effectively communicating spatial patterns. If site security is a concern, this balance is usually created by archaeologists in two

ways: mapping sites at low resolution or aggregating sites to grids or administrative boundaries. Often, maps of site locations are not accompanied by an explicit description of security concerns. This article introduces the methods available for obscuring archaeological site locations, the efficacy of those methods, and a recommendation for archaeologists to be mindful and explicit when selecting visualization methods.

Sensitive archaeological sites are defined by the U.S. Department of the Interior as those at risk for damage if their locations are disclosed (NPS 1997). Archaeologists must then determine whether a sensitive site has enough cultural significance or potential for future scientific discovery that its location should be obscured to protect it. It is important to note that current evidence points to local knowledge networks, rather than academic data, as being the source for terrestrial site locations for looters (Proulx 2013; Smith 2005). Although there is no evidence of a relationship between publication or OA repositories and terrestrial looting, archaeologists are ethically bound, and often also legally bound, to respect the cultural importance of a site by obscuring or not disclosing its coordinates.

The balance between ensuring the security of sensitive sites and sharing site location information is often implicit. However, Sarah Parcak's GlobalXplorer° web platform is one project that provides a high-profile example of archaeologists explicitly addressing visualizations available to the public and site security. The FAQ section of https://www.globalxplorer.org/ states that site maps will not be made publicly available unless deemed appropriate by the "governments and protection agencies involved." In addition, the images of the earth's surface that are presented to citizen scientists are not "linked to coordinates or other data that would expose the location of a site" (GlobalXplorer° 2019). The random presentation of these tiled images also prevents systematic use by potential looters.

Another example is DINAA, a repository for North American archaeological data. DINAA does not share actual site locations with the public but rather aggregates site counts. Its website indicates that the data are "publicly viewable as KML data with two forms of representation, at the level of US county and in a ~20km grid" (Anderson et al. 2011). The grid size was chosen as previous work provided precedent (DINAA 2013). DINAA's explicit discussion of data resolution for the purposes of obscuring data is relatively rare and emphasizes protection of the data, with the site going so far as to not store precise locational data at all. New regional trends are being discovered through grid aggregation (Anderson et al. 2017), although more localized spatial patterning is obscured. McCoy (2017) also explicitly incorporates the aggregation method when discussing his work with Geospatial Big Data. This article explores how other methods help communicate localized spatial patterning while continuing to protect locational data.

There are existing laws and recommendations in regard to the sharing of archaeological site locations; but in many contexts there is a great deal of autonomy in how to visually represent sites and no available archaeological literature on available techniques for obscuring sites while still communicating archaeological findings. Section 304 of the 2016 National Historic Preservation Act states that information should be withheld from the public if "disclosure could result in a significant invasion of privacy,

damage to the historic property, or impede the use of a traditional religious site by practitioners." This includes

> street addresses, highway and route numbers, Universal Transverse Mercator (UTM) or Geographic Information System (GIS) coordinates, electronic maps, and descriptions, including photographs and drawings, of the property's position in relation to local landmarks or natural features such that it could be found [Advisory Council on Historic Preservation 2016].

The Advisory Council on Historic Preservation also indicates that "if information is already in the public realm but with very limited accessibility, it does not mean that it can no longer be protected from further disclosure" (2016). This restriction applies only to properties that are listed in the National Register of Historic Places as determined by the keeper of the National Register.

The Cultural Resource Geographic Information System Facility Heritage Documentation Programs of the National Park Service present the complexity of maintaining geospatial data at a national scale in the Draft Set of Standards for Cultural Resource Spatial Data (NPS 2019). The NPS maintains geospatial data for "3000 cultural landscapes, 27,000 historic buildings and structures, 1,200 Ethnographic resources, 63,000 archeological sites, and over 500 American battlefields." Of concern is that the draft declares that "there are no standards for cultural resource spatial data" (NPS 2019). Attempting to collate or share information across administrative and agency boundaries without standards is modestly described as difficult. To begin to alleviate this issue, the NPS established the *Cultural Resource Spatial Data Transfer Standards: Guidelines for Use and Implementation* (NPS 2014). The standards include increased granularity in regard to restricting the distribution of cultural data, while noting that all archaeological data are restricted from being released as they fall under the 1979 Archaeological Resources Protection Act (ARPA; NPS 2014:19).

Section 9 of ARPA also prohibits public disclosure of information concerning the nature and location of archaeological resources on federal or Indian land that require a permit or other permission under the act for their excavation or removal. Disclosure is allowed as long as there is no risk of "harm to the archaeological resource" (ARPA, Uniform Regulations, Sec. 7.18). Beyond the designation of sensitive data via Section 304 of the National Historic Preservation Act or Section 9 of ARPA, state historic preservation offices vary in their policies for distributing archaeological site data. For example, the Nevada State Historic Preservation Office maintains the Nevada Cultural Resources Information System (https://shpo.nv.gov/services/nvcris), which provides restricted and unrestricted access levels to archaeological data. Researchers meeting the Secretary of the Interior's Standards for Archeological Documentation may request a subscription to restricted data via e-mail.

In 2018 the Geospatial Data Act was signed into federal law. The act formalized many of the groups and processes that govern spatial data at the national level, including the governance of the National Spatial Data Infrastructure (NSDI). The NSDI includes "the technology, policies, criteria, standards, and employees necessary to promote geospatial data sharing throughout the Federal Government, State, tribal, and local governments, and the private sector (including nonprofit organizations and institutions of higher education)" (Geospatial Data Act of 2018, Sec. 752),

established in 1994 by Executive Order 12906. The Congressional Research Service report on the act writes that the first goal of the NSDI is "to ensure that geospatial data from multiple sources . . . are available and easily integrated to enhance the understanding of the physical and cultural world" while guaranteeing protection of private and secure data (2018:6). Practical and appropriate workflows for communicating sensitive archaeological site locations are not yet available.

While this review focuses on laws and policies in the United States, there is a wide range of standards employed internationally. All archaeological sites are covered by the Society for American Archaeology's Principle 6, which encourages "taking into account" preservation and protecting sites when disclosing site locations. This article discusses methods for mapping archaeological sites that take into account these concerns and provides descriptions and assessments of geomasking methods to allow researchers to select appropriate geomasking techniques and parameters regardless of where they conduct research. This article calls on findings from the field of public health to explore the degree to which methods obscure actual site locations.

VULNERABILITY OF LOCATION-SHARING PRACTICES

Geographic information systems (GIS) are powerful technologies for collecting, managing, analyzing, and visualizing spatial information. Geospatial data include any data with a locational component. The precision and accuracy of geospatial products depend on researchers' decision making and the equipment used. As introduced above, archaeologists must balance the security of an archaeological site according to its importance and vulnerability, while meeting expectations for responsible, open science. This article does not provide techniques to assess a site's sensitivity but explores the available options for obscuring site locations and how those options balance security and scholarship.

A number of public health publications expose security flaws in common location-obscuring practices in visualizing spatial data. Brownstein, Cassa, and Mandl (2006) published a correspondence piece in the *New England Journal of Medicine* in which they called for guidelines for representing patients' homes to preserve anonymity. They suggested that the common practice of using lower-resolution maps was less effective than aggregating patients to administrative units (e.g., census tracts) or their preferred method of randomly changing a patient's location within a fixed distance. Over the next decade public health specialists conducted reverse geocoding experiments to test geomasking methods (Allshouse et al. 2010; Boulos et al. 2009; Brownstein, Cassa et al. 2006; Hampton et al. 2010; Seidl et al. 2017; Zandbergen 2014). The process of finding an address from a coordinate pair is referred to as reverse geocoding. Geomasking is a process of obscuring the location of real-world coordinates. After geomasking a set of coordinates, researchers found the probability that a patient's true location could be identified from available patient data, the distance between the observed and offset coordinates, and the population density of the area. Their work reveals a number of geomasking techniques that balance security and open science (sharing unobscured location data) in different ways.

Instead of patient or participant anonymity, archaeologists concerned about site security should consider how difficult it would be to identify a site from a geomasked map and how rigorous an attempt should be made to obscure sites. Only two geomasking techniques are routinely employed by archaeologists, low-resolution mapping and aggregation, with only one or two others having occasional representations in the literature. The following section describes geomasking methods that can be applied to archaeological sites. Each balances security and open data differently.

METHODS FOR VISUALIZING SENSITIVE LOCATION DATA

Aggregation shows the number of sites within a given boundary, with the most common boundaries being administrative units (e.g., counties; Figure 1) and grids (Figure 2). When purposefully and explicitly obscuring archaeological sites, this is the most common method employed. This method can show distributions over a large area but does not allow for spatial patterning at local levels.

Low-resolution maps indicate the actual observed location, with the assumption that the low resolution (e.g., a scale of 1:5,000,000) will make it difficult to find the real-world location (Figure 3).

Heat maps are representations of the density of observations; for example, the density of archaeological sites (Figure 4). Densities are calculated using a neighborhood, or kernel, the size of which is defined by the user.

Bounding boxes are hollow squares surrounding all the observations of a given area but do not depict the site locations (Figure 5).

Coordinate patterns without base maps are used to accurately convey the spatial relationships of observations but without the context of topographic or administrative features (Figure 6).

Random direction with a fixed radius allows the user to define a particular distance from which the geomasked point is shown away from the observed coordinate (Figure 7). The direction in which the geomasked point is placed is randomly selected via algorithm.

Random perturbation within a fixed radius allows a maximum distance from which a geomasked point may be offset from an observed point (Figure 8). The geomasked point will be placed in a random direction away from the observed point and at a random distance within the maximum distance specified by the user.

Random perturbation donuts are similar to random perturbation within a fixed radius but also include a minimum distance from which the geomasked points must be placed from the observed point (Figure 9). Overall, the donut method is generally the most effective at masking locations in public health studies while still providing a visual representation of spatial patterning.

Gaussian displacement resembles random perturbation within a fixed radius, with the exception that geomasked points are distributed in a normal distribution away from the observed point,
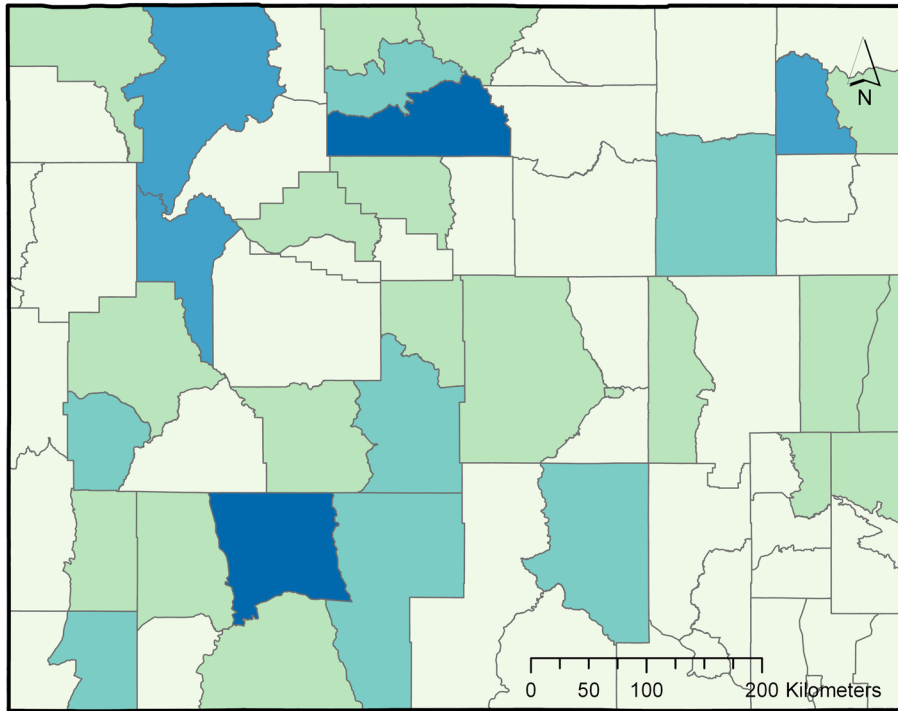
FIGURE 1. Sites aggregated to county boundaries. Darker color represents more sites. (State boundary data from Esri and TomTom North America 2019.)
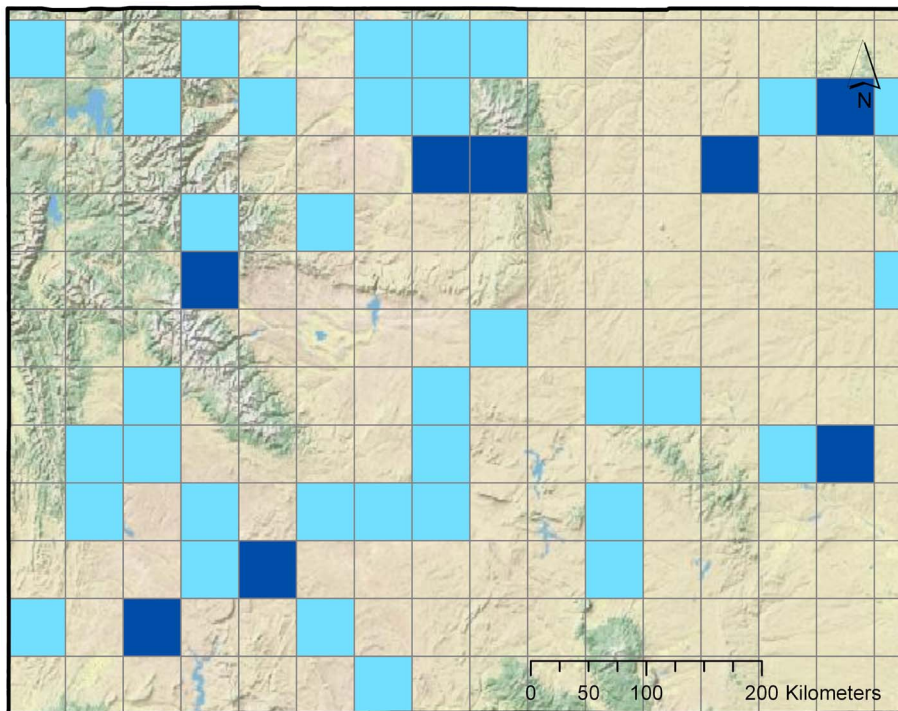

FIGURE 2. Archaeological sites aggregated by 700 km^2 grid cells. Dark blue cells represent the presence of two sites, and light blue represents one site. (State boundary data from Esri and TomTom North America 2019.)
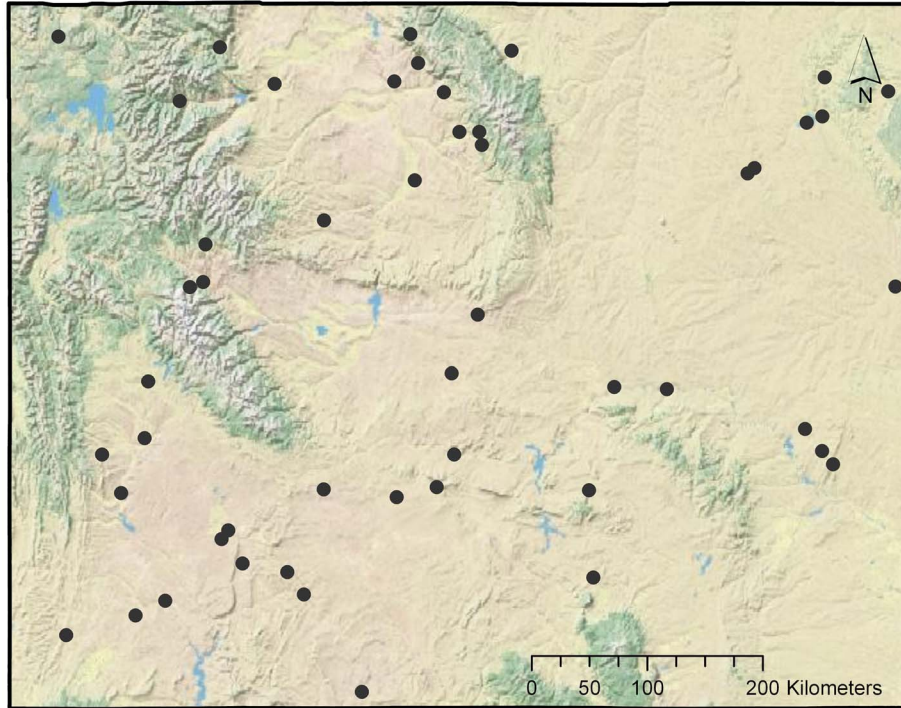
FIGURE 3. Archaeological sites mapped at the low resolution of 1:5,000,000. (Site data from University of Wyoming Department of Geography et al. 2017; state boundary data from Esri and TomTom North America 2019; topographic data from National Geographic Society and i-cubed 2019.)
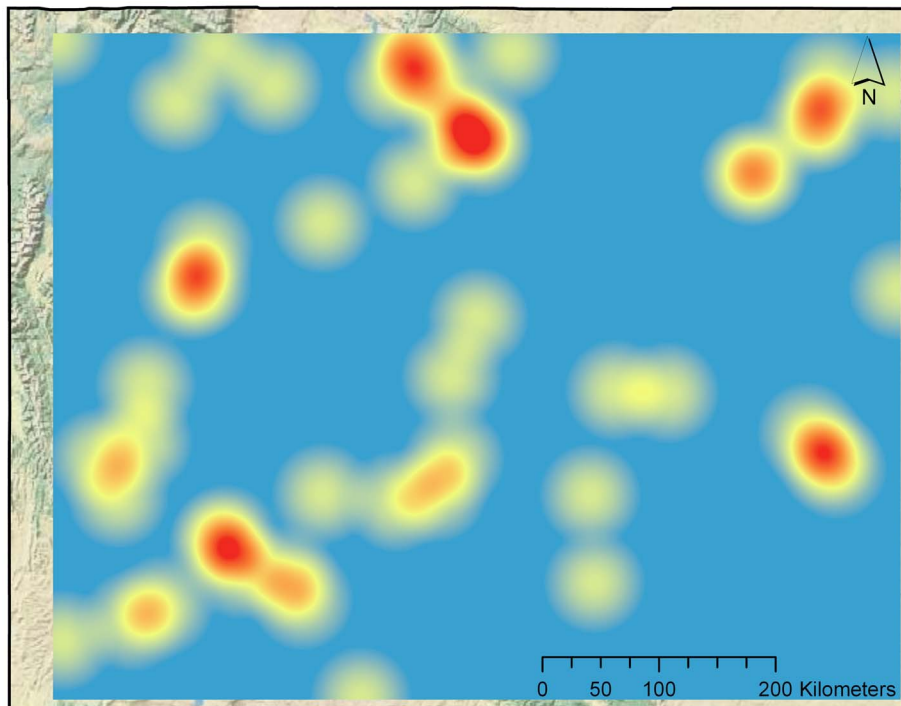

FIGURE 4. Archaeological sites depicted by heat map produced with a kernel size of 50 km. (State boundary data from Esri and TomTom North America 2019; topographic data from National Geographic Society and i-cubed 2019.)
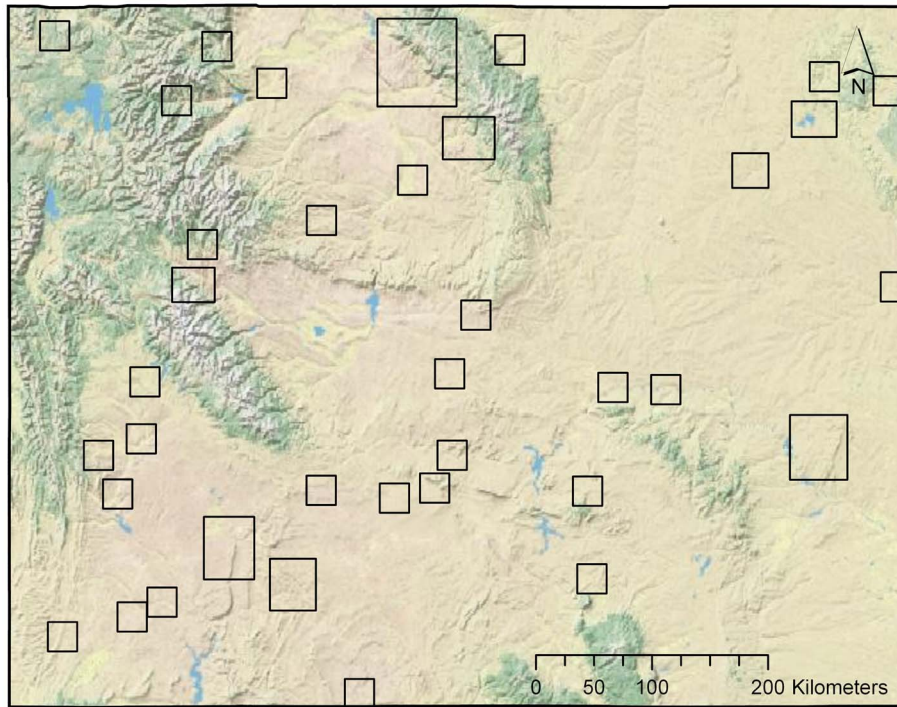
FIGURE 5. Bounding boxes drawn around individual and clusters of sites. (State boundary data from Esri and TomTom North America 2019; topographic data from National Geographic Society and i-cubed 2019.)
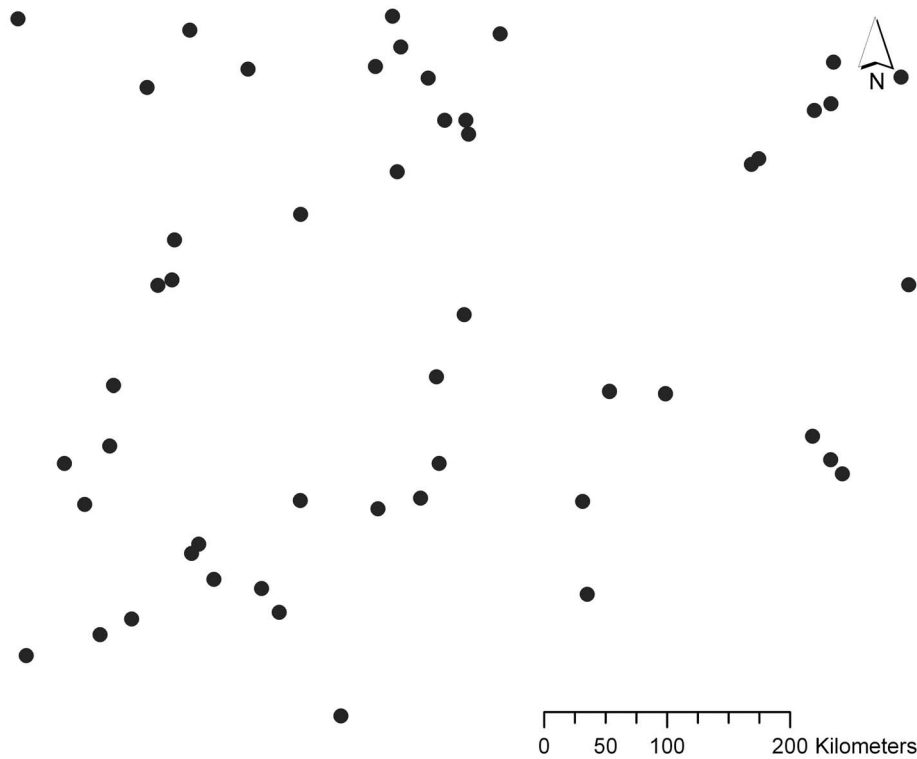

FIGURE 6. Locations of archaeological sites with the topographic base map removed. (Site data from University of Wyoming Department of Geography et al. 2017.)
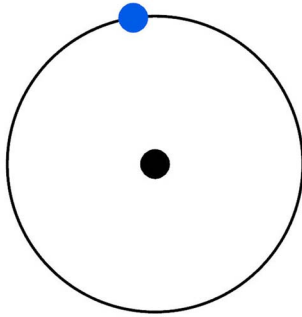
FIGURE 7. Random direction fixed radius. The geomasked point (blue) is placed in a random direction 1 km away from the observed point (black). The black ring represents possible locations for a geomasked point.
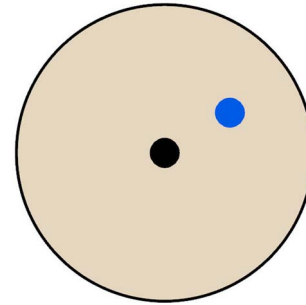

FIGURE 8. Random perturbation. The geomasked point (blue) is placed in a random direction at a random distance within 1 km of the observed point (black). Brown represents the area in which a geomasked point could be placed.
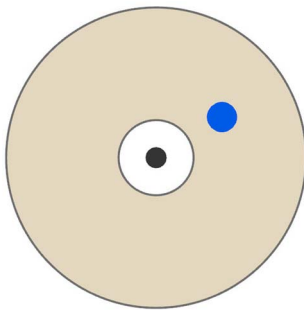

FIGURE 9. Random perturbation donut. The geomasked point (blue) is placed in a random direction in a random distance within 1 km of the observed point (black) at a minimum distance of 250 m. Brown represents the area in which a geomasked point could be placed.

with the maximum distance being set by a contextual variable (Figure 10). For example, in public health, the greater the population density (the contextual variable), the smaller the maximum distance from an observed point.

Gaussian donuts, also referred to as *bimodal Gaussian displacement*, are similar to Gaussian displacement but also set a minimum threshold at which the geomasked points can be placed from the observed points (Figure 11). Gaussian donuts are also effective in masking true locations while showing spatial patterning: the added dimension of adjusting the maximum displacement range allows for smaller offsets in densely populated areas.

Programming models, such as linear programming and digit switching, involve using code to systematically obscure real-world coordinates. Linear programming considers each observation and applies models that include the probability of successful reverse geocoding, constraints, and objectives. Digit switching changes two or more digits of coordinates in the Military Grid Reference System. Another code is used as a key to replace the original digits. By specifying the digits to be switched, the user has control over the maximum distances to which coordinates will be

geomasked. Illustrations of linear programming can be found in Wieland and colleagues (2008). Illustrations of digit switching can be found in Clarke (2016). Another example of applying models to randomly perturb coordinate locations is described in Fronterrè and colleagues (2018).

The *Voronoi* or *Thiessen method* identifies the center of polygonal areas, usually land parcels in public health studies, and creates Voronoi, also known as Thiessen, polygons. Voronoi polygons are the result of lines drawn through the midpoints between observed points. Geomasked points for each observation are placed on the closest part of the closest Voronoi edge. For parcel data, this method generally avoids placing a geomasked point near a polygon centroid. Illustrations of this method can be found in Croft and colleagues (2016).

Location swapping identifies neighborhoods with similar geographic characteristics to a point being geomasked. The original location of that point is then swapped with a location from one of those similar neighborhoods. Variations of this method, such as *location swapping with donut*, are explored in Zhang and colleagues (2017).

Scholarship and security are balanced differently in each of these approaches. First, skewed heavily in the direction of security are aggregation, heat maps, and bounding boxes. These techniques make it improbable that the actual sites will be identified, at the cost of entirely obscuring spatial patterning at the local level, for other archaeologists and the public. Also skewed heavily in favor of security are observed site locations published without a base map. Without topographical clues, identifying the actual sites would be very difficult. However, there is the risk that once one site is identified, the other site locations could be identified using the published spatial relationships.

Versions of donut masking strike a more calculated balance between scholarship and security. By placing a geomasked point at some distance from an observed point, researchers decrease the size of the area that would need to be searched to locate an archaeological site. Studies in public health show that the risk of reverse geocoding an observed location is much smaller with donut methods than with low-resolution mapping—a commonly deployed technique in archaeology and, prior to the
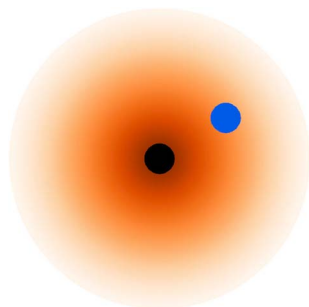
FIGURE 10. Gaussian displacement. A geomasked point (blue) is offset from the observed point (black) at a distance determined by at least one contextual variable, such as density of sites. The orange field presents the area in which a geomasked point may be placed, while the intensity of the orange indicates the likelihood that a geomasked point will be placed in that specific location.

FIGURE 11. Gaussian donuts. A geomasked point (blue) is offset at a distance determined by at least one contextual variable, such as density of sites, at a minimum distance of 250 m. The orange field presents the area in which a geomasked location may be placed, while the intensity of the orange indicates the likelihood that a geomasked point will be placed in that specific location.

last decade, the most commonly deployed in public health. Donut masking, some programming models, and Voronoi methods have also been explored for how well they maintain spatial relationships.

Comparison of geomasking techniques in archaeological contexts is necessary because the data types and the objectives of archaeologists differ from those of public health researchers. The concern in public health is that personal data could be connected back to an individual person. Making this sort of connection would usually involve using parcel data from a city or county to match identifying characteristics such as age, ethnicity, and sex to health data. The archaeological concern is to protect site locations and adhere to related laws and policies. In this regard, the archaeologist is interested in balancing the obscurity of the site (the size of the area that would require survey to identify a site), open science (providing access to real-world coordinates to repeat analysis and interpretation), and facilitating visualization in published works (geomasking coordinates to mimic observed spatial or geographic patterns).

METHODS FOR COMPARING GEOMASKING TECHNIQUES FOR ARCHAEOLOGICAL SITES

A comparison of geomasking techniques was conducted using 50 archaeological sites in Wyoming whose coordinates are publicly available (University of Wyoming Department of Geography et al. 2017). Techniques that do not result in a masked point pattern—for example, those resulting in a masked area (e.g., heat maps and aggregation)—were not included in the comparison. In addition, Voronoi methods and programming methods were not included. Voronoi methods require an underlying polygon layer from which to create centroids, which is more conducive to studying modern residential associations (e.g., parcel data) than archaeological sites. Programming methods may be tailored to archaeology, but no work has been completed in this area. Existing location-switching programs are also more conducive to modern
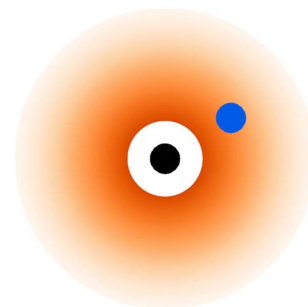
residential data where placing a geomasked point on an existing residence outside the study area is good practice, whereas placing a geomasked point on a known archaeological site outside the study area would not be good practice.

Low-resolution maps, random direction buffers with fixed radius, random perturbation donuts, and Gaussian donuts were applied to the 50 archaeological sites with varying parameters to produce 50 geomasked points for each type. Random perturbation and Gaussian methods without minimum offset "donuts" were not included, because the methods allow geomasked points to fall in close proximity or on top of observed sites. Low-resolution maps were tested by creating static maps of observed site locations at the scales of 1:100,000, 1:5,000,000, and 1:20,000,000 in Google Maps. These maps were brought into ArcGIS Pro, a commercial GIS platform that is in wide use in cultural resource management. Once in ArcGIS Pro, the maps were georeferenced, and then new points were placed on the observed points.

To create geomasked points using random direction fixed radius and random perturbation donuts, buffers were created at a maximum distance of 1,000 m from observed sites and random points were assigned either to the circumference of the buffer for fixed radius types or between the circumference and a minimal offset distance of 250 m. This was done in ArcGIS Pro but could easily be done in an open-source software such as QGIS. Random perturbation without a minimum distance was excluded as the geomasked points would occur in close proximity to the observed site.

Gaussian displacement donuts were created using the v.perturb GRASS command in QGIS (GRASS GIS 2020). This command allows users to specify the mean and standard deviation of a normal distribution of random points within a given area. In this case a minimum distance of 250 m was selected to prevent points from occurring in close proximity to an observed site. Standard deviations of 500 m and 750 m were tested, as these distributions approximate the 1,000 m maximum distance used in the fixed distance and random perturbation methods.

Once each geomasked location set was created using the above methods, three measures were recorded: the distance between
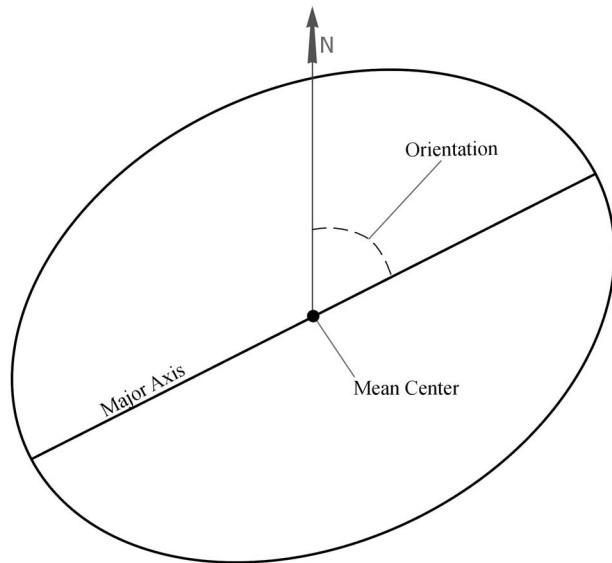
FIGURE 12. Example of a standard deviational ellipse and mean center, which are calculated using the x- and y-coordinates of a set of locations.

each observed and geomasked point, the minimum resulting area that would need to be searched to find the observed site based on the geomask method applied, and the Global Divergence Index (Gdi). The Gdi is one measure of how far displaced geomasked points are from the observed site locations (Kounadi and Leitner 2014; Seidl et al. 2017). The smaller the Gdi, the smaller the displacement. The index is calculated by first finding the mean centers and the standard deviational ellipses of observed and geomasked datasets, as proposed by Kounadi and Leitner (2014). The mean center is found by calculating the average x- and y-coordinates for a set of points. A standard deviational ellipse is created by using the standard deviations of the x- and y-coordinates from the mean center as the major and minor axes of the ellipse. Figure 12 depicts the parts of a standard deviational ellipse used in the calculation. Next, the divergences between the mean centers, orientation of the ellipses, and major axes of the ellipses are calculated. The three divergence values are then averaged to produce the Gdi:

$$\overline{GDi} = (Mdi, \ Odi, \ MAdi)$$

Mdi is the divergence of the mean centers. In addition to using the distance between the original and geomasked mean centers, the distance between the observed mean center and the farthest

point in the study area is also used. The study area for this project is the state of Wyoming, with the northeast corner of the state occurring farthest from the mean center of the observed archaeological sites:

$$Mdi = \frac{\text{distance of observed mean to geomasked mean}}{\text{distance of observed mean to farthest point in the study area}} \times 100$$

Odi is the divergence between the orientations of the ellipses. The orientation is the degree of the angle between the major axis and north:

$$Odi = \frac{\text{orientation of observed ellipse} - \text{orientation of geomasked ellipse}}{180} \times 100$$

MAdi is the divergence between the lengths of the major axes of the ellipses. In addition to the lengths of the observed and geomasked ellipses, the length of the longest major ellipse axis (Maximum.MA) that could be drawn through the entire study area is required. The length of the major axis for the geomasked ellipse is compared with either the difference between the maximum major axis length and the observed length or the difference between the minimum major axis length and the observed length. Whichever difference is greater is used for the comparison. The minimum possible major axis length for the study area is so close to zero that it does not significantly impact the resulting index value and is therefore not included in the denominator of the second calculation in Figure 13.

RESULTS

Distance between Observed Sites and Geomasked Locations

Descriptive statistics summarizing the distances separating observed and geomasked points are contained in Table 1. The greater the distance, the more the observed site is obscured; however, increased displacement will also affect the spatial patterning of the resulting geomasked dataset. Perhaps the most surprising result was that the commonly used low-resolution method of using a map scale of 1:5,000,000 resulted in geomasked points that were quite close to the actual points. Random perturbation and Gaussian displacements without a minimum distance were not included, as they would also place geomasked points in close proximity to, or possibility in the same location as, the observed sites. Mean and maximum distances for Gaussian

$$MAdi = \frac{|\text{Geomasked. MA} - \text{Observed. MA}|}{\text{Maximum. MA} - \text{Observed. MA}} \times 100, \ \text{if Observed. MA} \ \leq \ \frac{\text{Maximum. MA}}{2}$$

$$MAdi = \frac{|\text{Geomasked. MA} - \text{Observed. MA}|}{\text{Observed. MA}} \times 100, \ \text{if Observed. MA} \ > \ \frac{\text{Maximum. MA}}{2}$$

FIGURE 13. The Major Axis Divergence Index (MAdi) compares the difference in orientation of major axes of one-standard deviational ellipses formed by observed and geomasked points.

TABLE 1. Summary of the Distances (m) between an Observed Point and the Geomasked Point for Each Geomasking Method.

| Measure | Low Resolution | | | Random Direction, Fixed Radius | Random Perturbation Donut | Gaussian Donut | |
	1:100,000	1:5,000,000	1:20,000,000	(Radius of 1,000 m)	(Maximum 1,000 m; Minimum 250 m)	Minimum 250 m (SD 500 m)	Minimum 250 m (SD 750 m)
Mean	23.09	602.11	3,546.43	1,000.00	688.08	670.33	914.06
Median	21.03	564.83	3,523.52	1,000.00	718.91	617.90	807.43
SD	12.28	272.78	1,470.57	0.00	214.52	305.08	448.73
Minimum	4.79	51.40	285.33	1,000.00	454.80	257.16	285.86
Maximum	48.40	1,468.25	7,943.30	1,000.00	659.94	1,509.61	2,553.99

TABLE 2. Global Divergence Index and Total Search Area Resulting from Each Geomasking Method.

| Measure | Low Resolution | | | Random Direction, Fixed Radius | Random Perturbation Donut | Gaussian Donut | |
	1:100,000	1:5,000,000	1:20,000,000	(Radius of 1,000 m)	(Maximum 1,000 m; Minimum 250 m)	Minimum 250 m (SD 500 m)	Minimum 250 m (SD 750 m)
Global Divergence Index	0.00	0.07	0.31	0.02	0.02	0.02	0.08
Search area (m^2)	–	–	–	6,283.19	2,944,000.00	1,214,000.00	2,424,000.00

donuts were relatively large. Random direction buffers with fixed radius provide the greatest control in terms of distance between observed and geomasked locations because the distance is defined by the user.

Search Area

Table 2 lists the minimum area that would need to be surveyed to identify the site after geomasking if the searcher was aware of the geomasking parameters. For example, if a person knows that a geomasked location was created using the random perturbation donut method in which the maximum distance was 1,000 m and the minimum distance was 250 m, the greatest potential area that would need to be surveyed to identify the observed site would be 2,944,000 m^2 (total area of maximum buffer minus area of minimal search area). This search area is represented in brown in Figure 9. If the parameters of the geomask are unknown to the searcher, it would be impossible to delineate a quantifiable search area that would further obscure real-world site locations. Search area was not calculated for low-resolution methods, as they do not result in a quantifiable area.

Random direction with fixed radius produced the smallest search area because if the parameters are known, survey would only be required along the circumference created by the radius of the geomask. Random perturbation and Gaussian donuts produced larger search areas. Search areas for the Gaussian donuts were approximated using the mean distance found between observed and geomasked points because the user does not set the maximum distance. As with using absolute distance between observed and geomasked points, an increased search area further

obscures archaeological sites but may result in changing the spatial pattern observed.

Global Divergence Index

Table 2 also lists the Gdi for each geomasking method. Smaller Gdi values indicate more similarity between observed and geomasked points. Larger Gdi indicates greater divergence between observed and geomasked points. Georeferencing a map at the 1:100,000 scale resulted in a near 0 Gdi, indicating very close fidelity between the location of observed sites and geomasked locations. Random direction with fixed radius, random perturbation donut, and Gaussian donut with 500 m standard deviation also performed well, with Gdi values of 0.02. The 1:5,000,000 map and the Gaussian donut with a defined standard deviation of 750 m resulted in greater spatial divergence. Finally, the map at 1:20,000,000 resulted in a large divergence index of 0.31.

DISCUSSION

The random perturbation donut method performed best in terms of balancing security and maintaining spatial pattern fidelity. This method resulted in geomasked locations being placed at a controllable distance away from observed sites, a large search area, and low divergence in spatial location between observed sites and geomasked locations. The technique is simple and can be carried out in major GIS software, such as ArcGIS and QGIS. An ArcGIS Toolbox that allows the user to customize the donut parameters accompanies this article as supplemental material. While random direction with fixed radius resulted in a small Gdi, the small search

area makes it a less effective technique compared with random perturbation donut geomasking.

It is more difficult to control for maximum distance offsets with Gaussian displacement methods because points are distributed randomly on a normal distribution according to a user-specified variable. In public health, a smaller standard deviation would be used in more densely settled areas. This works because parcel data are already publicly available and it is useful to place a geomasked residential location on top of another residence to prevent reverse geocoding. In archaeological applications the Gaussian displacement may have less utility because there is not an immediate corollary to public parcel data. If a Gaussian displacement were to be used for an archaeological application, results from the Gdi analysis suggest that the smallest possible standard deviation should be used to minimize divergence. Georeferencing a static map with the scale of 1:100,000 resulted in the lowest Gdi, indicating little change in spatial divergence between observed and geomasked data; however, the very small minimum change in distance would make this a poor choice for protecting a site's location. The great variation of distances produced by maps scaled 1:5,000,000 and 1:20,000,000 produced relatively large Gdi values. The minimum distance of 51.40 m for 1:5,000,000 suggests that some geomasked locations may be placed very close to or even on top of observed archaeological sites.

This article is an introduction to primary geomasking techniques that assessed methods for use in archaeological applications by identifying the degree to which they obscure sites and maintain spatial patterning. Random perturbation donut masking was identified as a strong balance between maintaining site security and providing access to spatial patterning. Factors such as site density, topography, and contemporary infrastructure could be taken into account when selecting a geomasking technique and its parameters. For example, larger maximum offset areas could be chosen for areas with a high density of archaeological sites, with unique topographies, or that are significantly covered with contemporary infrastructure to prevent reverse geocoding. Being explicit in these choices, particularly when true locations are shared or when locations are heavily obscured, should be a standard in the field.

Seidl and colleagues (2018) present an interesting study in which they tracked the confidence of individuals in reverse geocoding a set of points to their original households before and after they were told that the points had been geomasked. They found that "frequent notifications that the points are masked" reduced individuals' confidence, "thereby lowering identification risk" (2018). The application of explicit geomasking discussion may have similar results in archaeology. If confidence that an archaeological site is easily findable from a geomasked location is put into question, one may hope that the risk of attempts to ground truth or otherwise endanger a sensitive site would decrease.

Additional recommendations come from another field concerned with data privacy, participatory data collection. Sensor-derived data are often collected via smartphones and other devices. Kounadi and Resch (2018:Table 7) compiled a list of recommendations to prevent data from being shared inappropriately. In addition to the techniques discussed above, they also suggest the following in relation to the dissemination of anonymized datasets:

- Avoid sharing multiple versions of anonymized datasets
- Avoid sharing anonymization metadata
- Create and share a risk assessment on sharing the anonymized data

These practices will help ensure that original site locations are not re-created after geomasking and that later data users are informed that the data are geomasked and that careful consideration was made to minimize risk.

Supplementary Materials

For supplemental material accompanying this article, visit https://doi.org/10.1017/aap.2020.9.

DonutGeomask.tbx is an ArcGIS Toolbox that allows the user to geomask a set of points using the random perturbation donut method.

Acknowledgments

Data Availability Statement

Original data created for this study included four shapefiles (doi:10.6082/uchicago.1932) derived from nonsensitive archaeological sites made publicly available through the *Wyoming Student Atlas* (University of Wyoming Department of Geography et al. 2017). The datasets created for this study are curated in Knowledge@UChicago, an institutional repository at the University of Chicago, found at https://knowledge.uchicago.edu/.

REFERENCES CITED

Advisory Council on Historic Preservation
  2016 Frequently Asked Questions on Protecting Sensitive Information about Historical Properties under Section 204 of the NHPA. Electronic document, https://www.achp.gov/digital-library-section-106-landing/frequently-asked-questions-protecting-sensitive-information, accessed January 8, 2019.
Allshouse, William B., Molly K. Fitch, Kristen H. Hampton, Dionne C. Gesink, Irene A. Doherty, Peter A. Leone, Marc L. Serre, and William C. Miller
  2010 Geomasking Sensitive Health Data and Privacy Protection: An Evaluation Using an E911 Database. *Geocarto International* 25:443–452.
Anderson, David, Thaddeus G. Bissett, Stephen J. Yerka, Joshua J. Wells, Eric C. Kansa, Sarah W. Kansa, Kelsey Noack Myers, R. Carl DeMuth, and Devin A. White
  2017 Sea-Level Rise and Archaeological Site Destruction: An Example from the Southeastern United States Using DINAA (Digital Index of North American Archaeology). *PLoS ONE* 12(11):e0188142.
Anderson, David, Eric Kansa, Sarah Kansa, Stephen Yerka, and Joshua Wells
  2011 Developing the Cyberinfrastructure for a National Archaeological Site Database. Electronic document, http://ux.opencontext.org/wp-content/uploads/2012/09/DINAA-NASD-Technical-Proposal-2011.pdf, accessed August 8, 2019.
Boulos, Maged N. Kamel, Andrew J. Curtis, and Philip Abdel Malik
  2009 Musings on Privacy Issues in Health Research Involving Disaggregate Geographic Data about Individuals. *International Journal of Health Geographics* 8:article 46. DOI:10.1186/1476-072X-8-46.

Brownstein, John S., Christopher A. Cassa, Isaac S. Kohane, and Kenneth D. Mandl
  2006 An Unsupervised Classification Method for Inferring Original Case Locations from Low-Resolution Disease Maps. *International Journal of Health Geographics* 5:article 56. DOI:10.1186/1476-072X-5-56.

Brownstein John S., Christopher A. Cassa, and Kenneth D. Mandl
  2006 No Place to Hide: Reverse Identification of Patients from Published Maps. *New England Journal of Medicine* 355:1741–1742.

Clarke, Keith C.
  2016 A Multiscale Masking Method for Point Geographic Data. *International Journal of Geographical Information Science* 30:300–315.

Congressional Research Service
  2018 The Geospatial Data Act of 2018. Electronic document, https://crsreports.congress.gov/product/pdf/R/R45348, accessed April 9, 2020.

Costa, Stefano, Anthony Beck, A. H. Bevan, and Jessica Ogden
  2013 Defining and Advocating Open Data in Archaeology. In *Proceedings of the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology*, pp. 449–456. Amsterdam University Press, Amsterdam, the Netherlands.

Croft, William Lee, Wei Shi, Jörg-Rüdiger Sack, and Jean-Pierre Corriveau
  2016 Location-Based Anonymization: Comparison and Evaluation of the Voronoi-Based Aggregation System. *International Journal of Geographical Information Science* 30:2253–2275.

Digital Index of North American Archaeology
  2013 DINAA Sensitive Data Security Measures and SHPO Collaboration. Electronic document, http://ux.opencontext.org/archaeology-site-data/dinaa-sensitive-data-security-measures-and-shpo-collaboration/, accessed August 8, 2019.

Esri and TomTom North America
  2019 USA State Boundaries. Electronic document, https://www.arcgis.com/home/item.html?id=540003aa59b047d7a1f465f7b1df1950, accessed February 7, 2019.

Fronterrè, Claudio, Emanuele Giorgi, and Peter Diggle
  2018 Geostatistical Inference in the Presence of Geomasking: A Composite-Likelihood Approach. *Spatial Statistics* 28:319–330.

GlobalXplorer°
  2019 Frequently Asked Questions. Electronic document, https://www.globalxplorer.org/faq, accessed April 9, 2020.

GRASS GIS
  2020 GRASS GIS 7.8.2. GRASS Development Team. Electronic document, http://grass.osgeo.org, accessed February 5, 2020.

Hampton, Kristen H., Molly K. Fitch, William B. Allshouse, Irene A. Doherty, Dionne C. Gesink, Peter A. Leone, Marc L. Serre, and William C. Miller
  2010 Mapping Health Data: Improved Privacy Protection with Donut Method Geomasking. *American Journal of Epidemiology* 172:1062–1069.

Huggett, Jeremy
  2014 Promise and Paradox: Accessing Open Data in Archaeology. In *Proceedings of the Digital Humanities Congress 2012*, by Clare Mills, Michael Pidd, and Esther Ward. Studies in the Digital Humanities. Digital Humanities Institute, Sheffield, UK. https://www.dhi.ac.uk/openbook/chapter/dhc2012-huggett, accessed April 9, 2020.

Kounadi, Ourania, and Michael Leitner
  2014 Spatial Information Divergence: Using Global and Local Indices to Compare Geographical Masks Applied to Crime Data. *Transactions in GIS* 19:737–757.

Kounadi Ourania, and Bernd Resch
  2018 A Geoprivacy by Design Guideline for Research Campaigns That Use Participatory Sensing Data. *Journal of Empirical Research on Human Research Ethics* 13:203–222. DOI:10.1177/1556264618759877.

McCoy, Mark D.
  2017 Geospatial Big Data and Archaeology: Prospects and Problems Too Great to Ignore. *Journal of Archaeological Science* 84:74–94.

National Geographic Society and i-cubed
  2019 USA Topo Maps. Electronic document, https://www.arcgis.com/home/item.html?id=99cd5fbd98934028802b4f797c4b1732, accessed February 7, 2020.

NPS (National Park Service)
  1997 Secretary of the Interior's Standards for Archeological Documentation. In *Archeology and Historic Preservation: Secretary of the Interior's Standards and Guidelines*. Electronic document, https://www.nps.gov/history/local-law/arch_stnds_7.htm, accessed April 9. 2020.
  2014 *Cultural Resource Spatial Data Transfer Standards: Guidelines for Use and Implementation*. Cultural Resource GIS Facility, Preservation Assistance Programs. Electronic document, https://irma.nps.gov/Datastore/DownloadFile/489140, accessed February 5, 2020.
  2019 Draft Set of Standards for Cultural Resource Spatial Data. Cultural Resource Geographic Information System Facility, Heritage Documentation Programs. Electronic document, https://www.nps.gov/hdp/standards/crgisstandards.htm, accessed February 5, 2020.

Proulx, Blythe Bowman
  2013 Archaeological Site Looting in "Glocal" Perspective: Nature, Scope, and Frequency. *American Journal of Archaeology* 117:111–125.

Seidl, Dara E., Piotr Jankowski, and Keith C. Clarke
  2017 Privacy and False Identification Risk in Geomasking Techniques. *Geographical Analysis* 50:280–297.

Seidl, Dara E., Piotr Jankowski, and Atsushi Nara
  2018 An Empirical Test of Household Identification Risk in Geomasked Maps. *Cartography and Geographic Information Science* 46. DOI:10.1080/15230406.2018.1544932.

Smith, Kimbra L.
  2005 Looting and the Politics of Archaeological Knowledge in Northern Peru. *Ethnos* 70:149–170.

Strupler, Néhémie, and Toby C. Wilkinson
  2017 Reproducibility in the Field: Transparency, Version Control and Collaboration on the Project Panormos Survey. *Open Archaeology* 3:279–304.

University of Wyoming Department of Geography, Wyoming Geographic Information Science Center, and Wyoming Geographic Alliance
  2017 Selected Archaeological Sites in Wyoming (2013). In *Wyoming Student Atlas Online*. Electronic document, https://www.arcgis.com/home/item.html?id=6b4e33541a8a4338b7be994396df669c, accessed August 8, 2019.

Wieland, Shannon C., Christopher A. Cassa, Kenneth D. Mandl, and Bonnie Berger
  2008 Revealing the Spatial Distribution of a Disease while Preserving Privacy. *Proceedings of the National Academy of Sciences of the United States of America* 105:17608–17613.

Zandbergen, Paul A.
  2014 Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data. *Advances in Medicine* 2014. DOI:10.1177/1556264618759877.

Zhang, Su, Scott M. Freundschuh, Kate Lenzer, and Paul A. Zandbergen
  2017 The Location Swapping Method for Geomasking. *Cartography and Geographic Information Science* 44:22–34.

AUTHOR INFORMATION

Cecilia Smith ■ University of Chicago Library, University of Chicago, 1100 East 57th Street, Chicago, IL 60637, USA (ceciliasmith@uchicago.edu)