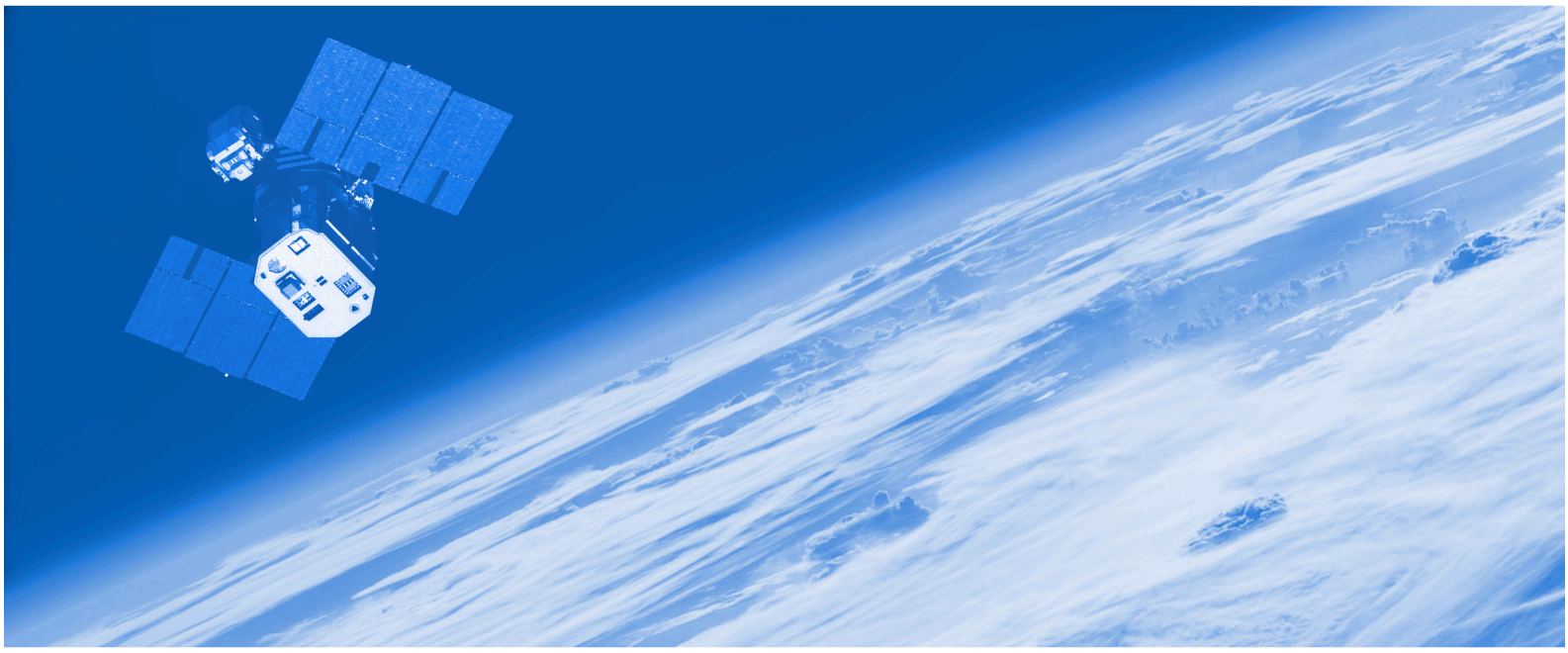
Location-Based Data in Crisis Situations

Principles and Guidelines

March 2019



AAAS | Scientific Responsibility, Human Rights and Law

Association for the Advancement of Science (AAAS). As a program of AAAS – the world's largest multidisciplinary scientific membership organization – SRHRL fosters and facilitates the responsible practice and application of science in the service of society. The Program is committed to promoting high standards for the practice of science and engineering; advancing the human right to enjoy the benefits of scientific progress and its applications; engaging scientists, engineers and their professional associations in human rights efforts; monitoring and enhancing assessment of emerging ethical, legal, and human rights issues related to science and technology; furthering the use of science and technology in support of human rights; and initiating activities to address the impact of developments at the intersection of science, technology, and law.

This report was prepared by:
Jessica Wyndham, Program Director, AAAS SRHRL
Ellen Platts, former Staff Assistant, AAAS SRHRL
Jonathan Drake, Senior Program Associate, AAAS SRHRL

Acknowledgement
The authors wish to acknowledge Dr. Susan Wolfinbarger who, together with Dr. Mark Frankel, conceived of and initiated this project. They also acknowledge the participants in the three workshops who came from academia, civil society, government, industry, international non-governmental and multilateral organizations, and professional scientific societies. These participants, together with multiple reviewers from academia and civil society offered their experience and intellectual rigor to the Principles and Guidelines.

Primary support for this project was provided by the U.S. National Science Foundation through award number 1560948.

Disclaimer
The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the AAAS Board of Directors, its Council, or membership, or the National Science Foundation.

Contact
AAAS welcomes comments and questions regarding its work. Please send information, suggestions, and any comments to SRHRL at srhrl@aaas.org.

Cite as: AAAS Scientific Responsibility, Human Rights and Law Program, "Location-Based Data in Crisis Situations: Principles and Guidelines" (Report prepared by Jessica Wyndham, Ellen Platts, and Jonathan Drake), March 2019.
DOI: 10.1126/srhrl.aax3877

Location-Based Data in Crisis Situations: Principles and Guidelines

Background

This document is the product of three workshops hosted by the American Association for the Advancement of Science (AAAS) in Washington, D.C. in 2016 and 2017. Participants in the workshops came from academia, civil society, government, industry, international non-governmental and multilateral organizations, and professional scientific societies. Their aim was to address the ethical issues associated with rapid growth in the use of location-based data such as remotely-sensed imagery, geotagged social media posts, and electronic communications records by crisis response actors.

Purpose and Scope

Data relating to the location of infrastructure, resources, and people can have positive and negative applications. In all circumstances, there are potential risks and benefits associated with collecting, aggregating, representing, using, and storing such data. In the context of crises, however, the nature and significance of the risks and benefits will differ from a non-crisis context. The following principles and guidelines aim to fill a gap in ethical guidance for the generation, collection, analysis, dissemination, and use of location-based data in crisis situations.

"Location-based data" is information that contains or is associated with position. In addition to spatial information, it typically contains a temporal component and is often, but not always, generated by electronic devices that are "location-aware" through the integration of satellite or terrestrially-based positioning systems. Location-based data are frequently shared via the internet and social media, and range from highly granular datasets, (e.g., individual geotagged photographs) to highly aggregated ones, (such as total reports of an earthquake over an entire region). Data that are location-based include data associated with specific coordinates (at varying levels of resolution), street addresses, or other information that can be geolocated, and named locations that may or may not be ambiguous (e.g., a city name in a Twitter profile, or a unique building mentioned in the text of a document). These data can apply either to individuals or groups, and in crisis situations such data are increasingly ubiquitous and used by multiple actors.

Per this definition, the term "location-based data" encompasses, but is not limited to: (a) data specifically and deliberately collected in preparation for or in response to an event by volunteers, local actors, or institutions, which contains a location or geographic component; (b) "Volunteered Geographic Information" (VGI) – user-generated content, created by individuals or groups for a purpose other than a deliberate data collection effort and which contains spatial information in a way that can be used by others (e.g., social media feeds); (c) ambient data and other data collected, frequently as a secondary process and without the specific knowledge and/or consent of the individual or groups (e.g., CCTV feeds, mobile phone records); and (d) remotely-sensed imagery, including geolocated photographs.

"Crisis situations" are situations of conflict and other situations of violence and natural disaster, including those that result in violations of human rights and international humanitarian law, and/or the destruction of cultural heritage. Because these situations are inherently dynamic and complex, affecting people and involving location-based data in a multiplicity of ways, the judgement of whether a given situation constitutes a crisis is often challenging, and will always contain an element of subjectivity. For

the purposes of planning and executing a response, however, defining the crisis in time and space is a key component of planning an effective response. The principles and guidelines that follow are designed to operate within this framework, and should be applied accordingly.

Primary Audiences for This Document:

- Newcomers to the use of location-based data, particularly in crisis situations;
- Volunteer networks of technical experts as well as other non-traditional data collection teams;
- Academic researchers such as geographers, including those working alone and those working at the behest of another actor;
- Local partners who may have their own codes of ethics, but need to find shared points of understanding with collaborators coming into their communities; and
- The broader community of non-experts involved in aid and relief work, as well as participatory science efforts.

Secondary Audiences:

- Existing organizations that are already familiar with and involved in the use of location-based data in crisis situations, including humanitarian and human rights organizations, government agencies and other entities with dedicated individuals or teams conducting research using location-based data and professional societies and associations with a focus on such research.

Principles:

1. Do No Harm: Identify and minimize potential risks of location disclosure, particularly as they may affect the vulnerability of individuals and populations
2. Define Your Purpose: Ensure action is mission-driven and goal-oriented
3. Do Good Science: Employ scientifically rigorous and responsible methods
4. Collaborate and Consult: Engage with local partners
5. Give Access to Your Data: Share data openly, when safe and practicable

Principle #1: Do No Harm - Identify and minimize potential risks of location disclosure, particularly as they may affect the vulnerability of individuals and populations

A. *Know your specific context:* When locally contextualized, location-based data will inherently implicate and identify potentially vulnerable individuals, groups, organizations, and resources which will likely change over time. Risks specific to a particular crisis may include a context of violence, of discrimination against and/or targeting of specific gender, ethnic, religious, cultural or other groups, and/or economic and technological marginalization. These risks pertain to volunteers collecting data as well as to individuals and groups whose data are collected. Depending on the context, these risks might be significantly increased in the context of hostile state, military or paramilitary actors. The kinds of risks that can arise from failing to address the context specific to a particular location can include data generated in good faith being used maliciously. Be prepared to walk away from projects if it is clear that victims, bystanders, and response personnel cannot be adequately protected.

See Decision Tree: "Should I collect", Step 2.

B. *Assess the risks in a given crisis as it evolves*: Because most location-based data have an inherent temporal dimension, changing contexts under conditions of crisis have the potential to enable new forms of identification that can result in new risks and opportunities. When using location-based data in a crisis situation, each stage of the data management cycle - from data collection to analysis, communication, storage, archiving, and deletion - engenders different risks and ethical obligations. Risk assessments should be conducted periodically throughout the crisis as it evolves. They should consider harmful outcomes related to the (mis)use of location-based data that have occurred in similar situations in the past, as well as potential harms that are foreseeable in light of the current crisis. Consulting with local communities and leaders is a key component of this process, as they may be able to envision risks that outsiders cannot.

See Decision Tree: "Should I collect", Step 4.

One possible way to organize these parameters is by incorporating them into a framework known as a "risk matrix" that plots the probability of the harm taking place against the potential outcome's severity, with the risk represented by the product of these two factors. Such a matrix, shown in the appendix, can be helpful in identifying which types of events must be planned for in advance. While assigning numeric values to the severity and probability of possible harms is a somewhat subjective exercise, comparisons with similar response efforts in different locations or at different times may provide helpful guidance for this purpose. It is likewise important to recognize which combinations of likelihood and severity constitute "red lines" that must not be

crossed, and to consider that the vulnerability of various groups is not static; it may emerge or recede over the course of a crisis. Likewise, be aware that the risks to individuals may differ from the risks to groups. In the event that a large degree of uncertainty exists concerning the probability and/or severity of potential adverse outcomes, caution is warranted; responders should consider whether a particular activity is vital to achieving the desired outcome. When conducting such an assessment, however, one must not lose sight of the potential benefits of pursuing a course of action, nor of the fact that parameters may influence decision-making that defy categorization as "risk" or "benefit".

C. *Proceed according to the risks associated with each stage of the data cycle in the context of the crisis*:

i. *Collection:* Select volunteers carefully, and take steps to protect their privacy. Ensure that data collected are consistent with the requirements and context of the situation, consider potential sources of error in the data, and whether those errors may add to risk. Protect the privacy and identity of subjects, and ensure volunteers are trained to do the same. Engage with local partners and, at minimum, ensure they are aware of the data collection effort and why it is being undertaken. Be aware that some location-based data –particularly user-generated data– may have been geotagged inadvertently. If volunteers submit data directly to the response effort, ensure that they are aware of the full nature of the data they are sharing, and allow them to opt out of submitting certain metadata if desired. If metadata is included, special care must be taken to determine whether the risks associated with the data may be enhanced when combined with other freely available information (e.g., identifying populations that prefer to remain hidden). If so, thought should be given to whether the data or its collection strategy might be modified in such a way that mitigates the risk while preserving the usefulness of the data to the crisis response (e.g., masking the data). The collection effort should be continuously monitored and evaluated to ensure that the data being acquired are still necessary. If not, the collection should be suspended.

See Decision Tree: "Should I collect", Step 3.

ii. *Sourcing:* When relying on location-based data collected by others, identify its provenance, assess its quality and determine its use based on the origin of the data. Evaluate the potential for false or spoofed location data, and evaluate the motivations that might be behind a release. Even if accurate, some data may have been selectively curated or released in order to advance a purpose unrelated to, or potentially at odds with, the response effort. When enlisting sources weigh the risks to the source against the risk of not having the data.

See Decision Tree: "Should I collect", Step 4.

iii. *Analysis*: Recognize that combining multiple datasets often enables insights that are more actionable than the sum of their parts. In light of this, assess the risks and benefits of using analytical methods that can generate new location-based information where none was provided (e.g., methods that can determine the home location of a Twitter user based on analysis of the full set of their tweets, profile information, and other internet based information

See Decision Tree: "Should I share", Step 2.

about them). Recognize that the context of different crises (e.g., natural disaster vs. conflict) can affect the level of antagonism present in a situation, and that this may affect decision-making.

iv. *Communication*: Whether data are shared or visualized and, if so, which data, should be decided based on a broad assessment of the associated risks. Consider not only the risks to individuals and groups, but also to ecosystems and infrastructure. Ensure that communication and/or visualization of data does not reveal the locations of vulnerable populations, sites, and artifacts, and take steps to avoid putting any at additional risk. Even when the underlying data are obscured, visualizations may be sufficiently detailed to enable mis-use. If the use of such information is essential, ensure that informed consent has been obtained from individuals who generated or are most affected by the data. The type and level of consent given when the data were collected should be taken into account, as should the potential that the data's availability may change from being helpful to being harmful in the future. When communicating location-based data findings with other actors, decision-makers, or the public, consider the ways the location information may be used and perceived by various audiences, and aim to disseminate and clearly communicate the data in a way and at a locational precision that will take these factors into account while minimizing personal risk, retraumatization, misinterpretation, or misuse.

v. *Storage:* Ensure that data are stored securely, and if practical consider storing datasets separately so that they cannot be combined in the event of a data breach. Use encryption for the storage of any data that contain, or could be extrapolated into containing, personally identifying, demographically identifying, or other sensitive information such as cultural data. Ensure that encryption keys are entrusted to a limited number of trustworthy individuals, but not so few as to hamper operations. Make regular backups of data and store them in multiple locations, such that a mishap at one site will not result in catastrophic data loss, while remaining cognizant of the increased risk associated with maintaining multiple copies of the data.

See Decision Tree: "Should I collect", Step 4.

vi. *Archiving and deletion:* Location-based data need not be kept indefinitely. Once the crisis has passed, consider whether there is any benefit to maintaining the data in archival form, and whether those benefits outweigh the potential risks. Input from the affected community can be particularly helpful in making this determination (see principle #4).

Principle #2: Define Your Purpose - Ensure action is goal-oriented

A. Ensure that collection of location-based data is necessary: Given the emerging forms of both benefits and unintended consequences that can arise from location-based data, collection of location-based data (both generally, and specific types of location-based data) should only occur when it is necessary to the articulated goal of the project, taking into account the needs and interests of the local community as well as of the overall crisis response effort. Groups that are new to data collection in crisis might want to consider partnering

See Decision Tree: "Should I collect", Step 1.

See Decision Tree: "Should I share", Step 1.

with recognized and experienced humanitarian groups for the design of collection and analyses protocols.

B. *Avoid overcollection of data*: Data should be collected only if a clear plan exists for its use; "it might be useful later" is not a sufficient justification for collecting. Recognize that, depending on the specific nature of the crisis situation, the threshold of what constitutes overcollection may change. Conducting a review of existing data already collected by others may help in making this determination. If increased data collection is deemed necessary, however, it still must always further the project's stated goal, and take place subject to the guidance articulated in principle #1, above. If possible, however, avoid having to re-collect data. When possible, consider both the data's immediate purpose and its likely purpose in any further stages of the crisis.

C. *Consider the boundaries of the data collection effort:* As early as possible in the crisis response effort, identify the spatial and temporal horizons that will be applied to data collection. Establish a clear set of criteria that can be used to define the end of the crisis and associated data collection efforts, taking into account factors such as funding, resources, and timeline. Although the time at which the crisis began is often obvious, this is not always the case, in which case these same or similar criteria can often be used to define the time before which the data to be collected should not cover. Similarly, the extent of the geographic area affected by a crisis is often difficult to define precisely, however data collection efforts will often have to be bounded in order to be useful – the boundaries of the data collection effort and the broader crisis may not overlap. When defining the boundaries of both the crisis and the response, consider both the availability of data and the ways in which it is going to be applied in the response effort.

Principle #3: Do Good Science - Employ scientifically rigorous and responsible methods

A. *Verify Data Sources:* Collection and analysis of VGI and location-based data should lead to accurate and verifiable results that are relevant and actionable in the crisis context. Ideally, data that are used in the response effort should come from reputable sources which are transparent about their own methods of collecting and aggregating the resulting data. Wherever possible, verify the provenance and accuracy of third-party data sources. If these cannot be established, consider whether alternative sources of information may be available, and keep in mind that no information at all is often superior to inaccurate information.

B. *Use caution when employing experimental methods:* Conducting response efforts with scientific rigor does not preclude either innovative approaches or non-traditional methods, however these should be identified as such, and should only be employed following a thorough risk-benefit analysis involving both the response's scientific experts and the affected community. If the data or their analysis have the potential to be used in a legal context, consider the impact that using experimental methods may have on such a proceeding, versus more established methods that have already been subject to rigorous scientific scrutiny. Consider that the requirement for scientific rigor does not imply that practices must be complicated; simple methods can still be rigorous.

C. *Train volunteers engaged in the collection of data*: Methods and technologies involving location-

based data are frequently distinct from those associated with other data, and must be understood by those employing them. Before individuals begin collecting data, provide training to ensure that they are aware of the risks and responsibilities (both to themselves and the research subjects) associated with the task, making them aware of any and all standards for data, including chain of custody standards. Consider the ethical obligations of individuals playing different roles in the data management lifecycle.

D. *Act in accordance with recognized standards of ethical conduct:* Early on in the response, recognize existing humanitarian, human rights, and ethical frameworks that may be applicable to the investigation, and come to an agreed understanding about common standards. Ethical research practices should be adopted in order best to serve the public good, including privacy and confidentiality, and the protection of human subjects. Consider that ethics and legality may not always align, and be prepared to address such discrepancies. Consider the legal environment of your volunteers. Understand the frameworks and standards guiding the activities of other individuals and organizations operating in similar situations, and determine the standards guiding the collection, analysis, communication, and sharing of location-based data in that context.

E. *Consider potential sources of bias:* Relying on location-based data, particularly when gathered in a way that relies on the technological capacities and access of local communities and individuals will inevitably give rise to inherent bias in the data and risk empowering some segments of society while perpetuating or exacerbating the marginalization of others. Such biases are often amplified in crisis situations. Assessing the extent of this bias and, where possible, correcting for it will be critical to achieving successful outcomes in a crisis situation. For example, if connectivity is known to have been degraded in certain neighborhoods of a city, reports coming out of that zone might be given more statistical weight than those originating in areas where communications infrastructure remains intact. Recognize that these biases may be temporal as well as spatial, and evaluate the relevance of information accordingly. Identify, be respectful of, and assess how the cultural context may impact the investigation and its data. When sharing data or publishing or otherwise communicating information derived from the data, clearly communicate the level, source, and kind of bias in the data, along with any associated uncertainty. In addition to bias, make every possible effort to perform accuracy assessments which will enable the quantification and communication of error rates associated with the data.

F. *Submit to peer review:* The methods that a given investigation relies upon should be clearly identified and open to peer and community review, even if that is not immediately possible in a crisis situation. Provided that data collection takes place in a responsible manner, rigorous scientific analysis and peer review can take place at any time, including after the crisis has abated. Such review might, for example, accompany reports to funding organizations.

Principle #4: Collaborate and Consult - Engage with local partners

A. *Engage community actors:* Recognize that the people best qualified to understand the needs of a community in the wake of a disaster are the people living there. Contact community leaders, explain who you are, and the methods you are contemplating applying – they may already have ideas that you have not considered. Manage community expectations regarding data collection and be realistic about what you are attempting to achieve. Whenever possible without compromising safety, consult and collaborate with these individuals and their communities to best understand their context and needs, and incorporate their input into the design of your work and methods. Recognize that trusted community networks may be able to solve dilemmas –ethical and otherwise– that you are unable to. It may not always be necessary to collaborate, however the process of disclosure regarding methods, data, funders, et cetera is highly effective in establishing and maintaining trust. Strongly consider putting such disclosures in writing.

See Decision Tree: "Should I collect", Step 2.

B. *Build local capacity*: Work with subject matter experts to identify the local populations and partners relevant to the data collection effort. When safe and practicable, engage with local partners to develop any new data collection tools, define data requirements, gather baseline data in advance of potential disasters (known as "data preparedness"), and perform gap identification, data validation, and verification. In particular, engage with local partners to understand the opportunities as well as limitations, risks, and biases (including those that may be introduced by the partners themselves), presented by using location-based data in the context of the crisis. These partners can be extremely helpful in identifying contextually appropriate solutions to difficult issues. If local capacity exists, it should not be duplicated or undermined.

C. *Obtain Informed Consent:* Particular attention should be paid to the ethical issues associated with obtaining consent in the midst of an emergency or ongoing conflict, recognizing that the nature and form of consent may differ depending on the role and relationship of the person or group of individuals with regard to the data. When initial consent is obtained, the data collected under that consent should be subject to ethical prescriptions regarding the use of those data for a new purpose. Existing ethical codes such as those described in principle #3 may be useful in providing specific guidance in this regard.

See Decision Tree: "Should I collect", Step 3.

D. *Promote community resilience and responsible stewardship of data through the communication and repatriation of knowledge*: Local partners engaged in and/or affected by the collection of location-based data have a right to know how that data is being used, and should be granted access to any data collected by and about them, as well to as any insights and analysis derived from these data, taking the community's level of data literacy into account. They also have the right to rectify false, inaccurate, or incomplete data collected about them, to remove themselves and their associated data from the data collection systems at any time, and to have input regarding what will happen to that data after the investigation (deletion, archiving, etc). They should also have recourse to a defined mechanism for raising concerns or

See Decision Tree: "Should I share", Step 3.

making complaints about the data collection effort. This mechanism should be resilient enough to remain accessible even after funding for the response effort has ended. These conversations should include a discussion of the tension between releasing the data publicly and keeping it protected, and the risks and benefits of each course of action. This dialogue must take place in a manner appropriate to the level of data literacy present in the community, elevating it where possible. Researchers should make every effort to validate their general findings and impressions with the local community before they leave. All considerations regarding community access to an stewardship of data should take the local legal context into account. After the crisis has passed, the community must be involved in the assessment phase of the response as described below.

E. *Conduct assessments post-response:* Following the intervention and together with local partners, evaluate the response effort's use of location-based data in the context of measurable outcomes. Consider both positive and negative outcomes, and consider how the availability of location-based data –or lack thereof– affected these results. In instances where location-based data were helpful, determine whether the level of granularity was excessive, sufficient, or inadequate, and ensure that this information is available to the planners of future response efforts.

Principle #5: Give Access to Your Data - Share data openly, when safe and practical

A. *Share data, but assess the risks and accept the consequences*: Data should only be shared once an analysis of the potential consequences of its dissemination has been completed, taking into account the nature of the data to be shared and the individual and/or group with which it would be shared. Define who is accountable for data-related harms, and establish mechanisms for addressing such harms which involve the local community. Know that people will misinterpret maps and visualizations, and design them to minimize misinterpretation. Actors tasked with implementing the sharing of data must be trained in current data security practice and understand the implications of sharing the data.

See Decision Tree: "Should I share", Step 2.

B. *Assess the nature and form of the data to be shared:* The risks associated with the storage and sharing of data are likely to differ based on the type of data and its level of granularity, and the policies associated with sharing that data should be defined accordingly. Sharing of data should take place in the context of a clearly defined use case. In no instance should personally identifiable information be provided, including names, home addresses, phone numbers, IP addresses, or other obvious identifiers. Always remember that location-based data has the potential to be used as personally identifiable information even if it contains no such information explicitly. The form in which aggregated data are shared should reflect best practices and ensure that technical and administrative safeguards make re-identification difficult, if not impossible.

C. *Assess the individuals and/or groups with which data will be shared:* Greater levels of disclosure, autonomy, and access to data may be allowed for highly trusted data recipients with strong, data security, audit, and access control processes and whose goals in using the data coincide with the purposes for which it was originally collected.

D. *Ensure that the roles and responsibilities of key personnel related to data security are clearly defined:* When a project is initiated, a single individual or group should be designated to coordinate on all things related to data protection and privacy. Depending on the size of the project, a person or group responsible for data acquisition and licensing may also be designated. In both cases, these individuals or groups should be able to respond rapidly to critical events related to data security.

See Decision Tree: "Should I share", Step 4.

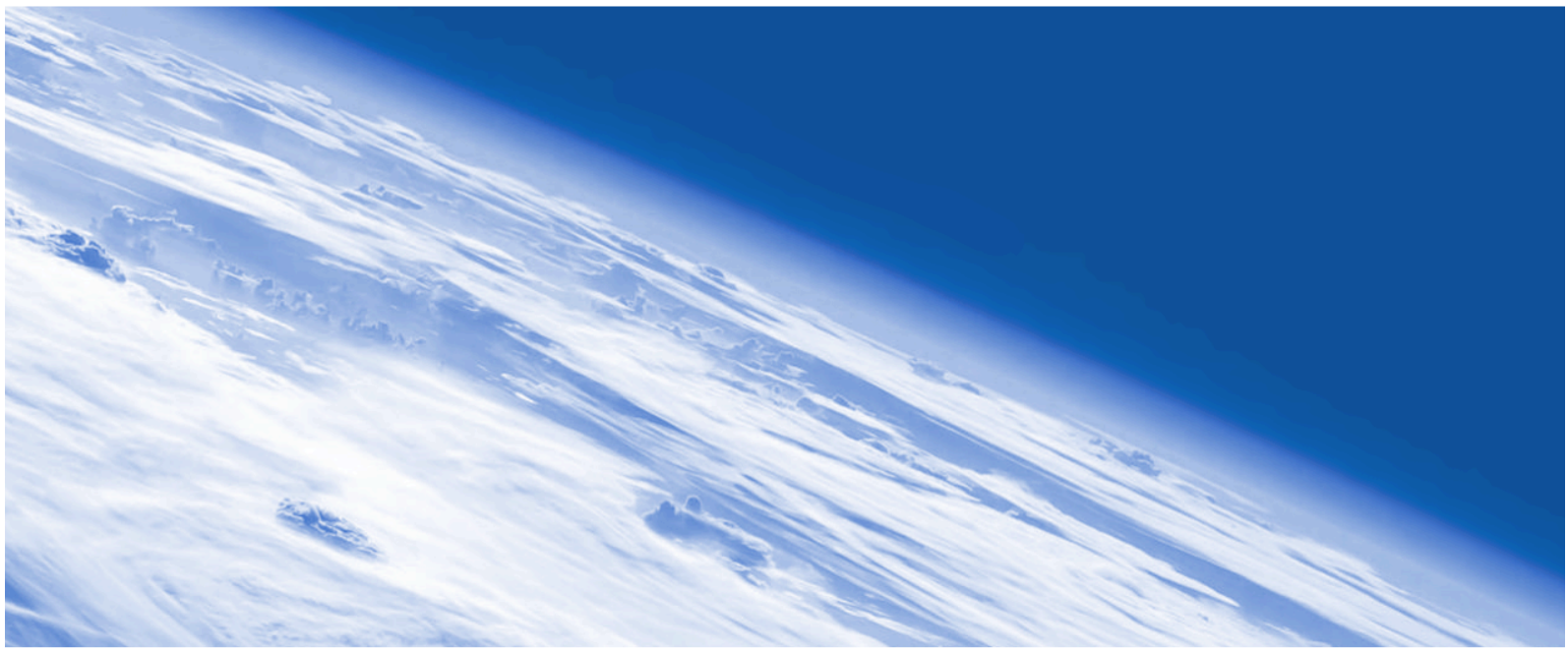
E. *Have a plan in place to respond to a data breach:* Take measures to prevent location-based data in your possession from being hacked or shared accidentally. Discuss mitigation strategies in the event of a breach prior to beginning the collection effort, taking into account the specific nature of the data being collected. Have a communication plan in place, and ensure that the contact information for all relevant stakeholders is readily available throughout the organization. These should include, at a minimum, leaders in the communities with which you are working, as well as any relevant local, regional, and national level civil authorities. Be aware of the laws in place in the host country regarding personal information, and have at least one team member tasked with ensuring compliance.

F. *Take extra caution in the context of a violent conflict*: Extra caution should be taken in collecting and sharing location-based data in situations where groups on the ground are in conflict. In such situations, location-based data may be more likely to lead to serious negative consequences for vulnerable populations if adversarial groups gain access to the data. This may include consulting with data protection experts and/or establishing a system of paths and gateways that allow data to be transferred effectively only in a deliberate and controlled way. Recognize that some data may be too sensitive to be shared until after an event is over.

Appendix: Sample Risk Matrix Associated with Events Affecting Staff in a Hypothetical Response

	Low Danger (1)	**Medium Danger (2)**	**High Danger (3)**	**Severe Danger (4)**
Low Probability (1)	Disorganized attempts to spoof data (1)	Organized attempts to spoof data (2)	Disinformation campaign targets response (3)	Foreign military intervention (4)
Medium Probability (2)	Random Phishing Attacks (2)	Local population mistrusts response (4)	Targeted phishing attacks (6)	Targeted violence against response (8)
High Probability (3)	Volunteers collect data of varying quality (3)	Chain of command disrupted within security forces (4)	Disinformation campaign targets vulnerable populations (9)	N/A (12)
Certain (4)	Dataset contains errors/omissions (4)	Telecommunications disrupted (8)	N/A (12)	N/A (16)

Each cell in this matrix represents a possible event that could impact the response effort. The position of the event in the matrix is defined by the danger represented by the event and the likelihood of it taking place, with the risk defined as the product of those two factors. This is a simplified example; in an actual response, multiple matrices may be necessary for different phases of the operation, and multiple types of events may occupy the same cells in the matrix. In this particular matrix, values of twelve pass the "red line" of unacceptable risk.



AAAS | Scientific Responsibility, Human Rights and Law