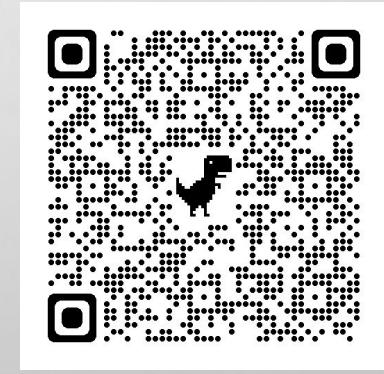


# Data Protection Workshop



Humanitarian  
OpenStreetMap  
Team



Link to google slides



# Objectives for today



Be aware of:

- HOT's approach to Data Ethics and Protection
- The processes and tools we've developed to address and implement the Policy

Presentation: 30 minutes

Scenarios: 20 minutes

Discussion: 30 minutes

Understand:

- How you can make use of it!

# HOT's protection framework



## What is it?

A practical guide and tool for entry level protection and risk assessments

Data collection and mapping project in an open data context

Support community-led open mapping exercises

## What is it NOT?

A comprehensive document covering all aspects of protection for experts

Operational safety and/or security

Data storage and server security

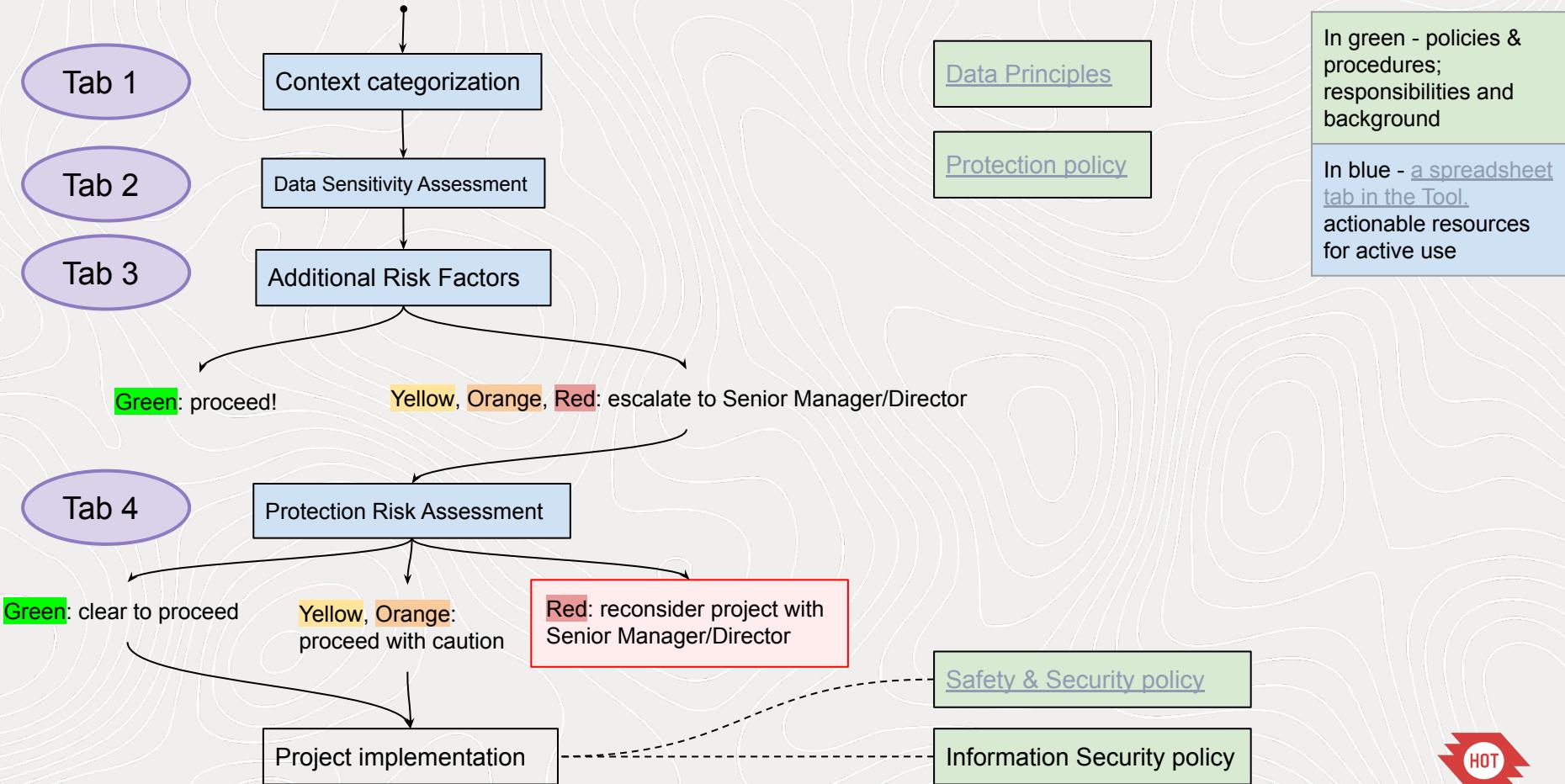
# Why do we need this?



Open mapping communities and program teams face unique challenges in how we work with data:

1. Using the OpenStreetMap platform allows for greater reach and accessibility of data, which is at the heart of our work. However, this can lead to unintended exposure to risk and the potential for misuse of data. As a result, we need to identify and mitigate any potential risks and harm.
2. Data we contribute to OpenStreetMap will be available online for a long time. With that in mind, we need to anticipate potential future risks and harm.

# Data Ethics & Protection Process Overview



# Data Principles

## Open and accessible

Contribute to and advocate for open data, first and foremost OpenStreetMap, and enable anyone to make effective use of this data.

## Useful and usable data

Collected and contributed OSM data should meet a purpose that's well defined and described, and conform to given data quality standards so it can be understood and used within and outside of HOT's Impact Areas.

## Inclusive and representative

Ensure communities at all levels can access and work in the OSM ecosystem by how HOT prioritizes, creates, and structures data collection, tech, and community resources.

## Ethical data and protection

Go for the most meaningful collaboration in data management and planning. Uphold policies and guidance to minimize risk of harms, including impact assessments and informed consent for any data collection or use, based on the people and communities we work with.

# Data Principles

## Open and accessible

1. Contribute data to open data sources. First and foremost to OpenStreetMap under the [ODbL](#) - the license selected for OSM data which allows free sharing, use, and adaptation.

## Useful and usable data

1. Be deliberate in what data we collect and generate and for what purposes, and align this with the pathways to impactful use of data.
2. Align data collection and mapping with the priorities of communities and partners.

## Inclusive and representative

1. When creating mapping tasks, consider who and what we're including, and who we're omitting by how we define the area of impact and the data models.

## Ethical data and protection

3. Conduct risk and data impact assessments to assess and minimize risk of harm for all projects together with the relevant communities. In situations where conflict is a factor, follow the "[Conflict zones mapping policy](#)".
4. Ensure everyone at HOT is compliant with the application of data and protection principles.

# Open and accessible

---

## What this means:



Contribute to and advocate for open data, first and foremost OpenStreetMap, and enable anyone to make effective use of this data.

## What we do:

1. Contribute data to open data sources. First and foremost to OpenStreetMap under the [ODbL](#) - the license selected for OSM data which allows free sharing, use, and adaptation.
2. Ensure (re)usability of data we contribute to OSM by providing ample documentation and metadata.
1. Advocate and push for open, [ODbL compatible](#), licensing of existing datasets from partners.
3. Provide easy ways of accessing and using OSM data for a variety of users, and explain how OSM data can be used and incorporated.
4. Make data available in easy to understand and use formats - including digital formats and platforms ([HDX](#)), mobile applications, and paper maps.

# Useful and usable data

---

## What this means:



Collected and contributed OSM data should meet a purpose that's well defined and described, and conform to given data quality standards so it can be understood, used, and re-used within and outside of HOT's Impact Areas.

## What we do:

1. Be deliberate in what data we collect and generate and for what purposes, and align this with the pathways to impactful use of data.
2. Align data collection and mapping with the priorities of communities and partners.
3. Improve the quality of OSM data and tagging schemas and standardization, including localization and contextualization.
4. Ensure we follow OSM guidelines and procedures, such as organized editing, and import processes.
5. Provide technical resources and services, support, and case studies (including measurement of the usability of data) for Impact areas use cases.

# Inclusive and representative

---

## What this means:



Ensure communities at all levels can access and work in the OSM ecosystem by how HOT prioritizes, creates, and structures data collection, tech, and community resources.

## What we do:

1. When creating mapping tasks, consider who and what we're including, and who we're omitting by how we define the area of impact and the data models.
2. Design, create, promote, and use technology and tools that are accessible and usable. Give specific attention to more vulnerable and/or minority groups to reduce barriers to access.
3. Commit to reduce inequities to contribution and participation before, during, and after mapping - such as access to devices, mobile data, digital and map literacy, opportunities and economic viability of volunteering time, etc.
4. Prioritize localization and accessibility of tools, documentation, and tech by providing multiple relevant languages.

# Ethical data and protection

---

## What this means:



Go for the most meaningful collaboration in data management and planning. Uphold policies and guidance to minimize risk of harms, including impact assessments and informed consent for any data collection or use, based on the people and communities we work with.

## What we do:

1. Apply a workable process of engagement and informed consent across our activities, including for remote sensing and digitization (while accounting for an ever more sophisticated technological landscape).
2. Ensure respect for the beliefs, cultures, lifestyles, and choices of communities in the decision making around data collection and sharing, and in our use of collected information. Engage partners in the ethical use of information.
3. Conduct risk and data impact assessments to assess and minimize risk of harm for all projects together with the relevant communities. In situations where conflict is a factor, follow the "[Conflict zones mapping policy](#)".
4. Ensure everyone at HOT is compliant with the application of data and protection principles.

# “Do No Harm”



HOT—and associated open mapping communities—are responsible for ensuring that people and communities are not harmed by:

- Our activities
- The data we create
- The data we keep

*Note that this Protection Framework addresses harm to people and communities, **not** risk to ourselves, our staff, or mappers. It is distinct from Safety and Security, which is an important but separate topic!*

# What is harm? How can it come about?



People or communities can be harmed by open mapping activities and/or data by

- Increasing vulnerability to or likelihood of attacks
- Stigmatization
- Loss of privacy/revealing of Personally Identifiable Information
- Economic losses

Can you think of others?

# Thinking About Harm to Communities



We're picking a couple real mapping projects to think about:

WHO is impacted?

WHOM or WHAT is impacting them?

HOW are they being affected?

HOW does our DATA or ACTIVITIES contribute to the impact?

1

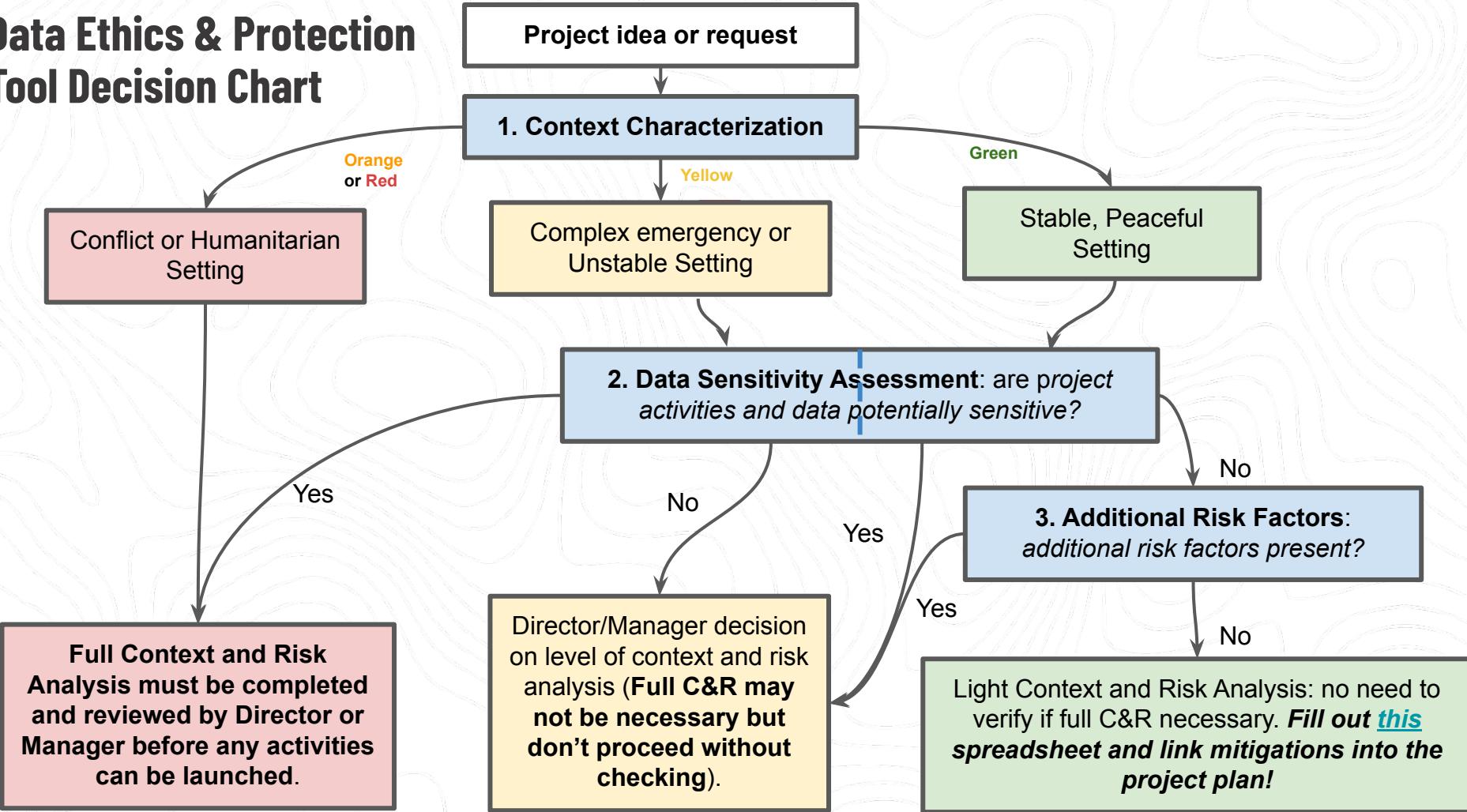
2

3



## Context Categorization & Assessment

# Data Ethics & Protection Tool Decision Chart



# Context Characterization

## Declared Interstate War

- WW2, Iran/Iraq, Eritrea/Ethiopia

## Undeclared Interstate War

- Russia/Ukraine 2022

## Recognized Civil War

- South Sudan independence war

## Unrecognized Civil War

- Darfur rebellion, Eastern DRC, Mali, Myanmar

## Internecine Violence

- Isis, Mai Mai in DRC, Burkina Faso, Mindanao in Philippines

## Complex Emergency

- Floods in Pakistan or Rohingya settlement (humanitarian context combined with natural disaster)

## Lawless Areas

- Northern Mexico drug gangs
- Cité de Soleil in Haiti
- Tribal areas in Pakistan

## High Criminality

- Urban slums such as Kibera, Karachi, Rio, or Mexico City

## Conflict-Adjacent areas

- Generally peaceful and stable but hosting/containing people fleeing or affected by nearby conflict (borders of Tanzania/Burundi, Thailand/Burma)

## Entrenched Poverty/Corruption

- Informal areas/slums without notable internal violence such as informal areas of Dar es Salaam

## Poverty

- Lack of resources due to underdevelopment rather than deliberate malfeasance. Generally characterized by subsistence farming (rural Tanzania, Bali)

## Stable, peaceful, developing

- May still be poor but things are improving, government, population, other actors reasonably aligned.

# Data Sensitivity

Not all data has equal potential for harm! Consider at least these aspects:

***Are you collecting personally identifiable information (PII)?***

***Are you collecting data considered sensitive by the community, project team, organization, or government?***

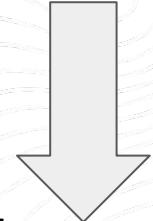
***Are you collecting data about people's health or medical status?***

***Are you working with or mapping vulnerable or marginalized groups?***

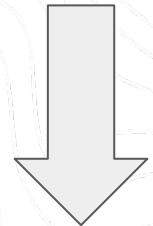
***Are you sharing otherwise unavailable data with other actors, including governments, who may use it to make specific decisions about individuals or communities?***

***Are there other actors or agencies with a vested interest or stake in map data (disputed boundaries for example)?***

**Sensitive Data!**



If you answer "Yes" to any of these questions:



# Additional Risk Factors

A given country may be generally peaceful and stable, but may require more than a light protection risk assessment if specific contextual factors are present.

*Is there an election or transfer of political power going on?*

*Will work be done in an area of sub-national conflict? Some countries have defined sub-regions in conflict while the larger area is generally peaceful.*

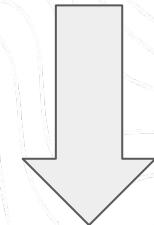
*Will work proceed during major disruptions such as crop failures, economic collapse, or prolonged natural disasters?*

*Will work proceed during an outbreak involving significant mortality, quarantines, and other heavy public health measures/restrictions, such as for Ebola or peak waves of Covid-19?*

*Are there large influxes of people from conflict-affected areas? A peaceful area may be at risk during such influxes.*

*Is this a new context for open mapping, with no established OSM or similar community?*

If you answer "Yes" to any of these questions:



Risk Factors Present!

4



## Risk Assessment and Mitigation

# Identify and Rate Potential Harm

Potential Harm	Vulnerability (1-5)	Likelihood (1-5)	Impact (1-5)	Risk	Mitigation	Residual risk after mitigation
Targeted attacks guided by data revealing ethnicity	5	3	5			
Informal neighborhoods bulldozed	3	2	3			
People seeking HIV care identified and stigmatized	4	3	5			

## On Vulnerability, Likelihood, Impact



How we quantify and prioritize risk:

Vulnerability	People's level of exposure (time and number of people) to this risk
Likelihood	How likely is the risk to materialize?
Impact	How bad will it be if it happens (1 = inconvenience, 5 = many deaths)?

The tricky part here is the difference between Vulnerability and Likelihood. Vulnerability is basically the number of person-hours exposed to the risk, and likelihood is the chance of it happening at any given time to someone who is exposed.

# Multiply Vulnerability x Likelihood x Impact to give Risk

Potential Harm	Vulnerability (1-5)	Likelihood (1-5)	Impact (1-5)	Risk	Mitigation
Targeted attacks guided by data revealing ethnicity	5	3	5	75	
Informal neighborhoods bulldozed	3	2	3	18	
People seeking HIV care identified and stigmatized	4	4	5	80	

1	1	1	1	Low
1	1	2	2	Low
1	2	2	4	Low
2	2	2	8	Low
2	2	3	12	Low
2	3	3	18	Medium
3	3	3	27	Medium
3	3	4	36	Medium
3	4	4	48	Medium
4	4	4	64	High
4	4	5	80	High
4	5	5	100	High
5	5	5	125	High

# Consider Mitigation Measures

Potential Harm	Vulnerability (1-5)	Likelihood (1-5)	Impact (1-5)	Risk	Mitigation	Residual risk after mitigation
Targeted attacks guided by data revealing ethnicity	5	3	5	75	<ul style="list-style-type: none"> <li>Do not collect data that reveals ethnicity</li> <li>Secure data with encryption</li> </ul>	
Informal neighborhoods bulldozed	3	2	3	18	<ul style="list-style-type: none"> <li>Ask community and stakeholders if there's a policy to suppress informal neighborhoods—if so do not map them.</li> </ul>	
People seeking HIV care identified and stigmatized	4	4	5	80	<ul style="list-style-type: none"> <li>Anonymize all health data before sharing</li> <li>Secure data with encryption</li> </ul>	

# Rate Residual Risk (what's left after mitigation)

Potential Harm	Vulnerability (1-5)	Likelihood (1-5)	Impact (1-5)	Risk	Mitigation	Residual risk after mitigation
Targeted attacks guided by data revealing ethnicity	5	3	5	75	<ul style="list-style-type: none"> <li>Do not collect data that reveals ethnicity</li> <li>Secure data with encryption</li> </ul>	Low (if not collected) High (if secured)
Informal neighborhoods bulldozed	3	2	3	18	<ul style="list-style-type: none"> <li>Ask community and stakeholders if there's a policy to suppress informal neighborhoods—if so do not map them.</li> </ul>	Low to Medium
People seeking HIV care identified and stigmatized	4	4	5	80	<ul style="list-style-type: none"> <li>Anonymize all health data before sharing</li> <li>Secure data with encryption</li> </ul>	Medium to High

# Discussion



## Questions per group:

- What is the risk rating for your scenario?
- What are the most impactful risks (2-3)?
- What are the mitigations you'd put in place?
- **Would you go ahead with this project?**

# Questions



From our side:

- Are you interested in applying or using these tools?
- Anyone that would like to help review and trial?

# Resources

- Protection Risk Assessment Template -  
[PLEASE COPY Data Ethics and Protection Tool](#)
- GitHub repository:  
[https://github.com/hotosm/data\\_protection\\_project/](https://github.com/hotosm/data_protection_project/)
- HOT Mapping in Conflict guidance:  
<https://www.hotosm.org/updates/mapping-in-conflict/>



PLEASE\_COPY\_Data\_Ethics\_and\_Protection\_Tool

File Edit View Insert Format Data Tools Extensions Help

F3 A B C D E

**0 Proposal and Scope**  
Provide background information that is required for informed review and approval of the risk assessment.  
 Completed

**1 Context Categorization**  
After completing 1. Context Categorization, please indicate the category of context the project will operate in. The checkbox here corresponds to the single checkbox in the Context Categorization.  
 GREEN  
 YELLOW  
 ORANGE  
 RED

**2 Data Sensitivity Assessment**  
After completing 2. Data Sensitivity Worksheet, please indicate if the project will involve sensitive data. If any checkboxes are ticked in the Data Sensitivity worksheet, the Sensitive Data Present checkbox must be ticked.  
 No Sensitive Data  
 Sensitive Data Present

**3 Additional Risk Factors**  
After completing 3. Additional Risk Factors, please indicate if the project will be subject to additional risk factors. If any of the checkboxes are ticked in the Additional Risk Factors worksheet, the Additional Risk Factors Present checkbox must be ticked.  
 No Additional Risk Factors  
 Additional Risk Factors Present

If all of the above checkboxes are GREEN, your project is likely business-as-usual Open Mapping, without unusual Ethical Data and Protection risk. Since this framework is not intended to be an obstacle to business as usual, you may proceed by yourself to complete 4. Protection Risk Assessment.

If any of the checkboxes are YELLOW, ORANGE, or RED, please consult your Senior Manager or Director for guidance.

**4 Protection Risk Assessment**  
After completing 4. Protection Risk Assessment, please indicate the highest remaining risk score after mitigation. Only the highest score must be entered here, but the scores before mitigation need not be considered.  
 1-12 (Low)  
 13-48 (Medium)  
 49-125 (High)

*If any of the risks remain RED after mitigation, please consult your Director for guidance.*

**5 Due Diligence Check**  
 Due Diligence Check completed

**6 Signoff and Approval**  
(if there is a second reviewer)



## Example Scenarios

# Scenario: Ebola outbreak



**Scenario:** During an Ebola outbreak, contact tracing is a key intervention to stop the spread of the disease, and mapping is key to contact tracing (medics need to know where the village or neighborhood the patient came from is). This is required to determine patient origins, identify infection hotspots, and to run contact tracing.

**Project proposal:** collect knowledge and (missing) names of population centres, along with key health infrastructure, in several provinces in western Uganda to serve as the basis for health interventions.

# Scenario: Satellite/UAV imagery program



**Scenario:** to support healthcare work in north-eastern Nigeria, it's critical to have up to date information on population distribution, movement, and accessibility. A donor is offering to fund a project to explore addressing these challenges via advanced AI/ML capabilities.

**Project proposal:** use AI/ML feature detection on a satellite imagery mosaic to derive building and road datasets in Borno state, Nigeria. Align and import these datasets with OpenStreetMap to complement existing data, and align it with existing data on healthcare and other primary services.

# Scenario: Evacuation route mapping



**Scenario:** the hasty retreat of foreign troops from Afghanistan has left many people scrambling to leave the country. To assist organizations helping for evacuation, a university is requesting assistance collecting and publicizing up to date information.

**Project proposal:** review and complete road networks. Identify viable escape routes, and crowdsource potential hazards such as police checks and roadblocks. Publish collected data on OpenStreetMap in combination with scenario specific web maps.

# Scenario: Informal Community mapping



**Scenario:** for years, Dar es Salaam has been plagued by large scale pluvial flooding in the increasingly urbanized Msimbazi River delta. Informal settlements are covering what used to be floodplains along river banks, and are devoid of even the most basic services. Large infrastructure investments are being planned to increase flood resilience, and identify

**Project proposal:** document flooding sites, occurrences, return periods, and severity. Map informal settlements, including the presence (and absence) of key services such as drinking water, sanitation, healthcare, and education.

# **Scenario: indigenous communities land-use mapping**



**Scenario:** An indigenous community in Peru is advocating to receive land deeds and protection against deforestation as land grabbers and drug cartel are encroaching on the forest.

**Project proposal:** Use satellite imagery to review the deforestation level currently along changes over time, and identify the land use on the given territory. Map key infrastructures to facilitate the recognition of the rights of the community on the land.