

Datasets

Ushahidi and our community provides datasets for research purpose. This is a document to determine what will be the best process to open datasets for research, but be respectful of privacy and security guidelines.

Outline for software company and deployers

purpose of the dataset and potential uses

what is the policy for releasing data

what needs to be anonymized

what is the policy for releasing data

what is the data storage policy if released to the Ushahidi community

methodology for cleaning data

what needs to be anonymized:

1. Name (be it first, last or family name)
2. Phone numbers
3. Email addresses
4. twitter handles

What are the data guidelines to be followed for sharing and guidelines

methodology for cleaning data

What are the data guidelines to be followed

For researchers:

- 1) purpose
- 2) why this was cleaned/filtered
- 3) process

Use Case: Haiti

Document process for Haiti data cleanup

do an internal check - eg. random number generated and spot check