Guidelines for Annotating Tweets

In this task, we will annotate tweets sent during disasters. We will use a tool called GALLOC (https://geoai.geog.buffalo.edu/iGALLOC/) to do the annotation.

1. Data Annotation Task

We will annotate four things from a tweet: (1) location description text; (2) location category; (3) spatial footprint; and (4) topic of the tweet. It's possible that one tweet may not contain any location description; in such a case, we only need to annotate (4). In the following, we describe each of the four.

(1) Location description text:
Location description is a piece of text in the tweet that describes a location. For example:

- *#HurricaneHarvey family needs rescuing at* 1280 Grant St. in Cypress, Texas 77249*, 2 elderly, one is 90 not steady in her feet*
- *#Harvey LIVE from* San Antonio, TX*. Fatal car accident at* Ingram Rd.*, Strong winds.*

A tweet may contain one location description only, two or more location descriptions, or no location description at all.

(2) Location category:
For each location description, we will select from a dropdown list of 10 possible categories. These categories and their examples are provided as below:

Category	Location description examples
C1: Door number addresses	● #HurricaneHarvey family needs rescuing at <u>*1280 Grant St. in Cypress, Texas 77249*</u>, 2 elderly, one is 90 not steady in her feet ● Papa stranded in home. Water rising above waist. HELP <u>*812 Wood Ln, 77828*</u> #houstonflood
C2: Roads and highways	● #Harvey LIVE from San Antonio, TX. Fatal car accident at <u>*Ingram Rd.*</u>, Strong winds. ● 9:00AM update video from Hogan St over White Oak Bayou, <u>*I-10, I-45*</u>: water down about 4' since last night. . .'
C3: Road and highway segments	● Streets Flooded: <u>*Almeda Genoa Rd. from Windmill Lakes Blvd. to Rowlett Rd*</u>. #HurricaneHarvey #Houston ● HELP! A pregnant lady is stuck in her car on <u>*I-45 between Cypress Hill & Huffmeister exits*</u> #harvey
C4: Highway exits	● Need trailers/trucks to move dogs from Park Location: Whites Park Pavillion off I-10 <u>*exit 61*</u> Anahuac TX ● <u>*Townsend exit*</u>, Sorters road and Hamblen road is flooded coming from 59 southbound #HurricaneHarvery #Harvey2017
C5: Intersections (a	● Guys, this is <u>*I-45 @Main Street*</u> in Houston. Crazy. #hurricane #harvey. . .

river and a road; two roads; crossing)	● Major flooding at _Clay Rd & Queenston_ in west Houston. Lots of rescues going on for ppl trapped…
C6: Natural features	● Frontage Rd at the river #hurricaneHarvey #hurricaneharvey @_San Jacinto River_ ● _Buffalo Bayou_ holding steady at 1000cfs at the gage near Terry Hershey Park
C7: Other human-made features	● If you need a place to escape #HurricaneHarvey, _The Willie De Leon Civic Center_: 300 E. Main St in Uvalde is open as a shelter ● Houston's _Buffalo Bayou Park_ - always among the first to flood. #Harvey
C8: Admin units and neighborhoods	● #HurricaneHarvey INTENSE eye wall of category 4 Hurricane Harvey from _Rockport, TX_ ● A picture of the entrance to the _River Oaks_ neighborhood in Houston off Kirby Drive.#Harvey #HurricaneHarvey https://t.co/oGBVf0iyzN
C9: Multiple areas	● Anyone doing high water rescues in the _Pasadena/Deer Park area_? My daughter has been stranded in a parking lot all night ● FYI to any of you in _NW Houston/Lakewood Forest_, Projections are showing Cypress Creek overflowing at Grant Rd
C10: Uncertain locations	● Talking distributed energy today & ran into @KISSOnline before they light up _St.Paul_ for #theChildrenMatter Hurricane Harvey benefit! _(reason: multiple places have this name; cannot resolve it to a specific location because the location type is unknown)_ ● College Station-St. Thomas canceled due to #Harvey. Cougs to host St. Pius X at _Cougar Stadium_ at 7:30 Friday. @TheCougarWay _(reason: there exists multiple Cougar Stadiums in the study area)_

For Category 10 _Uncertain Locations_, it refers to the situation that the text is likely a location description but it is difficult to determine its actual spatial footprint for various reasons.

(3) Spatial footprint:
For each location description, we will identify its spatial footprint (i.e., spatial representations on a map) which can be points, lines, or polygons. The GALLOC tool will assist the process of identifying spatial footprints, and more details are provided later.

(4) Message topic:
For each text message, we will choose its main topic from the dropdown 10 topics below. If a message is about multiple topics, we will choose the most relevant or most urgent one. For example, if a tweet is about both "P2: Sympathy and support" and "P4: Displaced people and evacuations", we will choose "P4: Displaced people and evacuations".

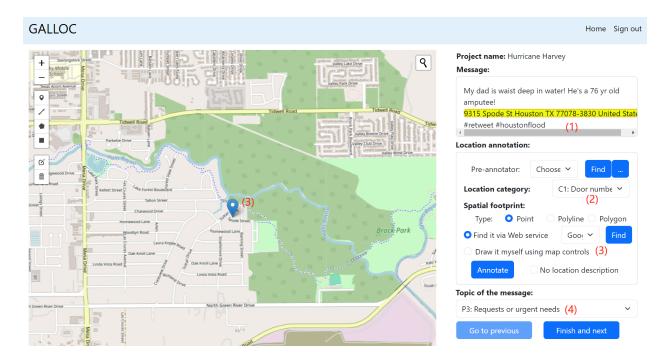| _Topic_	_Topic description_
P1: Information and advice	Messages asking or sharing disaster information and asking or providing advice for others

P2: Sympathy and support	Messages expressing emotional support, prayers, or well wishes
P3: Requests or urgent needs	Messages requesting urgent help, assistance, or expressing a critical need
P4: Displaced people and evacuations	Messages reporting on people displaced from their homes or updates about evacuation statuses
P5: Injured or dead people	Message reporting individuals who are injured or dead due to the disaster
P6: Missing or found people	Messages reporting individuals who are missing or have been found during a disaster
P7: Damages and disruptions	Messages reporting damages to infrastructure or facilities, as well as disruptions such as power outages, road closures, or business shutdowns
P8: Rescue, volunteering, or donation effort	Messages offering help or services, including rescue support, volunteer work, and offering or requesting donations of money, goods, or supplies
P9: Pets and animals	Messages about pets and animals, e.g., lost or found pets, animals in need of rescue, or updates on the condition and safety of animals
P10: Other	Messages that do not fit into the above categories

2. Data annotation tool GALLOC

We provide a brief explanation for using GALLOC below. More details about GALLOC can be found in the user manual at:

https://github.com/geoai-lab/GALLOC/blob/master/User_Manual_GALLOC.pdf

For a tweet like in the screenshot below, we will perform the data annotation in the following four steps to annotate the location description text (1), assign a category (2), find the spatial footprint (3), and set the message topic (4).

For choosing the geometry type of spatial footprints, here is a reference table.

Location category	Geometry type
C1: Door number addresses	Point
C2: Roads and highways	Polyline
C3: Road and highway segments	Polyline
C4: Highway exits	Point
C5: Intersections	Point
C6: Natural features	• Polygon for areal features (e.g., natural lakes, mountains); • Polyline for linear features (e.g., rivers); • Point for small features (e.g., mountain peaks);
C7: Other human-made features	• Polygon for areal features (e.g., reservoirs, parks); • Polyline for linear features (e.g., canals); • Point for small human-made features (e.g., restaurants);
C8: Admin units and neighborhoods	Polygon
C9: Multiple areas	Polygon

In general, more detailed spatial footprints (e.g., polygons and polylines) are preferred over simple point-based footprints for larger geographic features (e.g., states, cities, rivers, and parks). Also, human annotators are recommended to first use the Nominatim map service to find more

detailed spatial footprints. If Nominatim can't return a result, a rough drawing (e.g., a rectangle or a polygon) or a point representation is acceptable.

Two map services are provided to help automatically create spatial footprint:
- Nominatim: can return points, polylines, and polygons. Note that since there can be different geographic features with the same name, it prioritizes features based on its algorithm, often giving preference to those within the current view extent. A useful trick is to first zoom to the geographic area where you suspect the feature is, and then do the search.

- Google Maps: returns points only. Google Maps are especially suitable for finding door number addresses and point-based geographic features. Google Maps is not prioritizing the current view extent and it just returns the result of the best match.

3. Example-based specific guidelines

In the following, we provide some specific guidelines based on real examples to show how data should be annotated. In the examples, characters in green are correct annotations and characters in red are incorrect annotations.

(1) Annotate a location description as a whole piece not separate entities
- *This is at Periwinkle Way & Casa Ybel Road. #hurricaneIan*
- *This is at Periwinkle Way & Casa Ybel Road. #hurricaneIan*
- *#Floridians: if anyone is thinking of traveling on I-75 between Port Charlotte and Venice, please note I-75 is closed both ways #HurricaneIan*
- *#Floridians: if anyone is thinking of traveling on I-75 between Port Charlotte and Venice, please note I-75 is closed both ways #HurricaneIan*

(2) Do not annotate punctuations and prepositions that do not belong to a location description; Do annotate punctuations that are part of a location description (e.g., "St.")
- *#HurricaneHarvey family needs rescuing at 1280 Grant St. in Cypress, Texas 77249 2 elderly, one is 90 not steady in her feet*
- *#Harvey LIVE from San Antonio, TX. Fatal car accident at Ingram Rd., Strong winds.*
- *175+ roads flooded so far in #Houston via @houstontranstar*
- *#PortAransas was hit hard by #HurricaneHarvey.*

(3) Annotate direction words only when they are part of a location description
- *Power crews have just cleared a tree on Highway 39 South and that section is now fully reopened. #HurricaneIan*
- *A big high school football matchup in South Florida is being delayed due to #HurricaneIan*
- *The wind keeps shifting east and south...*
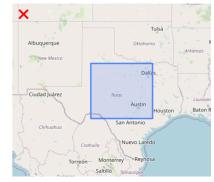- *is the San Jacinto bridge, north of Humble, TX open? HurricaneHarvey Harvey HarveyFloo*

(4) Annotate city name and state name as a whole
- The aftermath of Hurricane Harvey, said, long road to recovery. Prayer needed. *Houston, Texas*. #harveyrelief

- FOXNEWS: WATCH: POTUS's full remarks at a church in *Pearland, Texas*, used as a distribution center.
- FOXNEWS: WATCH: POTUS's full remarks at a church in *Pearland*, *Texas*, used as a distribution center.

(5) When the Nominatim map service can't return an accurate spatial footprint, we can draw a rough polygon, rectangle, or polyline (for line-based features) to represent their spatial footprint.

- My son's school will be "adopting" a school in Texas to help with their recovery.

(6) For location descriptions that contain a place name and its address, annotate them as two separate descriptions.

- *If you need a place to escape #HurricaneHarvey, The Willie De Leon Civic Center: 300 E. Main St in Uvalde is open as a shelter*
- *Free Gumbo for #HurricaneHarvey victims at the M.O. Campbell Center on 1865 Aldine Bender.*

(7) People sometimes say place names with their types omitted (especially when referring to roads). Please annotate those place names.

- *Both Memorial and Allen Parkway flooded. #hurricaneharvey* (note that "Memorial" here refers to "Memorial Dr." in Houston)

(8) Do not annotate chain hotels (e.g., "Holiday Inn"), chain restaurants (e.g., "McDonald's"), chain groceries (e.g., "Kroger"), and other chain stores that we can't determine the specific store

(9) Do not annotate locations that are only part of a hashtag:

- *Let's do better! #TexasFreeze #HelpTexas*

(10) For informal location descriptions rather than official names, still annotate them. If we can find their spatial footprints, provide the footprint geometries (e.g., drawing the geometry); otherwise, indicate that they are "Uncertain Locations".

- #HHS will provide on-site medical care to people sheltering at #*Houston's convention center*
- *Joel Osteen's giant Houston church is not open to Hurricane Harvey victims*