Several datasets available at <http://crisisnlp.qcri.org>

* most useful might be CrisisMMD dataset (resource 5), which contains both text and images, and is labeled wrt informative/non-informative, and also several humanitarian categories. I have the dataset myself, in case you can’t download it (I had some complains about this from other students).

Several datasets available at <https://crisislex.org/data-collections.html> but most useful

* CrisisLexT6 is labeled wrt relevant (on topic) versus non-relevant (off-topic). We have used CrisisLexT6 frequently in our work, as it’s pretty large.
* CrisisLexT26 labeled with respect to informative/non-informative, source of information (including eyewitnesses) and also wrt some situational awareness categories. The dataset contains tweets in several languages, and it’s small overall.

Actionability dataset available from:

* <https://github.com/GateNLP/emina/tree/master/actionability>

Some data related to utility companies/media, crawled directly from the Twitter pages of the companies/media targeted. Potentially useful for sentiment analysis or information about power loss/recovery (I have a collaboration along these lines with the group who maintains this repository).

<https://github.com/DisasterMasters/TweetAnalysis>

Incident-related datasets (these events are not considered disasters)

<http://www.doc.gold.ac.uk/~cguck001/IncidentTweets/> (I have these datasets if we need to use them)