Twitter Analysis for Quick Flood Response

Inception Report

Twitter Analysis for Quick Flood Response

Inception Report

Date:
29 April 2016

Radboud Universiteit Nijmegen

TABLE OF CONTENTS

1. Introduction

1.1. Background

Citizens and communities around the world are affected by floods each year. The people affected increasingly share their observations and their needs through digital media. But to date, there is no organized analysis of this data, so that its value remains untapped. For example in Jakarta, when it floods, people share 900 tweets per minute about their situation, including large numbers of observations on water depths, impact and needs assessments and even observations of dike breaches. This happens around the world from Philippines (85.000 tweets total) to Pakistan (82.000 tweets) to South-Korea (50.000 tweets) to Detroit (20.000 tweets). And this happens every day for countless much smaller events such as in rural parts of Indonesia (e.g. Indramayu, 750 tweets last March). All this is extremely valuable information for risk identification using the on-the-ground perspective. But at present it is not analyzed and not used for flood response.

The related technical challenge is that the content produced on-the-ground is text-based, unstructured and comes in large numbers. Flood events are described by hundreds of thousands of observations online per day, while some of its most valuable information is hidden in only a few text lines. The raw data cannot be connected to the procedures and technology of water and disaster managers (Twitter itself expressed their concern in a meeting with FloodTags, about how their raw data is “too raw” to connect to water managers). It actually needs analysis and translation to flood managers’ practice to become valuable.

The Red Cross Red Crescent Climate Centre (RCRCCC) and the Philippine Red Cross (PRC) are interested in the information from the ground to identify where, when and what risks are observed on-the-ground. But they are unable to do so for its technical challenge and lack of tools.

In this project we propose to develop tools that will enable the harvesting and practical use of social media information in disaster response, in direct collaboration with humanitarian response organizations who will be using the information. We propose to develop and sustain a minimum viable* webservice that collects data from Twitter, uses co-developed filters and enrichments for analysis, allows for improvement by academics through an open-innovation approach and is configurable by end-users for different use-cases (through an online browser interface, no software will have to be installed locally). The webservice will be hosted and maintained by FloodTags, a social enterprise. This project will be implemented together with the Red Cross Red Crescent Movement, but the resulting webservices will be available for all humanitarian organizations involved in disaster response.

** Minimum viable means that the webservice is operational, while improvements can be added to its core.*

The open source project FloodTags interprets online (social) media data for development issues. In The Philippines, FloodTags will set-up an online data service for (citizen-based) flood response of the Red Cross. The tool interprets tweets by using natural language processing and uses combinations with external data incl. GSMaP NRT (satellite mapping), MODISlance (MODIS based flood maps) and DEM

1.2. Partners and external relations

This project will be executed by a core team of organisations (the consortium) and relate to a number of external organisations and initiatives (external relations).

External relations are potential partners or end-users, who will be informed about the project and offered an API-key + access to the web interface. With this, they can at any time connect to the data, download enriched tweets and monitor events via their own software or via our front-end. **Some external relations can become partners in time.** They will, besides the API, get access to an SDK that helps them implement the software locally and develop new scripts for further filtering and enrichments.

Since recent enthusiasm from Indonesia, we reserved a separate overview for Indonesian external relations. If we have time left in the project (or are able to materialise additional resources), **we will configure the tool (partially) for Indonesia as well**, so that these possible partners and end-users can benefit from it too.

Hereunder an overview:

Partners	Description
The Red Cross Red Crescent Climate Centre (RCRCCC)	Global Counterpart for FloodTags IT in this project
The national societies of the Red Cross in The Philippines (PRC)	National Counterparts for FloodTags IT in this project
FloodTags	Developer FloodTags main architecture. Project lead.
Deltares	Knowledge partner for hydrology and hydraulics. Parallel R&D project: Building reliable flood extent maps on the basis of Twitter (on Dutch subsidy)
Radboud University	Knowledge partner for natural language processing. Parallel R&D project: Event Information Extraction using Social Media (PhD Promotion)
VU University Amsterdam	Knowledge partner for damage and response. Parallel R&D project: Assessing Flood Hazard and Impact using Social Media (PhD Promotion)

External relations Philippines	Description
Pagasa	National Meteorological Office. Possible science partner. PRC obtains meteo data from Pagasa in case of emergencies.
Project NOAH	Project for the Nationwide Operational Assessment of Hazard. Currently being transferred to Pagasa.
University of Philippines	Possible science partner. Jurjen has had contact in the past with Prof. Rhodora Gonzalez. Recontact through Dinand Alkema of ITC.
Netherlands Red Cross (NRC)	Reached out to align programs, achieve mutual goals. NRC has multiple programs in The Philippines.
Quiapo Church	Organises flood response from their catholic church. Possible end-user. Jurjen has had contact in the past with their rector Msgr. Jose Clemente Ignacio. Recontact through Dinand Alkema of ITC.
Other possible partners	Including Department of the interior, Office of Civil Defense and others. Will be followed up in coming months.

External relations Indonesia	Description
UN Global Pulse	Offered assistance for Indonesian configuration. Has focus on empowerment of local organisations.
The national society of the Red Cross in Indonesia (PMI)	The Red Cross in Indonesia. Are interested in a configuration for Indonesia.
National Disaster Management Organisation Indonesia (BNPB)	National disaster management organisation of Indonesia. Partners in the project worked with BNPB before. BNPB is looking for similar system. Opened contact (via Noel Pitoy).
Jakarta Flood Relief (JFRD)	NGO assisting a number of RW's during floods in Jakarta. Very local, ground-up organisation. Showed interest to use FloodTags. Are willing to annotate content.
Data Science Indonesia	Organises Bootcamp for young professionals. Showed interested to use FloodTags as case for two capstone projects. Are willing to annotate content and improve filters.

Other external relations	Description
ActivityInfo	ActivityInfo (AI) is an official UN volunteer reporting tool, that collects and analyses data for humanitarian organisations during crises. FloodTags and ActivityInfo are looking for funds to connect both products, so that humanitarian responders of the UN can complement reporting with real-time Twitter data. This will support decision making on re-allocation of funds and resources (during operations).

More external relations may be added during the project.

2. Approach and Workplan

2.1. Project phasing

In this project we will use a Rational Unified Project approach (RUP) which consists of the phases Acquisition, Definition, Elaboration, Production and Transition. To comply to Worldbank definitions, we renamed “Acquisition” to become “Inception”, which results in below project phasing:

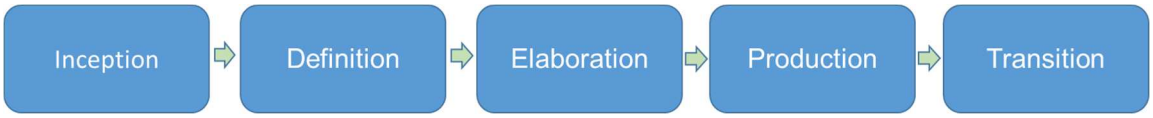

Figure 1: Rational Unified Processing (RUP)

In each of the phases, we discuss five domains:

- Societal: What is the societal problem and is it being solved?
- Functional: What does the software do, to solve the problem
- Technical: How does the software work, to solve the problem
- Graphical: What does the software look like
- Content: What does the software contain in terms of data

In RUP this looks like this:

	Inception →	**Definition** →	**Elaboration** →	**Production** →	**Transition**
Societal					
Functional					
Technical					
Graphical					
Content					

Within each phase there will be iterations based on input from project team and stakeholders. The iterations in the phase “Production” will be guided by Agile Development Principles. Per phase:

- Inception: Problem description translating into the scope of the project. The scope is divided into societal, functional, technical, graphical and content scope.
- Definition: Definition of the specific requirements to the software. Set-up of “Acceptance Criteria” to the project.
- Elaboration: The requirements are elaborated in a design of the solution, again divided over societal, functional, technical, graphical and content. The acceptance criteria are translated into a test plan.
- Production: Realisation of the software on the basis of the requirements. It will be component based and developed using agile software development principles.
- Test and Transition: The tested software is transferred to the production environment, users are trained, start of regular maintenance.

Just for illustration, hereunder an example RUP matrix for this project:

	Inception →	Definition →	Elaboration →	Production →	Transition
Societal	Rapid response saves lives	Rapid assessment with social media	-	-	SLA on rapid assessment SAAS
Functional	Flood overview, observations	Show tweets, show extent	Map, table, clickable	Sprint and test functions	-
Technical	Web-IT	Standards, protocols	Collect, enrich, store, give out	Sprint and test technical	-
Graphical	Easy to use, responsive	Material design, Wordpress	UX design	Webdevelop-ment and test	-
Content	observations, elevation, hydro	Twitter, DEM, Radar	Tweets/ s, DEM raster X, Y	Implement	SLA external data/ scripts

2.2. Workplan

Hereunder the main phases and activities with their durations. Each activity is delivered to be approved by either the Project leader, the Counterpart and/ or the Client. The main deliverables for the Client are Inception Report and Final Report.

Phase	Activities/ Deliverables	By whom	To be approved by
Inception (2 months)			
	Problem description	Counterpart & project leader	Counterpart
	Societal, functional, technical graphical and content scope	Counterpart & project group	Counterpart and project leader
	Activity planning	Project leader	Counterpart and project group
	Inception report	Project leader	Counterpart and Client
Definition (2 weeks)			
	Societal requirements (what does the solution contain to solve the problem)	Counterpart & project leader	Counterpart and project leader
	Functional requirements (what does the software need to do)	Counterpart and project team	Counterpart
	Technical requirements (how is the software built)	Counterpart and IT-architect	Counterpart
	Graphical requirements (what does the front-end of the software look like)	UX designer	Counterpart
	Content requirements (what data does the software use)	Counterpart and project team	Counterpart
	Acceptance Criteria (Memo)	Counterpart and project leader	Counterpart
Elaboration (2 weeks)			
	Societal design (description of the SLA)	Project leader	Counterpart
	Functional design (for each component of the software i.e. relevancer, flood extent, flood impact, flood alert)	Counterpart, component leaders and project leader	Project leader
	Technical design for each component	IT-architect, component leaders and project leader	Project leader
	Graphical design for each component and the software as a whole (incl. UX design)	UX designer	Counterpart and Project leader
	Content design for each component	Component leaders	Project leader
	Click demo (for acceptance of functional and graphical design)	UX designer	Counterpart
Production (3 months)			
	Production in Sprints (per component and the integrated software as a whole)	Project team	Counterpart
	Testing of the end-product (against “Acceptance Criteria)	Counterpart and project leader	Counterpart
	Refinement	Project team	Counterpart
	Test report	Project leader	Counterpart
	Acceptance of the project (societal, functional, technical, graphical, content)	Counterpart	Counterpart
Transition (3 months)			
	Signing of the SLA	Counterpart and FloodTags	
	Start of regular maintenance	FloodTags	
	Final report	Jurjen Wagemaker	Client

* Approval of all deliverables marked with ‘Counterpart’ will be approved by The Philippine Red Cross.

The runtime of the project, divided over five phases, is set at 8 months with deadline 31 October. The complete work schedule distributed over the next eight months can be found in Annex 2.

2.3. Organisation and staffing

Hereunder the final structure and staffing of the project organisation, as well as the internal communication and travelling arrangements.

Core team

Project leader	Jurjen Wagemaker
Red Cross Coordinator	Erin Coughlan
Quality control	Hessel Rosbergen

Project group

Coordinator (Engagement)	Donna Lagdameo
IT-Architect	Ron Boortman
Advisor/ Programmer (Flood Extent)	Arnejan van Loenen
Advisor/ Programmer (User Analysis)	Martijn Wieriks
Researcher/ Programmer (Flood Extent)	Dirk Eilander, Patricia Trambauer
Researcher/ Programmer (Event Detection)	Abbey Waldron
Advisor (Flood Impact)	Jeroen Aerts
Researcher/ Programmer (Flood Impact)	Jens de Bruijn
Researcher/ Programmer (Text Mining)	Ali Hürriyetoğlu, Erkan Basar
Local Project members	Team of the national society of the Red Cross

Task assignments and level of effort can be found in Annex 1.

External suppliers

To fit the functional possibilities to the end-users, we may hire a professional User Experience (UX-) Designer. This will be decided after the Definition phase when we have a clear view of the acceptance criteria. A second external supplier is Amazon, where we will set-up a dedicated webserver for this project.

Project meetings

Kick-off and Wrap-Up meeting	Project group
Definition and Transition meeting on location	Project leader + Coordinator + Local project members
Monthly Progress meeting	Core team
Other meetings	On need to meet basis

Travelling

Project members are based in The Netherlands, USA and The Philippines. To optimise communication while not spending too much budget on travelling:

- All project group and core team meetings will be done through skype
- There will be one visit by project leader and coordinator on location, probably at the end of Production (around the Acceptancy Test)
- Other meetings can be either through skype or in person

2.4. Task assignment of the PRC

The Philippine Red Cross is the main counterparts for this project. They are supported by the RCRCCC.

In the inception phase (two months)

In the inception phase, the project team and PRC define the **problem statement** and detail the **scope of the solution**. The PRC involvement needed includes

- Elaboration of the societal problem (what is the problem that needs fixing, e.g. the time-lapse to flood response is too large) and the scope of the solution (what kind of solutions are considered, e.g. real-time impact assessment using citizen observations). *Setting the societal scope is typically done by the management of the organisation.*
- Elaboration of the functional scope, where we describe what the software would need to do (e.g. create overview of reliable tweets on certain location, allow new queries to the database etc.). *This part is typically done by the scientific partners in the project team in communication with the head of disaster management and disaster experts in the organisation (and approved by management).*

- In the technical scope it is described how the solution technically should work (e.g. protocols to connect to existing Client software). *This is typically done by the IT architect in the project team in communication with the head of ICT in the organisation (and approved by management).*
- The graphical scope comprises what the solution graphically should look like (e.g. an online Dashboard interface with map) . *This is typically done by a graphical designer in communication with the head of disaster management and other disaster experts in the organisation (and approved by management).*
- Content scope: Conclusively it is described what is the content of the solution (e.g. what topics/ key words need monitoring and for which regions must it be applicable). *This is typically done by the scientific partners of the project team in communication with the head of disaster management and/ or management.*

The scope is described in chapter 3 of the inception report. The inception report (with the scope) will be approved by the PRC.

In the definition phase (two weeks)

In the definition phase, the project team and PRC define the requirements to the solution and set-up the **Acceptance Criteria** to which the software will be evaluated at the end of production (per domain i.e. societal, functional, technical, graphical and content). The same organisational involvement is needed from the PRC as in the Inception Phase. When ready, the Acceptance Criteria are approved by the management of the PRC. *This phase requires involvement of PRC management, head of disaster management, head of IT and head of purchasing department*

In the elaboration phase (two weeks)

No involvement of PRC needed.

In the production phase (three months)

In the realisation of the software, user input is required to:

- **Annotate** clusters of observations (on the basis of which custom classifiers can be developed). This will be done using a separate user-friendly interface. The expected amount of time input for PRC will be 5-10 days.
- **Test and accept** (or reject) individual sprint results and the overall result. The overall result is tested against the "Acceptance Criteria" from phase 2.

This phase requires involvement of management, head of disaster management and head of IT.

In the transition phase (one-two months)

The tested software is transferred to the production environment. If applicable, an SLA is signed and becomes effective, end-users are trained and **regular maintenance** is started. External relations are contacted and, if opportune, offered an API-key so that they can benefit from the same data. *This phase requires the involvement of PRC management, head of disaster management, head of IT and head of purchasing.*

	Inception ⇨	Definition ⇨	Elaboration ⇨	Production ⇨	Transition
Societal (PRC management)	Problem and scope of solution	Addon specific requirements		Approval of the SLA conditions	Manage SLA
Functional (disaster manager)	Addon to scope of the solution	Addon specific requirements		Acceptance tests	
Technical (IT manager)	Addon to scope of the solution	Addon specific requirements		Annotation & Acceptance tests	
Graphical (disaster manager)	Addon to scope of the solution	Addon specific requirements		Acceptance tests	
Content (disaster manager)	Scope of the data to be used	Distribution of metadata		Distribution of the actual data	SLA external data/ scripts

Figure 2: The involvement of the PRC per phase

2.5. Task assignment of the Scientific Partners

The scientific partners are Deltares, Radboud University Nijmegen and VU University Amsterdam. Hereunder their involvement in the project per phase:

In the inception phase (one month)

In this phase, the scientific partners show **what they can offer in the project**. This is divided into

- Societal scope: How is the solution of partner offered (e.g. as open source script, SAAS etc.)
- Functional scope: What contributions can partner make to solve the problem
- Technical scope: How would that solution technically work?
- Graphical scope: What are the suggestions for the graphical interface to the solution?
- Content scope: What content would the contributions need to run?

The contributions of the scientific partners for inception are integrated in this reports' chapter 3. On this basis the counterpart will define the project and set the acceptance criteria.

In the definition phase (two weeks)

The counterpart describes the Acceptance Criteria for a successful project. This will be done in communication with the scientific partners (to make sure scientific partners can fulfill expectations, or whether there is need for downsizing). The involvement of the scientific partner is relatively small.

In the elaboration phase (two weeks)

On the basis of the requirements from definition phase, the scientific partner **designs their part of the project**. The design will include all five domains:

- Societal design: E.g. what conditions would apply for their part of a SAAS
- Functional design: E.g. a user can see what hospitals are likely affected
- Technical design: E.g. python script as part of the FloodTags pipeline
- Graphical design; E.g. as table in material design
- Content: E.g. using Twitter feed with queries 'damage OR flood'

In the production phase (three months)

In this phase the software is **realized according to Agile Development Principles**. There will be sprints of two weeks that are concluded with a user test and acceptance. Not all partners will be working for three months consecutively and **partners make their own planning for the delivery of the software**, in communication with the IT architect (Ron Boortman). All code is shared via a bitbucket account.

Note that for practical reasons, some partners may choose to already start with some elaboration and production, before counterpart has officially approved the final Acceptance Criteria. There is a risk though, that work needs to be redone if the final requirements are different than expected.

In the transition phase (one-two months)

The tested software is transferred to the production environment. The SLA is signed and becomes effective, end-users are trained and regular maintenance is started. The role of scientific partner depends on the **arrangements made in an SLA**.

3. Phase 1 in RUP: Inception

Hereunder the following text.

As mentioned in section 2.1, the project is divided over five domains. They are:

- Societal: What problem are we solving. Is the problem being solved
- Functional: What does the software do, to solve the problem
- Technical: How does the software work, to solve the problem
- Graphical: What does the software look like, visually
- Content: What does the software contain in terms of data

Hereunder we describe the first step (inception) of each of these domains. As mentioned under section 2.4, the results of the inception are described in the inception report and approved by the PRC management.

3.1. Societal scope

Citizens and communities in the Philippines are affected by floods each year. The people involved increasingly share their observations and their needs through online media. The Philippine Red Cross is interested to analyze these observations and needs for better flood response, but they are unable to do so for its technical challenge and lack of tools. The solution that we are looking for fulfills at least the next requirements:

- The software will be used by The Philippines Red Cross (PRC) to better monitor and respond to floods in the Philippines and is compatible with other software in their respective operation rooms.
- The software will be operated by the disaster management team of the PRC and managed by the IT department of the PRC.
- Sustaining the software must be cost-efficient for the budget of a) National Society of the PRC possibly supported by other Red Cross entities such as APZ, NRC or RCRCCC b) via DREF appeals per disaster in case of exceptional events (as official tool under the federation) or c) on other budgets.
- The software can be used by new local partners or end-users in the Philippines (organisations can connect easily to the software) and is scalable to other countries where the Red Cross is active.

Societal Definition (preview)

In the next phase, the above scope will be translate into requirements to the software. In below table we show a preview of the requirements per item:

Requirements of the software (preview to Definition Phase)	
Societal Scope	**Societal Requirements (options, draft)**
The software will be used by The Philippines Red Cross (PRC) to better monitor and respond to floods in the Philippines and compatible with other software in their respective operation rooms.	The software can only be used if it is explicitly clear about the reliability of the data. Any unclarity about reliability will hamper efficient response. Further, the software must support: 1) During "exceptional floods": Support the DREF request process. The software provides quick, objective and representative analysis of key indicators used in a DREF application. 2) During "normal floods": Support the National Society Response Preparation. The software provides quick, objective and representative analysis of key indicators used to: - Begin negotiating contracts to purchase supplies - Begin transporting supplies to the region 3) During "any floods": Support the National Society Response. The software provides quick, objective and representative analysis of key indicators used to - Analyse potential cascading effects - Warn communities downstream because an upstream community was just flooded - Promote hygiene behaviours before diseases spread - Suggest the best evacuation locations so that people go to the right place from the beginning and do not end up somewhere that is less than ideal
The software will be operated by the disaster management team of PRC and managed by the IT department of the PRC	The software must be easy to be used and accompanied by a manual and/ or training of the software.
Sustaining the software must be cost-efficient for the budget of a) National Society of the PRC possibly supported by other Red Cross entities such as APZ, NRC or RCRCCC b) via DREF appeals per disaster in case of exceptional events (as official tool under the federation) or c) on other budgets.	The solution is open software on GitHub so that it can be hosted and maintained by PRC itself or acquired as a service (offered by an external organisation). There are three universities and one knowledge institute who invested in-kind to the development of the software. The company FloodTags invested in-kind and in cash. More leverage is needed to scale the software.
The software can be used by new local partners or end-users in the Philippines (organisations can connect easily to the software) and is scalable to other countries where the Red Cross is active	The way the tool will be set-up needs to: - Stimulate and facilitate local partners to connect - Stimulate and facilitate local beneficiaries to take advantage - Stimulate and facilitate decision makers to connect - Motivate gender awareness and women involvement

3.2. Functional scope

The functionalities of the tools will comprise:

- Real-time national overview of 1. the most important (in terms of relevance and reliability) social media messages on floods and 2. the most important users in the network

- Alert system that sends out a warning (to PRC) when thresholds in incoming (relevant and reliable) social media messages are exceeded.
- Real-time analysis of the messages including:
    - Overview of the topical content of the message(s)
    - Overview of the likely/ potential flood impact following the social media message(s)
    - Overview of the likely damage or damage components following the social media message(s)
- Support active engagement with Twitter users
- The software is easy to connect to by new users

More details of the societal requirements are explained in below table:

Functional definition (preview)

Requirements of the software (preview to Definition Phase)	
Functional Scope	**Functional Requirements (options, draft)**
Real-time national overview of - the most important (in terms of relevance and reliability) social media messages on floods. - The most important users in the network	A user can see relevant and reliable tweets: - on a zoomable map containing locations of the tweets - with graph containing the amounts of tweets - in a table showing the text of the most important tweets - with possible selection of time and search keyword - with possible export of data to a csv or pdf - with expandable query to include specific (new) hashtags
Alert system that sends out a warning (to PRC) when thresholds in incoming (relevant and reliable) social media messages are exceeded.	A user receives a warning when - A flood has been detected with high certainty (high precision) - A flood has been detected with low certainty (high recall)
Real-time analysis of the messages: Overview of the topical content of the message(s)	A user is presented an overview of the content of the messages, by: - Classification of the tweets into predefined classes - A real-time word (combination) count - With selection of time and region
Real-time Analysis of the messages: Overview of the likely/ potential flood impact following the social media message(s)	A user is presented an overview of the possible flooded area, on the basis of the incoming tweets about a flood event
Real-time Analysis of the messages: Overview of flood damage and response activities	A user is presented an overview of the damage and response activities ongoing, on the basis of the incoming tweets about a flood event.
Support active engagement with Twitter users	Top list of users that may have important additional information on a flood or can help send out important messages on the topic.
The software is easy to connect to by new users	After registration and acceptance by FloodTags, new users can connect to and personalise data collection, data enrichment and data distribution for their own use.

3.3. Technical scope

We will use and further develop the existing FloodTags IT. The technical scope consists of four parts:

- Retrieval and saving of data: FloodTags monitors Twitter using a query. For the Philippines we currently monitor: baha OR bumabaha OR apaw OR pagbaha OR pag-apaw OR guho OR #landoph OR #ifrc. All tweets that come in are saved into two separate databases: 1) A long-term back-up storage where tweets will go directly into the back-up storage unmodified and 2) A query database which is optimised for resolving queries in real-time. This database contains a slimmed down version of the tweet together with applied enrichments.
- Enrichment of data: The enrichment process of incoming data is separated from the retrieval. This way, enrichments can take longer processing time without blocking the throughput of the system. Also, enrichment scripts can be hosted by third parties, which help FloodTags with saving computing power for its core business and helps third parties to protect their source code, if needed. The enrichment is done in a sandboxed environment to prevent modification of other processes. We distinguish tweet objects and event objects. Event objects are groups of tweets defined by query and time in which a threshold (for that query) was/ is exceeded.
- Connection to the API: The FloodTags IT offers data through an API that can be accessed via a webbrowser and command line. More information on the documentation and explanation of all the supported functions and parameters can be found on the webpage of the API.
- Connection to the SDK: This section is not finished yet and will be completed as the project develops. More information on how to get involved, the used standards and protocols, can be found in the online repository of this project: https://github.com/floodtags/GFDRR

Hereunder the software packages (and versions) used on the project:

Server side	Data-Exchange	Client side
LINUX • Yum package management (Amazon distribution) Python • 3.4 • Anaconda Node JS • 0.10.24 • ECMAScript 5 ElasticSearch • 1.4.5 MongoDB • 2.6.11	• JSON format • GeoJSON locations	HTML SCSS • 3.4.15 Jade JavaScript • ECMAScript 5 Router • Express 4 (Node JS)

Technical definition of the enrichments (preview)

Hereunder the further explanation of the technical scope per enrichment:

Technical scope of the enrichments (preview to Definition Phase, options, draft)	
Classifier on the basis of Natural Language Processing	On the basis of initial user input (annotating clusters of content in several rounds of user input), classifiers are built that can recognise relevant content in big datasets, and separate irrelevant content. This component allows to: - Support the process of clustering and annotation via a web-application - Create classifiers on the basis of clustering and annotation - Enrich incoming tweets with classes (using the classifier) - Create topical overview on the basis of classes and (combinations of) incoming words. *Input: Data (tweets) and user response (options, annotation)* *Output: Clustering and classification script, web application, classifiers, topical overview*
Geocoding	The tweets within an event will be NLP filtered and geocoded using Postagging and Nomatim. Possibly additional features will be used from the tweets, such as timezone ID. Alternatively, we will consider to use existing geolists from the Counterpart and have them adjusted by local experts. The administrative boundaries to be used will come most likely from COD (common operational datasets, see also https://www.humanitarianresponse.info/en/applications/data) *Input: Data (tweets, library and COD) + Postagging script* *Output: Script that enriches tweet with a geocode*
Filtering on the basis of DEM, GSMap and MODIS	The geocoded tweets are compared to three external databases for their hydrologic fit. This means that we compare the observations with the elevation model (flood observation less likely to be true on higher grounds), with precipitation data (flood observations likelihood to be true after extreme rains) and flood hazard areas (flood observations likely to be true in previously flooded regions). The external data consists of a corrected DEM, GSMap near-real-time precipitation and MODISlance flood maps. If possible, near-real-time river discharges will be added as an additional external dataset. The results will be added as enrichments to the individual objects. *Input: Data (tweets, DEM, GSMap NRT, MODISlance)* *Output: Script that enriches tweet with probabilities of being true positives*
Threshold setting	An open source script that will take a number of tweets in up to ten different streams in a to-be-agreed time interval (10min-1hour approx.) and return whether a threshold has been crossed. Per query there will be two separate alerting thresholds calculated, one "normal" 6 sigma equivalent, and one with higher efficiency and more false positives at 3 sigma equivalent. The thresholds will be calculated per hour of the day. In cases where there is insufficient data a default method will be used. The thresholds for all of the streams will be provided in GeoJSON format or similar. *Input: 10 queries* *Output: Two thresholds per query*
Potential Flood Extent	Likely/ potential flood extent is delineated by overlaying the observations within a detected event (where event = thresholds exceeded for a query) and the flood prone areas as derived for the filters (from DEM, MODIS, etc.) or from the Aqueduct Global Flood Analyzer (http://floods.wri.org/#/). *Input: Data (tweets, MODISlance)* *Output: Script that enriches events with a maximum potential flood extent*
Flood Impact and Response	Determine likely damage or damage components on the basis of - Maximum Damage = Potential Flood Extent x stage-damage curves x maximum damage and/ or - Damage components listed on the basis of mentions of key infrastructure (schools,

	hospitals, power plants) in (estimated) flood extent area. - Response activities listed on the basis of mentions of response activities (aid, assistance, NGO, support) - Possibly: Damage factor distribution on basis of Bayesian approach of GPS located tweets in flood extent, using ratio flood tweets/ total tweets, tweet classes, sentiment and pattern of posts, calibrated with historic damage *Input: Data (tweets and events). It would help to have a previous event well-described. Satellite imagery available a.o. at http://floods.unosat.org/geoportal/catalog/main/home.page* *Output: Open source scripts that produces real-time damage and response overviews on the basis of Twitter information and Flood Extent*
Social Network Analysis	User analysis will be used to: - Filter out spammers and low-credibility users: Enrich tweet with user information (high – low – no credibility) - Enrich tweets with qualification of influence and amplifier* - Create top list of users that can best be engaged with to 1) Validate flood information 2) Reach many people with a message *Top influencers: the people and companies that carry weight in flood and disaster management. Their characteristics are: - large amount of followers - high engagement in social media through retweets, social media shares - high amount of mentions/shares from other users Top amplifier characteristics; Power users who are 86% more likely to send tweets. Twice as likely to follow more than 20 brands. 54% more likely to use twitter on their mobile. And most likely to retweet or share your status updates or content. Their characteristics are: - high number of tweets - follows a high number of users (big accounts) - retweets/shares a lot *Input: Data (tweets)* *Output: Script that real-time enriches tweets with credibility of user + script that creates real-time user list with influence and amplifier ranking*

3.4. Graphical scope

There are several options for the graphical scope in the front-end. In any case the interface must reflect the considerations in the societal scope (easy to use, stimulates collaboration) and meet the functional requirements. As first step we will create a simple dummy front-end in Django and continue after the finalisation of the criteria (in the Definition phase).

3.5. Content Scope

The content is limited by the boundaries of the country The Philippines. The table below presents an overview of the content of the software:

Source data	
Twitter	On basis of text query at least including baha OR bumabaha OR apaw OR pagbaha OR pag-apaw OR guho. To be elaborated in Definition Phase.
Spatial data	DEM, GSMap NRT, MODISlance, COD and others. To be elaborated during Definition.
Output data	
Flood Extent, Damage, Overviews	Txt, tables, raster maps, shapefiles. To be elaborated in Definition.

ANNEX 1: TEAM COMPOSITION, TASK ASSIGNMENTS & LEVEL OF EFFORT

Name of Staff & Firm associated with	Area of Expertise Relevant to Assignment	Designation for this Assignment	Assigned Tasks or Deliverables	Location (3)	Number of Days
Jurjen Wagemaker (Employee, FloodTags)	Project management, citizen observations	Project leader	Project lead	The Netherlands (International)	18
Erin Coughlan (Employee, Red Cross Red Crescent Climate Centre)	Climate and disaster management	Red Cross Coordinator	- Coordinate between RCRCCC and the Red Cross on location - Deliver disaster management context - Assist transition of software to local Red Cross	United States (International)	3
Hessel Rosbergen (Sub-Consultant, FloodTags)	Project management of large ICT projects	Quality control	- Assist to project management and ICT set-up - Reviewer of key outputs	The Netherlands (International)	2
Donna Lagdameo (Employee, Red Cross Red Crescent Climate Centre)	Stakeholder engagement	Engagement Coordinator	- Coordinate between RCRCCC and the Red Cross on location - Deliver disaster management context for functional designs - Assist transition of software to local Red Cross	The Philippines (International)	5
Ron Boortman (Employee, FloodTags)	ICT Development	Architect/ Programmer	- Back-end design and realisation - Front-end realisation	The Netherlands (International)	42.5
Arnejan van Loenen (Employee, Deltares)	Flood information systems	Advisor/ Programmer	Design and realisation of flood extent mapping and validation	The Netherlands (International)	7.5
Jeroen Aerts (Employee, VU University)	Flood impact assessments	Advisor	Design and realisation of flood impact mapping	The Netherlands (International)	5
In-kind contributions					
Radboud University: Ali Hürriyetoğlu	Data Analysis	Researcher/ Programmer	Design and realisation of text mining scripts	The Netherlands (International)	30
Erkan Basar (employee, FloodTags)	ICT Development	Researcher/ Programmer	- Design and realisation of text mining scripts - Front-end development	The Netherlands (International)	5
SEPAM Consultants: Martijn Wieriks	User Analysis	Advisor/ Programmer	Design and realisation of user analysis scripts	Indonesia (International)	15
ApicturesAnalysis: Abbey Waldron	Event Detection	Researcher/ Programmer	Design and realisation of event detection	The Netherlands (International)	10
Deltares: Dirk Eilander + Patricia Trambauer	Flood information systems	Researcher/ programmer	Design and realisation of flood extent mapping and validation	The Netherlands (International)	30
VU University: Jens de Bruijn	Flood impact assessments	Researcher/ programmer	Design and realisation of flood impact mapping	The Netherlands (International)	30
Local staff (Employees of local Red Cross)	Local situation experts	Project members	- Description of Local Context - Data delivery, Training of classifiers - Testing of the software	Manilla, Philippines	30

ANNEX 2: WORK SCHEDULE + DETAILED ACTIVITY PLANNING

WORK SCHEDULE

No	Phase	Activities and *Deliverables*	Month																	
			March		April		May		June		July		August		September		October	
1	Inception	Problem description																
2		Societal, functional, technical graphical and content scope																
3		Activity planning																
4		*Inception report*																
5	Definition	Societal requirements																
6		Functional requirements																
7		Technical requirements																
8		Graphical requirements																
9		Content requirements																
10		*Acceptance Criteria (Memo)*																
11	Elaboration	Societal design (description of the SLA)																
12		Functional design (for each component)																
13		Technical design (for each component)																
14		Graphical design (incl. UX design)																
15		Content design for each component																
16		*Click demo and Acceptance*																
17	Production	Production in Sprints																
18		Testing of the Sprint deliveries																
19		Acceptance test																
20		Visit of TL to Manilla																
21		Refinement																
22		*Test report (Memo)*																
23	Transition	Signing of the SLA																
24		Start of regular maintenance																
25		*Final report*																

DETAILED ACTIVITY PLANNING

(ALSO AVAILABLE SEPARATELY)

No	Phase	Activities and *Deliverables*	Month															
			March		April		May		June		July		August		September		October	
1	Inception	Problem description																
2		Societal, functional, technical graphical and content scope																
3		Activity planning																
4		*Inception report*																
	FloodTags	Leads Inception Phase. Delivers Inception report			1-May													
	All partners	Contributions to the inception report (scope and activity planning)			15-Apr													
	RCRCCO PRC	Elaborate and agree to scope (inception)			1-May													
5	Definition	Societal requirements																
6		Functional requirements																
7		Technical requirements																
8		Graphical requirements																
9		Content requirements																
10		*Acceptance Criteria (Memo)*																
	Floodtags	Leads Definition. Delivers Acceptance Criteria						1-Jun										
	All partners	Contributions to the Acceptance Criteria (can partners deliver this)					15-May											
	RCRCCO PRC	Interviews with PRC, preparation of acceptance criteria					15-May											
	PRC management	Agree to acceptance criteria						1-Jun										
11	Elaboration	Societal design (description of the SLA)																
12		Functional design (for each component)																
13		Technical design (for each component)																
14		Graphical design (incl. UX design)																
15		Content design for each component																
16		*Click demo and Acceptance*																
	Floodtags	Click demo						1-Jun										
	All partners	Contributions to the click demo						1-Jun										
17	Production	Production in Sprints																
18		Testing of the Sprint deliveries																
19		Acceptance test																
20		Visit of TL to Manilla																
21		Refinement																
22		*Test report (Memo)*																
	FloodTags	API with realtime query data (RB, needs basic query)			1-May													
	Ron Bantman	First version front-end dummy data (EB)			1-May													
	Jurjen Wagemaker	Tweets from API are geocoded - first version (RB)			1-May													
	Erkan Basar	First version front-end real data (EB, needs API)					15-May											
	Abbey Waldon	Interface supporting creation of a high precision classifier (EB)						1-Jun										
	Martijn Willemsz	Interface supporting annotation for geocoding (EB)						1-Jun										
		Interface supporting creation of a high recall classifier (EB)						1-Jun										
		High precision classifier implemented (RB, needs classifier)								1-Jul								
		Improved geocoding implemented (RB, needs annotated input)								1-Jul								
		High recall classifier is implemented (RB, needs classifier)								1-Jul								
		Thresholds are set for ten queries (AW, needs detail queries)								1-Jul								
		Events are identified in realtime (RB, needs thresholds)								1-Jul								
		Events pipeline, events can be enriched (RB)						1-Jun										
		Topical overview implemented (RB, needs script AH)									15-Jul							
		Tweets are filtered on basis DEM GSMap (RB, needs script PT)										1-Aug						
		Tweets are filtered on basis of SNA (RB, needs script MW)										1-Aug						
		Events are enriched with extent (RB, needs script PT)											1-Sep					
		Events are enriched with damage response (RB, needs script JdB)											1-Sep					
		Events are enriched with results from SNA (RB, needs script MW)											1-Sep					
		Final front-end is implemented (EB, needs full API)													1-Oct			
		SDK is functioning for new partners and end-users (RB)													1-Oct			
	Radboud University	Ambiguity analysis, filtering and clustering (AH, needs basic query)						1-Jun										
	Ali Hurriyetoglu	Evaluation strategy for annotation high recall classifier						1-Jun										
		Generation of classifiers (AH, needs annotated clusters)								1-Jul								
		Generation of script for topical overview (AH)									15-Jul							
	PRC	Use interface to annotate tweets for precise query							15-Jun									
		Use interface to annotate tweets for geocoding							15-Jun									
		Use interface to annotate tweets for recall classifier							15-Jun									
	Deltares	Prepare dummy results for FloodTags API (PT)						1-Jun										
		Configure enrichment on basis DEM, GSMap and MODIS for PHL (PT, needs API with geocoded flood data)								1-Jul								
		Develop script for flood extent on basis MODIS (PT, needs event pipeline to work)										1-Aug						
	VU Amsterdam	Prepare dummy results for FloodTags API (PT)						1-Jun										
		Develop script damage response (JdB, needs API with damage data)										1-Aug						
23	Transition	Signing of the SLA																
24		Start of regular maintenance																
25		*Final report*																
	FloodTags	Start of regular maintenance, final report																1-Nov
	PRC	Test and accept the software														1-Oct		
		Training of end-users																1-Nov