I REACT

IMPROVING RESILIENCE TO EMERGENCIES THROUGH

ADVANCED CYBER TECHNOLOGIES

Open Data and Historical Products Module

Deliverable ID	D3.3
Work Package Reference	WP3
Issue	1.00
Due Date of Deliverable	30/11/2017
Submission Date	10/11/2017
Dissemination Level[1]	PU
Lead Partner	Terranea
Contributors	-
Grant Agreement No	700256
Call ID	H2020-DRS-1-2015
Funding Scheme	Collaborative

European Commission

I-REACT is co-founded by the Horizon 2020 Framework Programme of the European Commission under grant agreement n. 700256

[1] **PU** = Public, **PP** = Restricted to other programme participants (including the Commission Services),
RE = Restricted to a group specified by the consortium (including the Commission Services),
CO = Confidential, only for members of the consortium (including the Commission Services)

Prepared by	Reviewed by	Approved by
G. Zeug	W. Stemberger	C. Rossi

Issue	Date	Description	Author(s)
0.01	13/10/2016	Provision of template	W. Stemberger
1.0	10/11/2017	Version 1.0	G. Zeug, S. Illium

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 PURPOSE OF THE DOCUMENT

The goal of this document is to provide a release note of the Disaster Event Tracker developed in WP3 /Task 3.7).

1.2 STRUCTURE OF THE DOCUMENT

The document is organized as in the following:

- **Chapter 1** is the introduction and description of the document itself;
- **Chapter 2** summarises the related tasks and operation;
- **Chapter 3** is the technical service description.

1.3 ACRONYMS LIST

API	Application Programming Interface
ADRC	Asian Disaster Reduction Center
CRED	Centre for Research on the Epidemiology of Disasters of the University of Louvain in Brussels (Belgium)
GPE	Geopolitical Entities
IFRC	International Federation of the Red Cross
OFDA	Office of US Foreign Disaster Assistance
REST	Representational stateless transfer
UNDP	United Nations Development Program
UN-ISDR	United Nations International Strategy for Disaster Risk Reduction
UN-OCHA	The United Nations Office for the Coordination of Humanitarian Affairs

1.4 REFERENCES

ID	Title	Revision	Date
[RD01]	D 2.1 Report on Multi Hazard Requirement Analysis	2.0	17/10/2016
[RD02]	D 2.3 Report on design of the Big Data Architecture, Linked Data & Semantic Structure	1.0	10.10.2016
[RD03]	D 2.7 Technical Requirements and System Architecture	1.0	09.01.2017

2 TASKS & OPERATION

Current disaster events will be historical in the future. A disaster event tracker was developed to record events and to update the historical event dataset, developed in T3.3 automatically.

Three different data sources for information about disaster events were identified:

- ReliefWeb (https://reliefweb.int) is a service provided by UN OCHA. It is a leading online source for reliable and timely information on global crises and disasters. Addressing the humanitarian aid community as main users ReliefWeb provides disaster and development related information, including reports, maps, infographics and videos from trusted sources. ReliefWeb offers all its content through an API including also information about disaster events.
- GLIDE (http://glidenumber.net): accessing disaster information can be a time consuming and laborious task. Not only is data scattered but frequently identification of the disaster can be confusing in countries with many disaster events. To address both of these issues, the Asian Disaster Reduction Center (ADRC) proposed a globally common unique ID code for disasters. This idea was shared and promoted by the Centre for Research on the Epidemiology of Disasters (CRED) of the University of Louvain in Brussels (Belgium), OCHA/ReliefWeb, OCHA/FSCC, ISDR, UNDP, WMO, IFRC, OFDA-USAID, FAO, La Red and the World Bank and was jointly launched as a new initiative "GLIDE". On the website new disaster events are registered and related event information is collected. Events can be queried but the website does not provide an API for data access. Though, the database can be downloaded.
- ECHO Flash is a service provided by the Emergency Response Coordination Centre of the European Commission (http://erccportal.jrc.ec.europa.eu/ECHO-Flash). The website publishes daily news about disaster and crisis events worldwide. The website is in html format. No API or other download interface exists.

The disaster event tracker was designed to visit the three above mentioned service websites and in extension many more, to gather recent natural disaster events in Europe. It is capable of filtering and reducing description texts, find geopolitical entities (GPE) and retrieve geo-coordinates, provided through the Google Geocoding-API in EPSG 3857. When processed, all gathered and filtered event objects are uploaded to a database by REST-API access. The event tracker should be executed automatically daily, to yield reliable results. Only basic web scraping techniques and the Python[1] programming language were used to make the script compatible with low and medium scale Linux server back-ends. The final API access has been built in accordance to the provided documentation[2].

[1] https://www.python.org/
[2] https://bitbucket.org/mobilesolutionsismb/ireact-rest-api

3 TECHNICAL STRUCTURE

In this section, the technical structure in reference to **Error! Reference source not found.** is presented.

3.1 WEB-SCRAPING

The Event-Trackers Web-Scraping part handles all information access points, stored as string URLs. Since each of the current implemented services has to be hailed by an different interface, a headless RSS-Feed reader (DailyFlash), a headless browser with included JS support (GlideDB) and regular API requests (ReliefWeb) were used for information retrieval.

3.2 ENTRY-SORTING &TEXT-PROCESSING

After the information retrieval, it needs to be sorted in accordance to the provided REST-API specifications for the final POST request. Hence the varying source types, each of the services needs to be treated in its unique way. There is currently no automatic way in determining the needed fields. This must be manually adjusted by a human operator – and checked for correct operation regularly every two weeks. A so called *PostObject* has been implemented to serve this task.

When all the information has been sorted to the corresponding fields, the description or comment can be filtered for useful information. “Usefulness” in this case was defined by: shortened sentences that hold literal or real numbers that are neither dates, years or times. Additionally, the descriptions / comments are searched for so called GPEs and added to the location field. The natural text-processing library *Spacy*[3] provides the background architecture for determining additional locations. For more basic string tasks, python regular expressions by the package *re* were used. A final *PostObject*, holding a single event, is finished and stored in a *PostObjectCollection* with all the other events.

3.3 GEO-CODING

Just before accessing the provided REST-API, the Google Geocoding Service[4] can be used to retrieve additional information or just the coordinates of all GPEs for further processing. Please note that a Google API-Key has to be provided for billing purposes if the daily free-use limit of

[3] https://spacy.io/
[4] https://developers.google.com/maps/documentation/javascript/geocoding

2500 requests is exceed. A simple loop processes the stored GPEs for each *PostObject* in the *PostObjectCollection*.

3.4 API-COMMUNICATION

The React REST-API is accessed as described in its documentation. Credentials must be provided before application; otherwise the service will not accept any POST requests. After authentication the gathered PostObjects are simply pushed by POST request as json dictionaries. Error handling and re-authorization has been implemented as well.

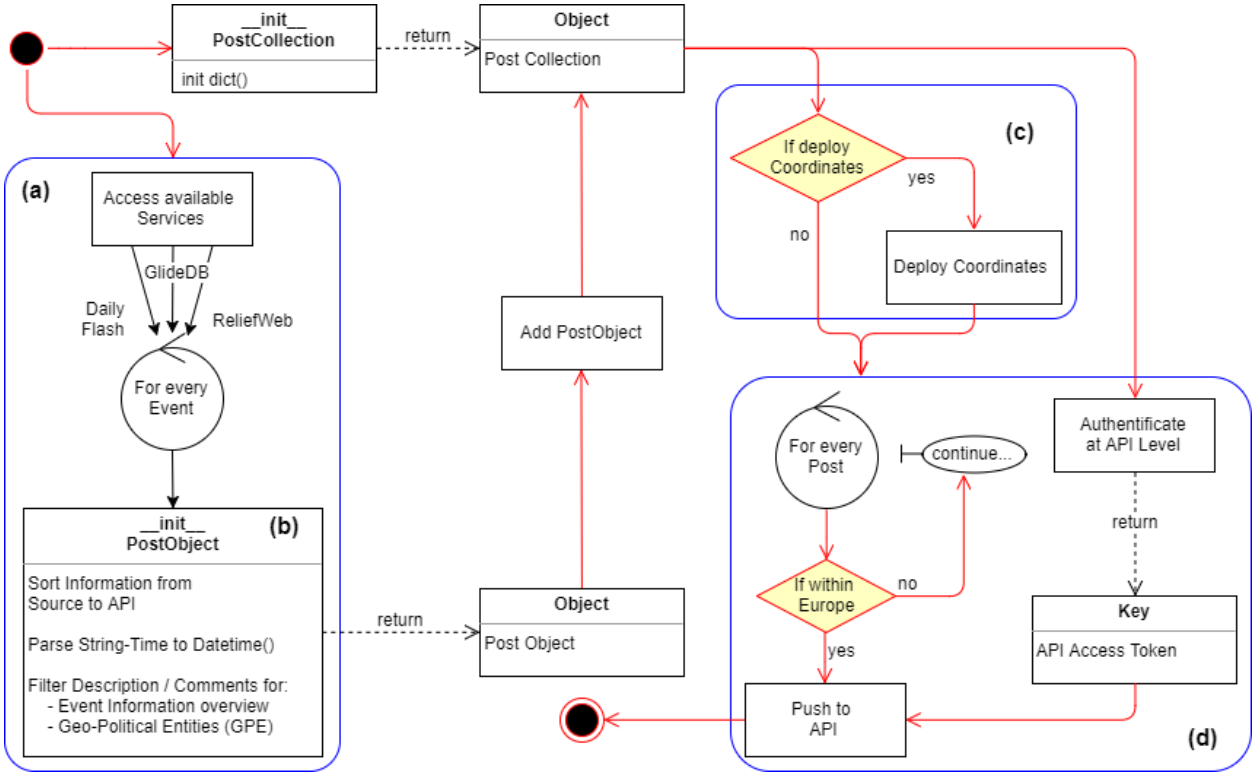

Figure 1: Event Tracker - Flowchart. (a): Web-Scraping; (b): Entry-Sorting &Text-Processing; (c): Geo-Coding; (d):API-Communication

END OF THE DOCUMENT