Social Media Analysis for Situation Awareness during Crises (SMASAC)

GRÉGOIRE BUREL[1], MAYANK KEJRIWAL[2] AND PRASHANT KHARE[1]

[1]Knowledge Media Institute, The Open University, Milton Keynes, UK.

[2]Information Sciences Institute, University of Southern California, Marina del Rey, CA, USA.

WebConf'18, Lyon, France.

23-27 April 2018.

SMASAC - Introduction

GRÉGOIRE BUREL, MAYANK KEJRIWAL AND PRASHANT KHARE

Organisers

Grégoire Burel

Research associate and data scientist at the Knowledge Media Institute (Open University) involved in the COMRADES European project and the lead developer of the Crisis Event Extraction Service (CREES).

Mayank Kejriwal

Computer scientist in the Information Integration group in the Information Sciences Institute (ISI) at the USC Viterbi School of Engineering and works on knowledge graph construction and information extraction for socially consequential domains.

Prashant Khare

PhD student at the Knowledge Media Institute (Open University). His research focuses on extracting knowledge from crowd-generated content in course of crisis/disaster situations.

Speakers

Grégoire Burel
Research Associate

Prashant Khare
Ph.D. Student

Pedro Szekely
Research Associate Professor

Tutorial Outline / Repository

Start Time	Duration (min.)	Session Name	Description
9:00am	30	Introduction	Introduction to situation awareness and social media use during crises.
9:30am	20	Data Collection	Data collection and basic filtering.
9:50am	30	Entity Extraction	Presentations of tools and methods for extracting entities.
10:20am	40	Coffee Break	Coffee break and technical support.
11:00am	30	Categorisation	Automatic classification of documents.
11:30am	30	Event Extraction	Tools and methods for event extraction.
12:00pm	20	Visualisation	Examples of some situation awareness platforms.

Slides and Code:
http://github.com/evhart/smasac-tutorial

Crisis Situation

A **crisis** (from the Greek κρίσις - krisis; plural: "crises"; adjectival form: "critical") is any **event that is going (or is expected) to lead to an unstable and dangerous situation** affecting an individual, group, community, or whole society.

Responding to crises rapidly is critical as they can involve:

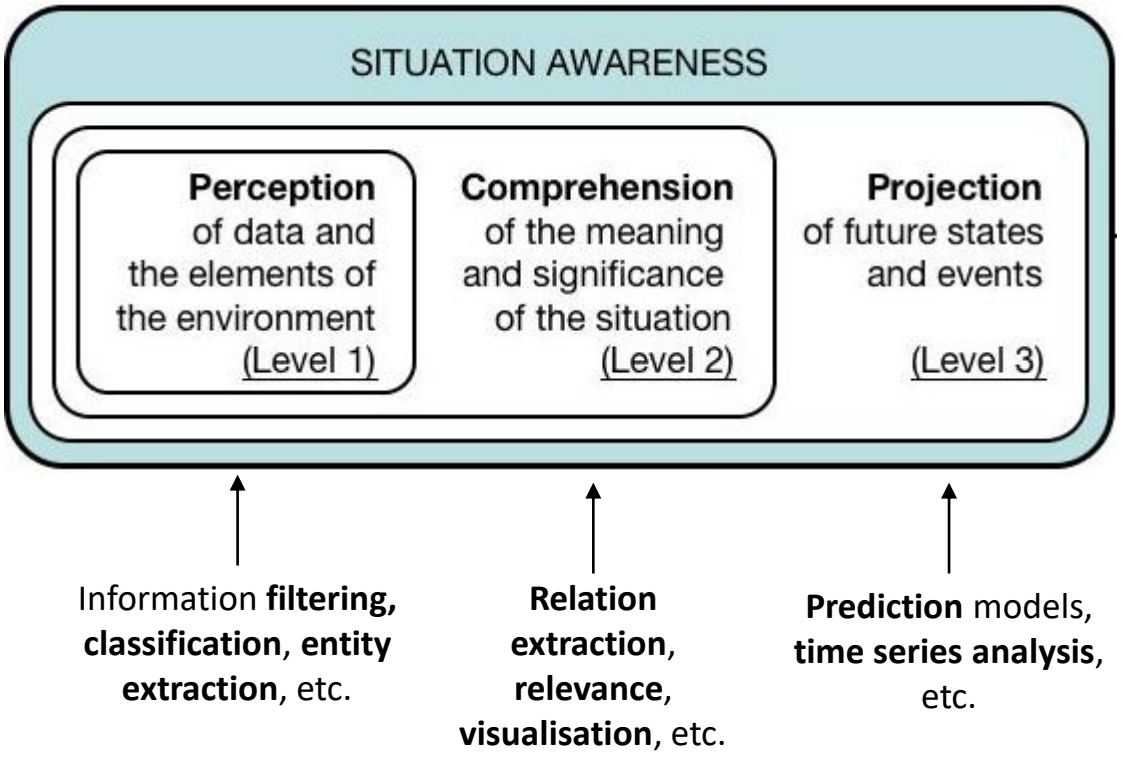
- **Personal injuries** (death, trapping, missing, etc.)
- Infrastructure **damage** (roads blocked, services, etc.)
- **Global issues** (civil arrest, conflicts, etc.)
- **Services overload** (availability and access of emergency services and to resources)

How do we better **understand a crisis situation in order to act accordingly**?

Situation Awareness (1)

Situation awareness is the **perception** of environmental elements and events with respect to time or space, the **comprehension** of their meaning, and the **projection** of their status**.**

Endsley's model of Situation Awareness (Endsley, 1995)

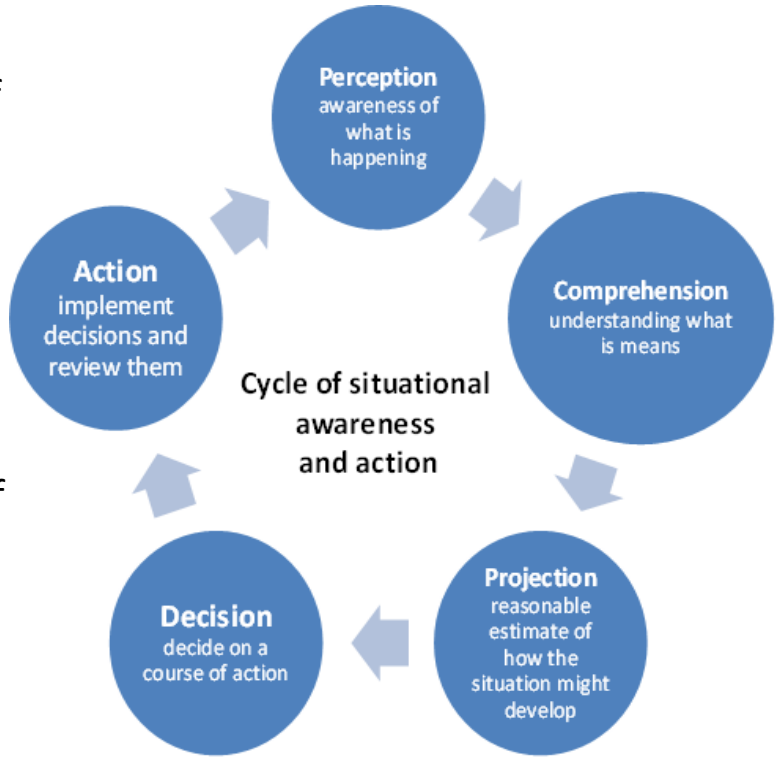


Situation Awareness (2)

Situation awareness is the **perception** of environmental elements and events with respect to time or space, the **comprehension** of their meaning, and the **projection** of their status**.**

Why do we need situation awareness during crises?

- Accessible knowledge can be **integrated to assess and cope with a situation** (Sarter 1991).
- Citizens and responders must be able to **collect reliable information and build an understanding** of the current situation and its evolution (Endsley, 1995).
- **Obtain accurate real-time and complete information** about a particular crisis situation (Winerman, 2009).



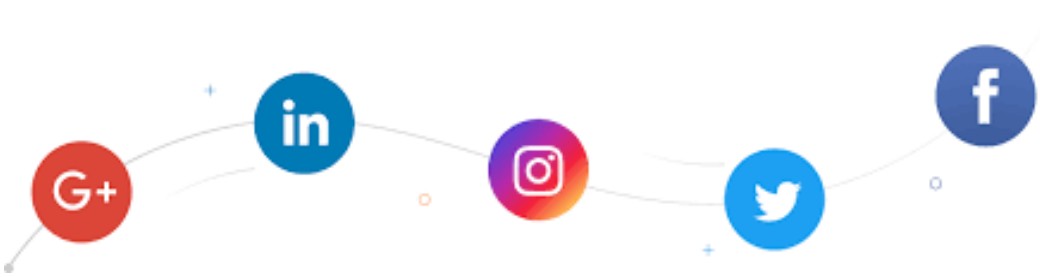
How do we **collect and process efficiently information during emergency situations**?

Social Media (1)

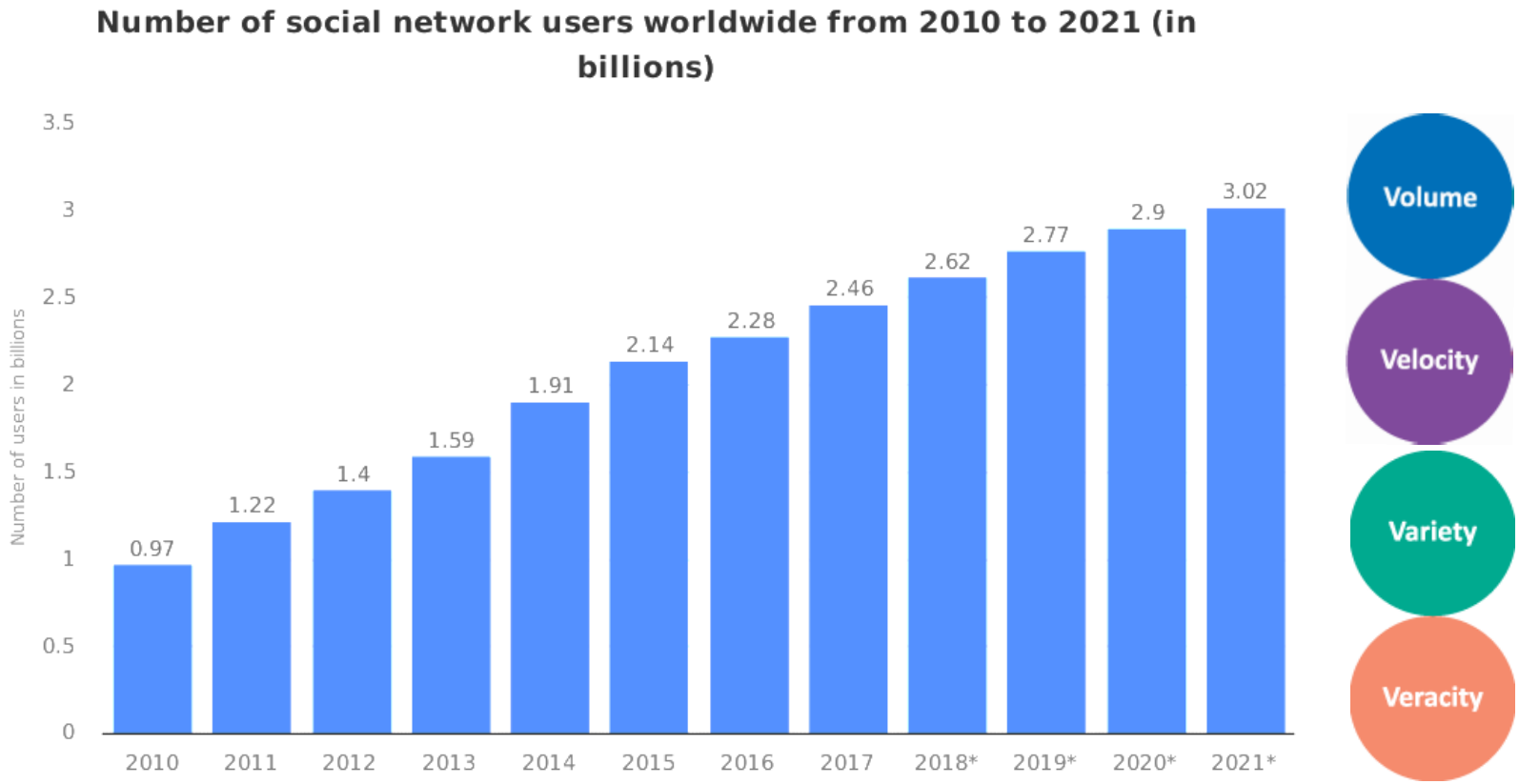
Social media consist of web technologies and platforms that **enable people to interact and engage** by creating content in a conversational and participatory manner

Social web applications are created to **encourage communication between people**, they tend to have some of the following attributes:

- **Identity**: who are you? (e.g., Facebook)
- **Reputation**: what do people think you stand for? (e.g. LinkedIn)
- **Presence**: where are you? (e.g., Four Square)
- **Relationships**: who are you connected with? who do you trust? (e.g., Facebook)
- **Groups**: how do you organize your connections?
- **Conversations**: what do you discuss with others? (e.g., Reddit).
- **Sharing**: what content do you make available for others to interact with?



Social Media (2)



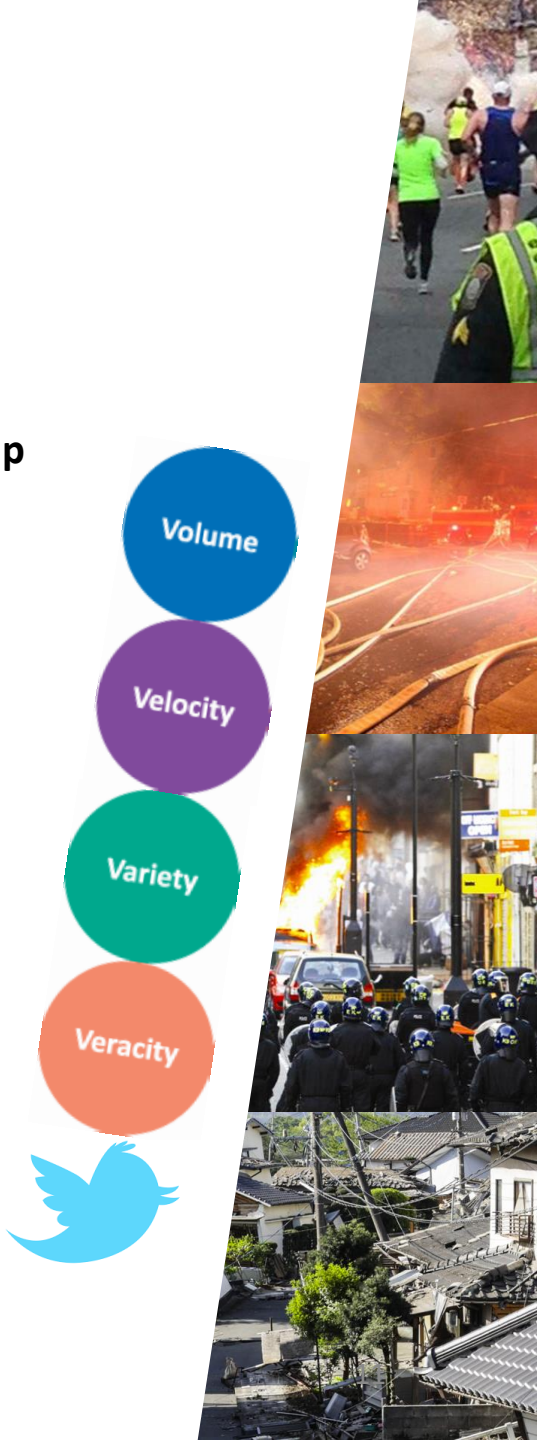
Social Media and Crises

Social media has become a common place for **communities** and **organisations** to **communicate** and **share information** during crises, to enhance their situational awareness, to share **requests** or **offers for help** and **support**, and to **coordinate** their recovery efforts.

Example of Twitter usage during crises (+200m active users / +400m posts a day):

1. During the 2011 Japan earthquake, **177 million tweets** related to the event were sent in one day.
2. The **news about the Boston bombings first appeared on Twitter**.

How do we **obtain situation awareness** from **noisy** and **large heterogeneous high-velocity real-time unstructured data**?



Hurricane Harvey – Twitter Usage (1)

Rockabye Baby Rentals
@RockabyeAnnie

Rescue needed [redacted] 3 adults & 3 Infants on Roof in Meyerland

10:37 PM - Aug 27, 2017

See Rockabye Baby Rentals's other Tweets

Shea Serrano
@SheaSerrano

if anyone has a boat in the area we need to get this mom and her tiny baby out as soon as possible twitter.com/jennyg11002/st…

2:17 AM - Aug 29, 2017

70 116 people are talking about this

BONEZ @Herringbonez · 27 Aug 2017

A NEW THREAD IS BEING CREATED AS OF NOW.

IN ORDER TO KEEP TRACK OF EACH INDIVIDUAL

1. NAMES/AGES
2. ADDRESS
3. PHONE #
4. EXACT LOCATION

34 72 62

Hurricane Harvey – Twitter Usage (2)

Annie Sullivan
@anniemsullivan
Follow

Kim Kardashian, Miranda Lambert and More Stars React to Hurricane Harvey: Stars like Kim… dlvr.it/PhqGTm fhfusa.us

Hector Moran
@djhectormoran
Follow

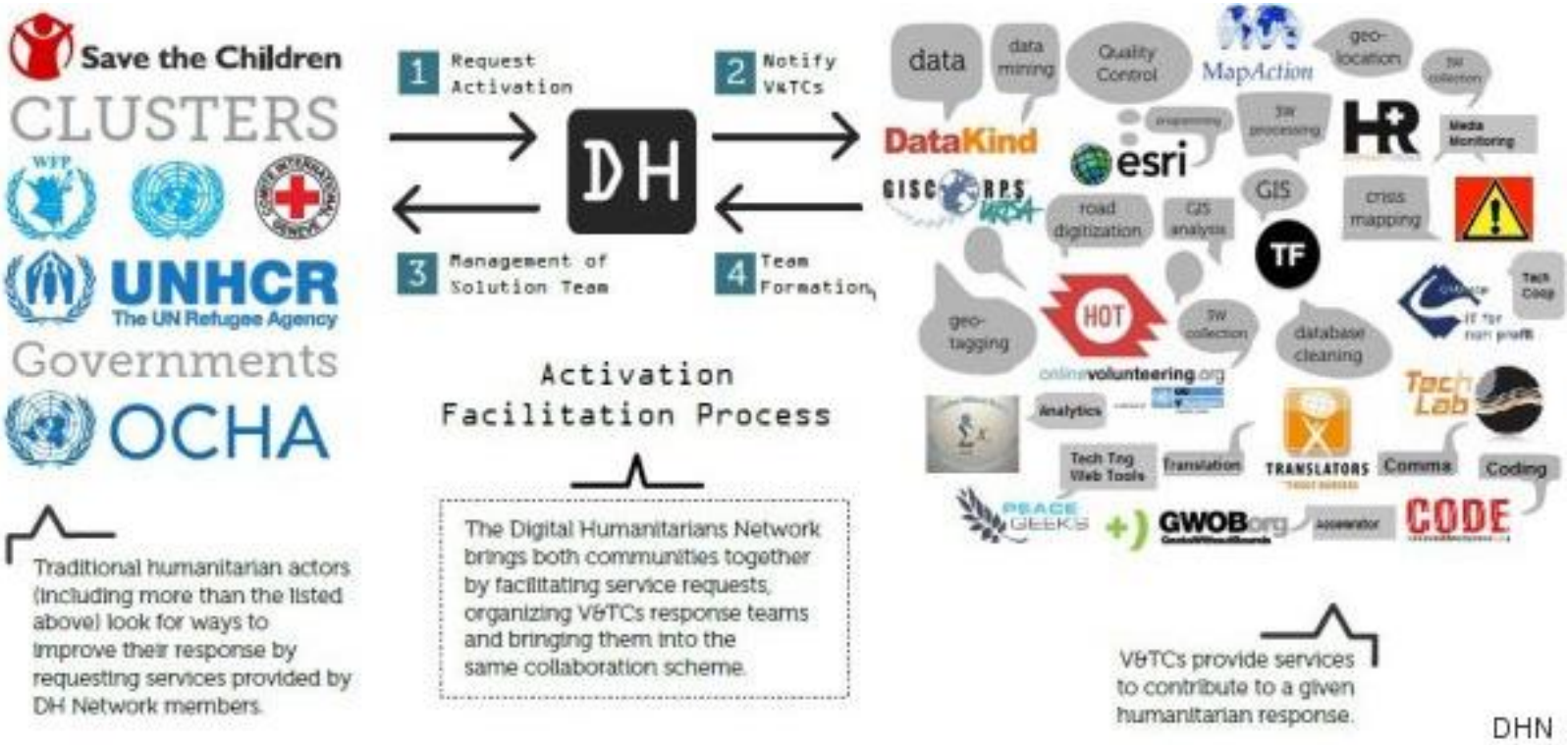
Harvey Express...
.
:
.
.
.
#electricscooter #razor #hectormoran #hectormoranontour #harvey…
instagram.com/p/BYRZbcfjEe0/

Ty
@RRihannaGrande
Follow

texas fellas, is it gay to be in this hurricane? i mean, you're literally getting blown by a dude named harvey..... sound a lil spicy 2 me

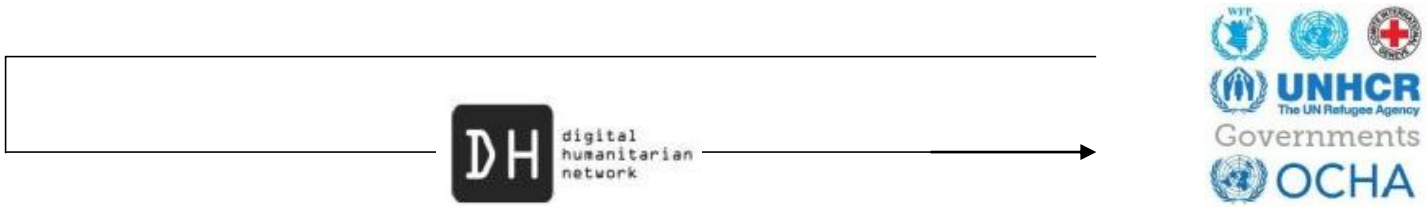
Digital Humanitarian Network et al.

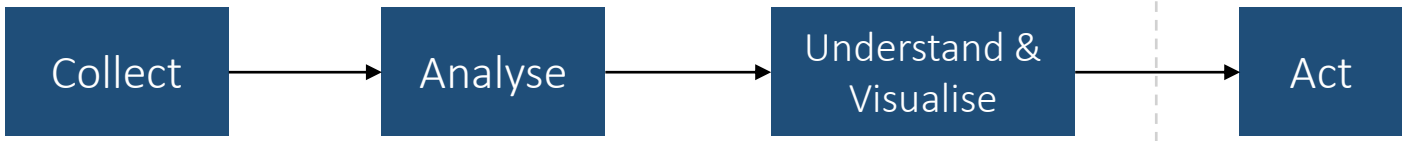
With the emergence of social media and access to other technologies (e.g., mapping tools and information management platforms), **digital volunteer networks** have been created to help data processing tasks. Nevertheless, low-level **tasks** (e.g., data collection, information extraction and filtering) **remain mostly manual labour**.



Digital Situation Awareness Pipeline

When dealing with information collected from various **online sources** (e.g., SM, mapping software, etc.), it is necessary to be able to **collect**, **represent**, **process** and **visualise** such information in order to better **understand a situation.**





Gather data from various information sources

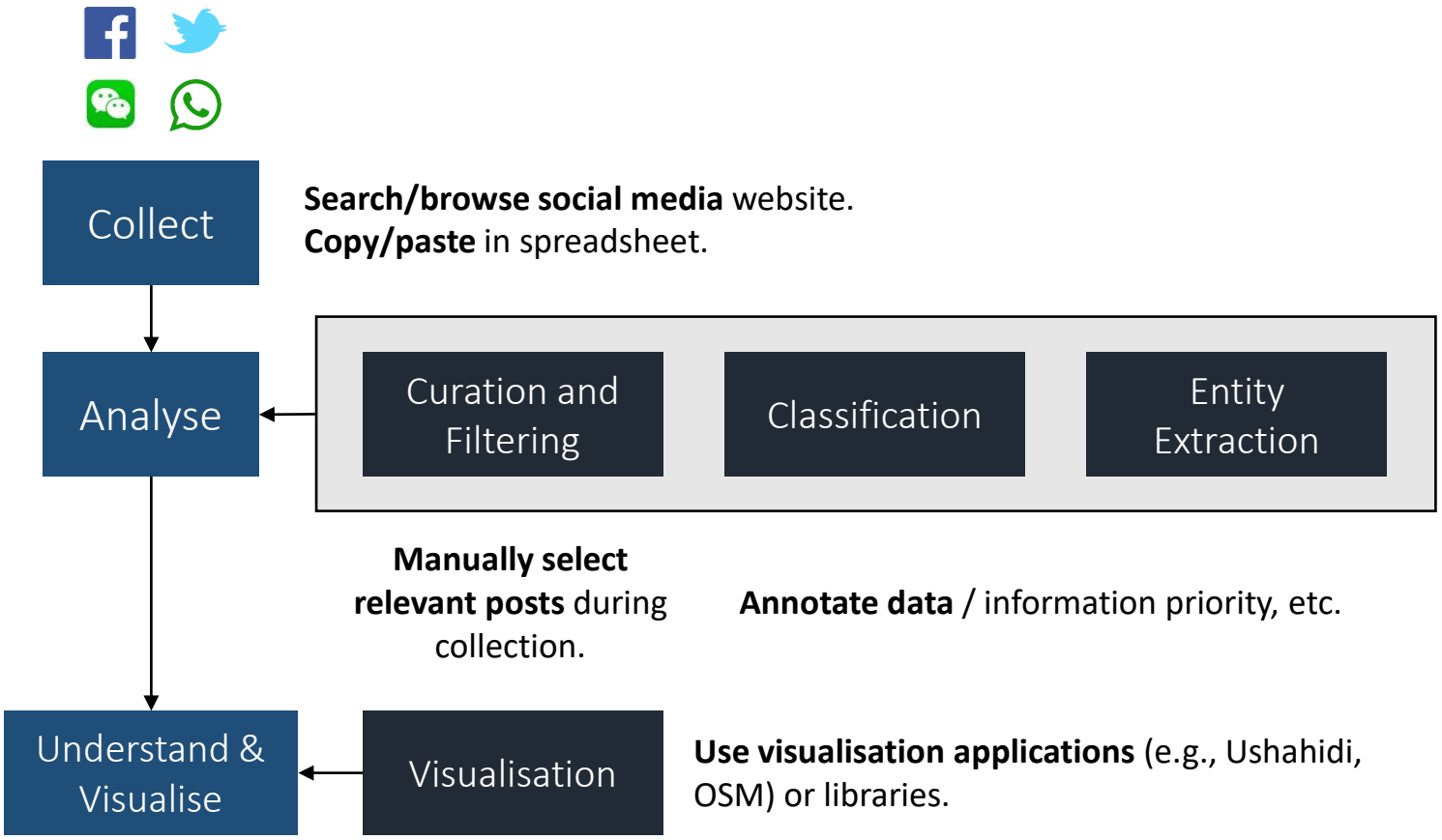
Extract key information, verify trustworthiness, classify, etc.

Connect and **Visualise** information.

Decide what to do and organise actors and resources.

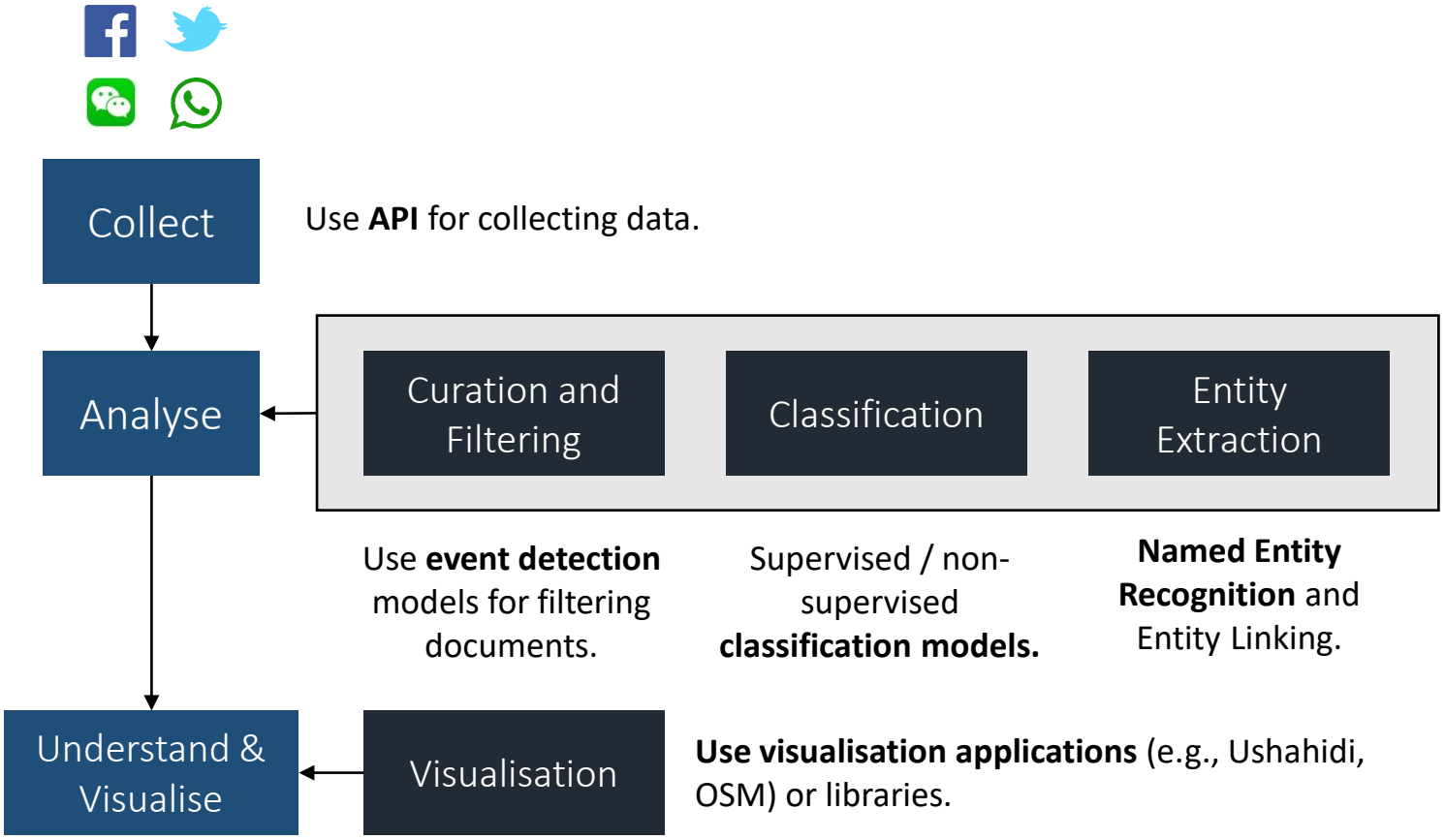
Typical (Manual) Processing Pipeline

During crises, **most of the data is processed manually** by crisis responders, **digital humanitarian**, organisations and individuals.



Automatic Processing Pipeline

Supervised or semi-supervised methods as well as **APIs** an automatic tools can be used for collecting data**.**



Automatic Data Collection and Analysis

Collect

Data can be easily collected without requiring manual browsing**:**

- Use streaming/search **API for collecting data.**
- Typically **restrict search** using keywords and location.
- **Reformat** data/extract basic information automatically for further processing (e.g., tokenisation, hashtags, URLs).
- **Store data** in database (or save in a file).

Analyse | Curation and Filtering | Classification | Entity Extraction

Relevant data can be filtered automatically using keywords as well as event classifiers**:**

- Use **trained classifiers** for identifying relevant documents and other relevant information**.**
- Dynamically re-train classifiers.

Extract important information from documents (e.g., locations, actors, etc.):

- Use Named Entity Recognition (NER) tools.

Data Visualisation

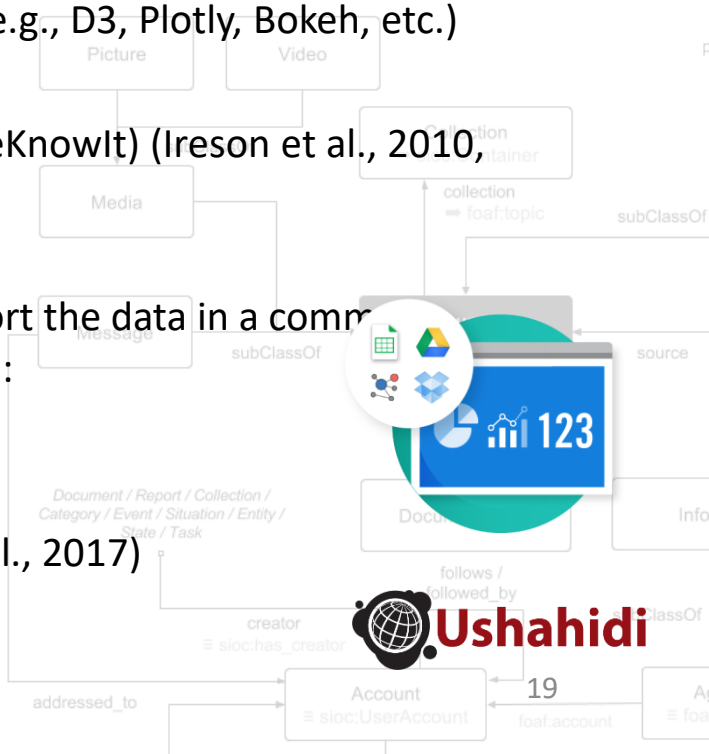
Understand & Visualise | Visualisation

Visualisation platforms or libraries can be used for visualising the processed data**.** For example, the following visualisation can be done:

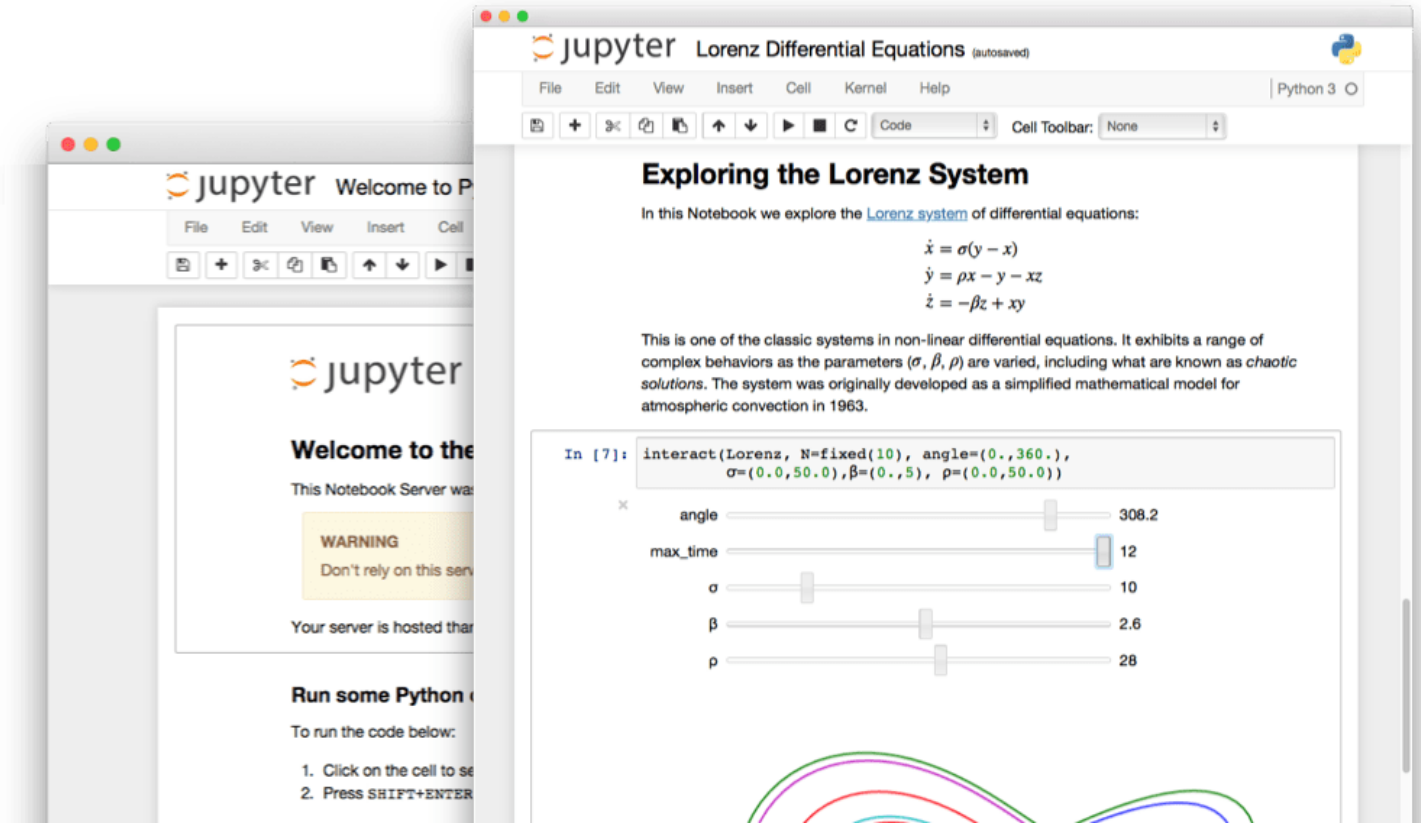
- **Import data in situation awareness platforms** (e.g., Ushahidi, Sahana).
- Directly visualise the data using **visualisation libraries** (e.g., D3, Plotly, Bokeh, etc.)
- Use **HDX headers** and Quick charts.
- Use **ontological visualisation** platforms (e.g., Topica/WeKnowIt) (Ireson et al., 2010, Cano et al., 2011).

Typically, the first step before visualisation would be to export the data in a comm
model or reformat the data according to the target platform:

- **CSV** or CSV with **HDX** headers (**HXL**).
- **JSON** format or custom format
- **Common ontology** (e.g., CURIO, DoRES, etc.) (Burel et al., 2017)



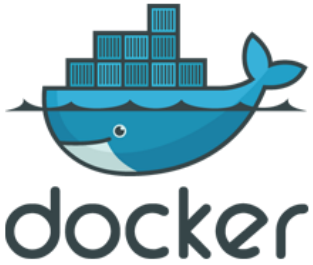
Hands-on Requirements



Software Requirements

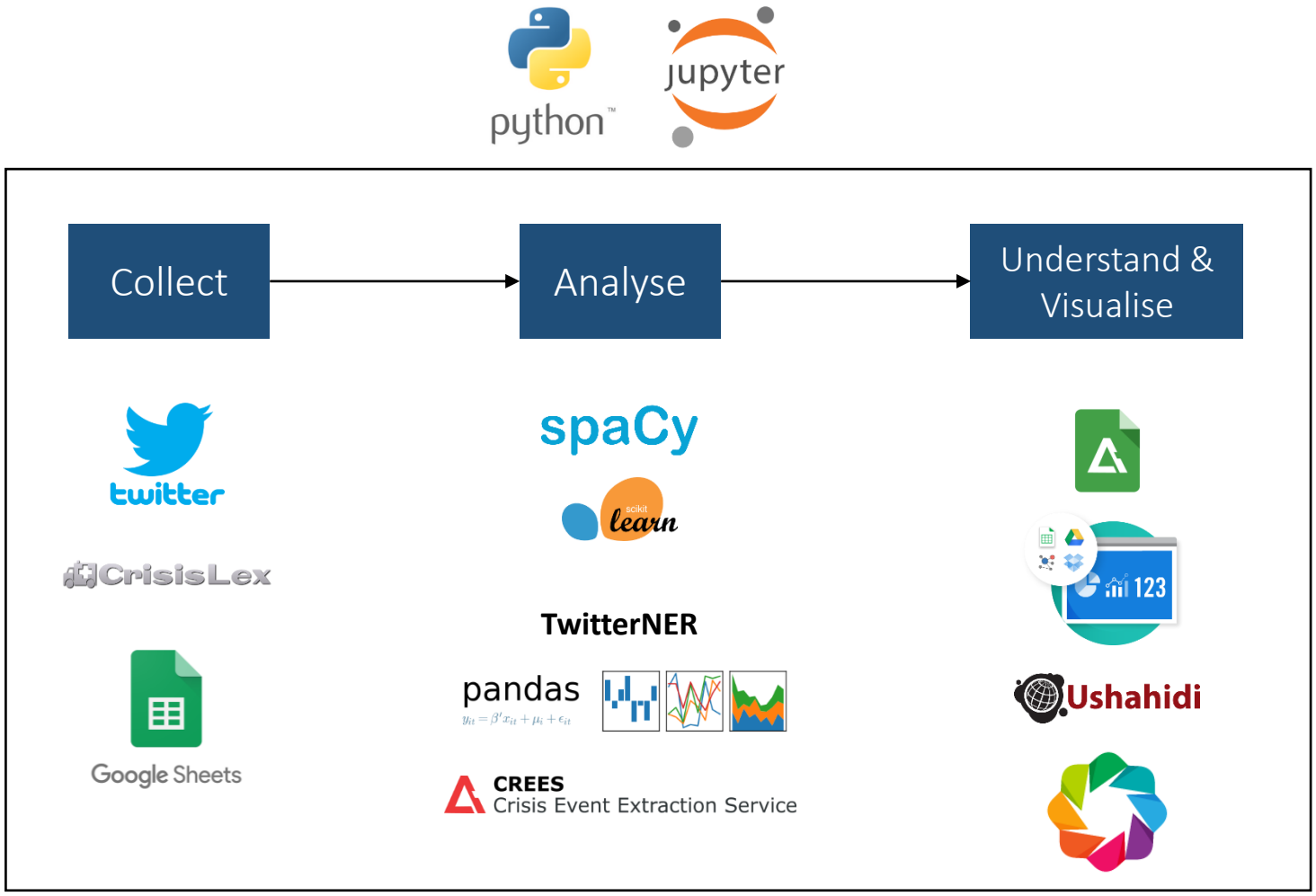
Local installation

Local installation
(via Docker image
or repo2docker)



Remote execution
(via mybinder.org)

Examples of Tools and Libraries



Used Data

The tutorial is focusing on **Twitter** data due the its availability**.** We use the **CrisisLexT26** data since it is already made available. **CrisisLex** (http://crisislex.org/) is a repository of crisis-related social media data and tools. It includes **collections of mostly Twitter crisis data and a lexicon of crisis terms**.

The **CrisisLexT26** dataset contains:

- Tweets from **26 crises**, labelled by informativeness, information type and source.
- **Data collected between 2012-2013**.
- **~250K tweets** (~28 000 annotated tweets).
- Crises includes: *Bombing, Building collapse*, *Crash, Derailment, Earthquake, Explosion, Fire, Floods, Haze, Shootings, Typhoon, Wildfire…*

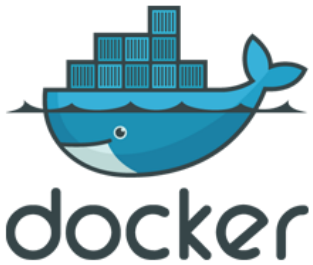
We also use a dataset about the **Las Vegas Massacre** :

- **~50K tweets** collected from the Twitter API in the aftermath of the Las Vegas massacre (2017).

CrisisLex

Tools/Workspace Configuration

Hands-on



or

Slides and Code:
http://github.com/evhart/smasac-tutorial