# Fusing Flight Data with Social Media Data

**Dr. Axel Tanner**
IBM Research Zurich
Säumerstrasse 4
CH-8803 Rüschlikon
SWITZERLAND

axs@zurich.ibm.com

**Dr. Albert Blarer**
armasuisse Science + Technology
Feuerwerkerstrasse 39
CH-3602 Thun
SWITZERLAND

albert.blarer@armasuisse.ch

**Dr. Vincent Lenders**
armasuisse Science + Technology
Feuerwerkerstrasse 39
CH-3602 Thun
SWITZERLAND

vincent.lenders@armasuisse.ch

## *ABSTRACT*

*Military decision making is often based on information systems with humans-in-the-loop which have to interpret data from multiple sources. This process can be very overwhelming for humans when the data sources produce large amounts of data.*

*We investigate in how far big data analytics and information fusion techniques can be used to support humans in handling large amounts of heterogeneous data and improve the understanding of unfolding events as part of the Observe and Orient steps of the OODA cycle.*

*Our work focuses on fusing data from two very different data sources: user-generated content from the social media platform Twitter and air traffic control data from the OpenSky sensor network. Our goal is to find and provide detailed information about events related to the aviation domain that are represented at the same time in both data sources. The challenge lies in fusing exact and explicit data from airplane communication versus the very broad and imprecise natural language used in Twitter.*

*To bridge the semantic gap between these sources, we develop an advanced information fusion model which allows us to use each source as trigger for events while enriching the data with information from the other source. Using real-world data that we have collected over several months, we show multiple cases of evidence that both sources provide mutual enrichment. This is done in an automated fashion but leads in general to a more loose and inexact relationship that requires suitable interpretation and understanding by humans. Still, the combination enhances the understanding and therefore is highly helpful as a basis for decision makers to assess the events as they unfold and act accordingly.*

# 1. INTRODUCTION

Information fusion is an important way to enhance the understanding of arising situations for decision making, possibly for automatic evaluation or filtering of situations, or for an enriched presentation to the human decision maker in the OODA loop.

In our work, we investigate how a data source with flight data is related to data from social media. As source for flight data we are using air traffic control data gathered by the crowd-sourced OpenSky sensor network[1], our source for social media data is Twitter[2]. As illustrated in Figure 1, these data sources have very different characteristics:

- OpenSky contains highly specific data about airplanes during flight extracted from ADS-B and Mode S messages gathered through crowd-sourced sensors. They are stored as so-called StateVectors that contain the identity of the airplane and the flight, as well as position, direction and status information at a specific time. Limitations: reception of messages depends on the availability of sensors within line of sight to the plane, therefore flight tracks are often not covered through the full flightpath.
- Twitter has around 100 million active users per day, who are sending ~500 million short messages (limited to 140/280 characters) per day about all possible topics that the users are concerned with. This can be interpreted as a source for *ambient awareness* of all kinds of events occurring around the world. The content of tweets is completely free-format and often contains abbreviations, sloppy expressions and mistakes, therefore are very noisy. Limitation: only 1% of the overall volume is publicly accessible through the free Twitter streaming API that is used for our research purposes.

Both sources deliver *Big Data* in the sense of volume and velocity, but where the flight data has a high veracity and low variety, Twitter data has a rather low veracity paired with a high variety.

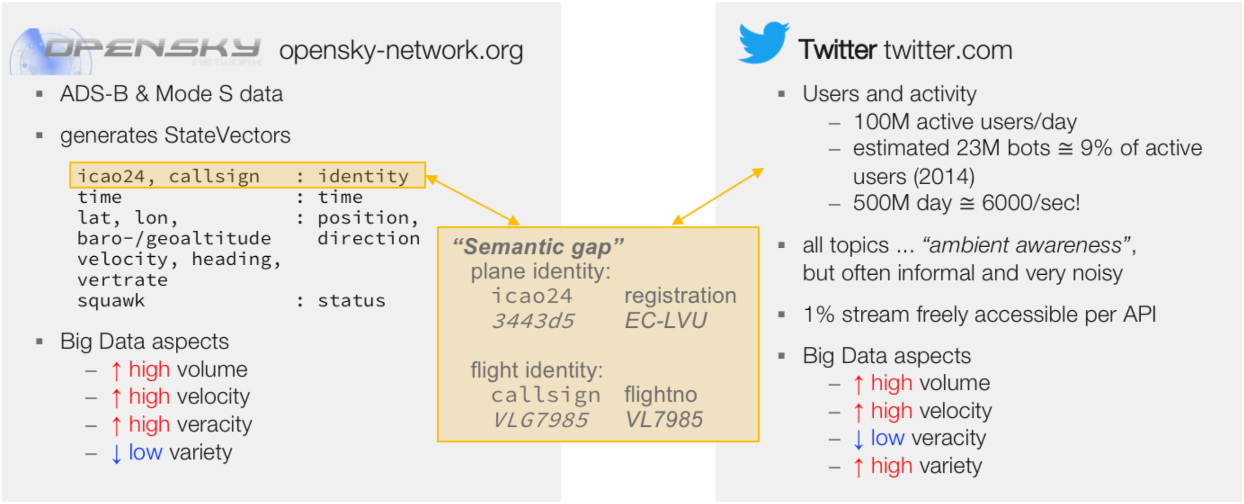


**Figure 1: Data sources and semantic gap**

[1] The OpenSky Network https://opensky-network.org/

[2] Twitter https://twitter.com/

Flights and airplanes are of general interest to many people, especially in cases of problems, be it delays or incidents. Intuitively, we therefore expect that many types of anomalies around flights and airplanes will get attention by people and lead to tweets with corresponding information.

In the following, we describe our experiments in fusing these data sources and our findings in the initial experiments. Section 2 describes related work, Section 3 the fusion model to bridge between the sources and Section 4 describes the possible triggering factors raising situations of interest. The paper concludes with a discussion of the results in Section 5.

## 2. RELATED WORK

The topic of data fusion arose originally in the context of multi-sensor environments. A good overview and classification of this field was given by Castaneda [1]. A lot of work targets specific contexts and environments, like the fusion of data to reach a higher level of cyber awareness (see, e.g., [2]). The importance of *context* for information fusion is extensively discussed and formalized in [3].

With the omnipresence of social media in our current world, it is natural to use such data as source and context to trigger and evaluate arising situations. E.g., Akbar et al. [4] include Twitter data in the context of IoT data for the prediction of congestions in Madrid. In this work, though, only the number of tweets in certain geographic regions is used as additional indicators. Deeper use of Twitter content, also in the context of traffic analysis, is made in [5] that is closest to our work.

To our knowledge, the investigations to join data from air traffic with social media data, as presented here, is still new.

## 3. BRIDGING THE DATA SOURCES

Our goal in this work is to identify anomalous situations in air traffic, using air traffic data as well as social media data to trigger and enrich the understanding of the events.

All data fusion efforts require to *bridge* between the different data sources, in order to identify what data is related to the same entity and situation. Especially in sources that are as disparate as in our case, this is challenging.

For example, Figure 1 demonstrates this *semantic gap* between the data sources: OpenSky data identifies airplane and flight by the *ICAO 24-bit aircraft address*[3] (the ID of the transponder, usually fixed throughout the lifetime of an aircraft) and the *callsign* (specific for the current flight, usually reused each day for the same flight of an airline). On the other hand, in human communication one uses most often the *registration number* specifying the airplane and the *flight number* to specify the flight.

Identifying the tweets mentioning a specific airplane or flight as seen in the OpenSky data requires translating between these data elements. Although, e.g., the callsign is often related to the flight number, there is no general mapping to translate these data elements to each other, therefore the mapping must be learned over time.

[3] See, e.g., https://en.wikipedia.org/wiki/Aviation_transponder_interrogation_modes

# 4. IDENTIFYING SITUATIONS

Figure 2 gives an overview of our data flow and data fusion model that combines data from Twitter and OpenSky. Due to the high volume and velocity, handling the data in real-time requires substantial computational resources.

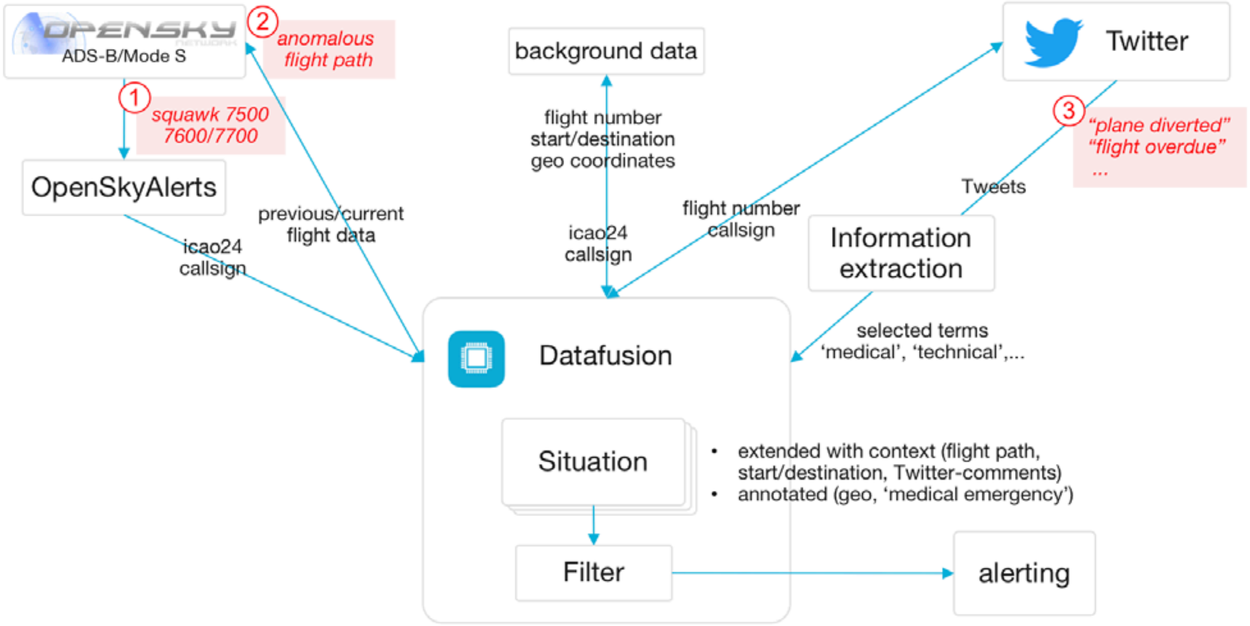


**Figure 2: Overview of data sources, flow and processing of data, and triggers for situations**

With two sources we can investigate different triggers for raising situations of potential interest: on the side of the flight data, either abnormal status codes (*squawk* codes) or anomalies in flight patterns are used as triggers, on the side of social media, we look for different phrases to trigger situations to further analyse. These scenarios are investigated in more details in the following subsections.

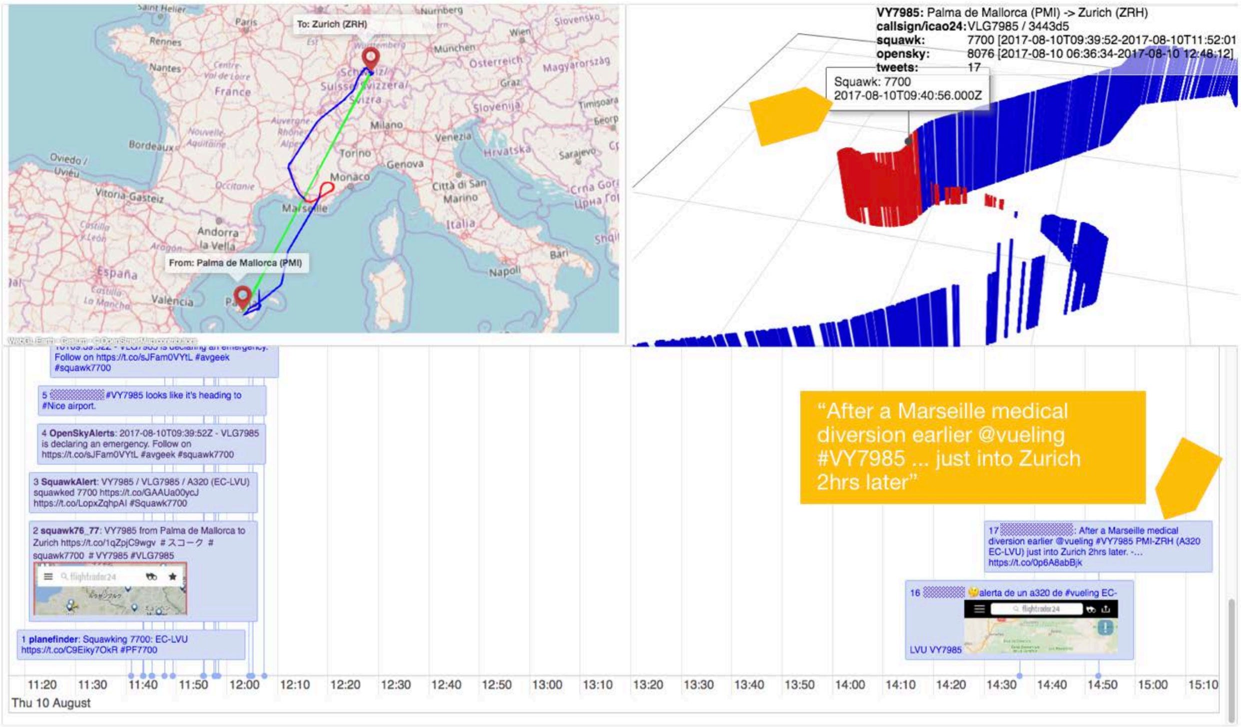
## 4.1 Trigger by anomalous status codes

Flight data in OpenSky contains for every entry a *transponder code* (also known as *squawk* code consisting of 4 octal digits)[4] that signals the current status of the aircraft. The meaning of the codes differs partially for different regions in the world, but the emergency codes *7500 (aircraft hijacking)*, *7600 (radio failure)*, *7700 (emergency)* are standardized world-wide.

The OpenSky network has created a special alerting API endpoint[5] that pre-processes the data to suppress spurious alerts resulting, e.g., from sending an alert when switching dials at the transponder.

[4] For more information, see, e.g., https://en.wikipedia.org/wiki/Transponder_(aeronautics)#Transponder_codes

[5] OpenSky Live Alerts https://opensky-network.org/api/alerts/all - also visible on Twitter https://twitter.com/openskyalerts



**VY7985:** Palma de Mallorca (PMI) -> Zurich (ZRH)
**callsign/icao24:** VLG7985 / 3443d5
**squawk:** 7700 [2017-08-10T09:39:52-2017-08-10T11:52:01
**opensky:** 8076 [2017-08-10 06:36:34-2017-08-10 12:48:12]
**tweets:** 17

Squawk: 7700
2017-08-10T09:40:56.000Z

Follow on https://t.co/sJFam0VYtL #avgeek #squawk7700

5 [redacted] #VY7985 looks like it's heading to #Nice airport.

4 **OpenSkyAlerts**: 2017-08-10T09:39:52Z - VLG7985 is declaring an emergency. Follow on https://t.co/sJFam0VYtL #avgeek #squawk7700

3 **SquawkAlert**: VY7985 / VLG7985 / A320 (EC-LVU) squawked 7700 https://t.co/GAAUa00ycJ https://t.co/LopxZqhpAI #Squawk7700

2 **squawk76_77**: VY7985 from Palma de Mallorca to Zurich https://t.co/1qZpjC9wgv #スコーク # squawk7700 # VY7985 #VLG7985

1 **planefinder**: Squawking 7700: EC-LVU https://t.co/C9Eiky7OkR #PF7700

"After a Marseille medical diversion earlier @vueling #VY7985 ... just into Zurich 2hrs later"

17 [redacted]: After a Marseille medical diversion earlier @vueling #VY7985 PMI-ZRH (A320 EC-LVU) just into Zurich 2hrs later. -... https://t.co/0p6A8abBjk

16 [redacted] alerta de un a320 de #vueling EC-LVU VY7985

**Figure 3: User interface showing flight and twitter data for an emergency case (squawk code 7700). The upper-left part shows the flight as planned with start and destination airports connected by a green line, together with the observed flight path in blue (normal squawk codes) and red (emergency squawk code). In the upper-right the same observed flight data is shown in a 3d representation. The lower part shows a timeline with tweets found for the given callsign and flight number.**

Figure 3 shows our experimental user interface for investigating alert situations in both sources. Triggered by an emergency status (squawk 7700 around 9:40 UTC on 2017-08-10) of the flight with callsign/icao24 VLG7985/3443d5, our system fetches registration and flight number as well as planned start and destination from online background data[6]. The system learns that this is the flight VY7985 Palma de Mallorca to Zurich and starts ad-hoc to search and collect tweets for this data. Flight and Twitter data are gathered and represented visually.

The Twitter timeline shows within a few minutes after the emergency occurs some known bots announcing the alert via Twitter (including *OpenSkyAlerts* with tweet 4), but also some individual users wondering about the changed route (e.g., tweet 5). Not many other tweets are observed until around 3 hours later, when some user confirms that a medical emergency happened during the flight, leading to an intermediate landing in Marseille.

Observing the tweets for multiple different emergency events, we find multiple Twitter bots that very quickly after an emergency squawk send tweets (like the Twitter users *planefinder*, *squawk76_77*, *SquawkAlert* besides *OpenSkyAlerts* - see Figure 3 lower-left). Only few bot-like addresses are observed to send updated information later on, judged from the varying format of tweets seemingly by human users.

[6] Various (paid and free) services are available allowing mapping between callsigns and flight numbers, and resolving scheduled starting point and destination, e.g., planefinder.net, flightaware.com, flightrader24.com. Geographic coordinates of airports can be found, e.g., in DBpedia.

We use this trigger to start collecting more potentially related tweets (as opposed to the general stream of 1%), as all tweets for the last 7 days are accessible through a REST search API (as long as the resulting number of tweets stays below some limits).

Caveat: identifying tweets of interest via flight number is not perfect due to the semantic ambiguity, especially for short flight numbers. For example, during our experiments we had an alert for the valid flight number *TN8* that is also the abbreviation for a TV channel in Nicaragua – drawing many unrelated tweets into our collection.

As expected, even the content of the correctly related tweets found in the context of such emergency events are very mixed, containing questions mixed with additional information *(“What happened to VY7985, Vueling airbus from Palma to Zürich? Landed on MRS”*) or more solid information (*“After a Marseille medical diversion earlier @vueling #VY7985 ... just into Zurich 2hrs later”*), though always in free text form. This might already be helpful for human decision makers, but ideally one could extract more explicit information automatically.

We investigated several different possibilities:

- Annotations:
  - Explicit matching for airport codes (extracted from DBpedia[7]). Caveat: some airport codes are regular words (e.g., the Anderson Regional Airport in North Carolina, US, has the airport code AND)
  - Detecting flight numbers (per regular expression). Caveat: the general scheme allowed for flight numbers is quite general and will catch many other entities.
  - Detecting special terms (like ‘medical’, ‘technical’, ‘radio’, ‘emergency’, ...)
  - Named Entity Recognition (NER) with DBpedia-spotlight[8]
- Summarization of tweets, removing repeated information and ranking top-tweets via a scheme based on word-frequencies using the NLTK package[9], giving higher weights to annotated terms.

The DBpedia-spotlight annotations are especially helpful, a wide range of entities is recognized, and corresponding categories are returned. In our context, categories for geographic places, airports, airlines and aircrafts are especially interesting (some examples shown in Table 1). Again of course, not all terms describe a unique entity (e.g., *BA* is recognized as *British Airways* as well as *Bachelor of Arts*).

**Table 1: Example terms recognized by DBpedia-spotlight**

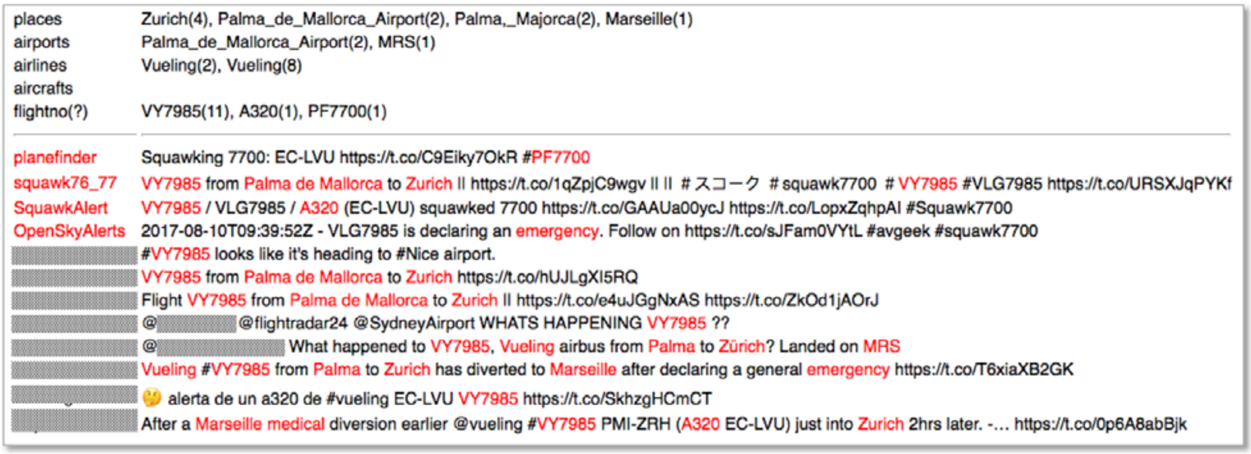
| **Example terms** | **Class** | **DBpedia category** |
|---|---|---|
| ‘Marseille’, ‘Aylesbury’ | geographic places | DBpedia:Place |
| Palma de Mallorca’, ‘MRS’ | Airports | DBpedia:Airport |
| ‘Vueling’, ‘British Airways’, ‘BA’ | Airlines | DBpedia:Airline |
| ‘A-380’, ‘Airbus 319’, ‘B787’ | Aircraft | DBpedia:Aircraft |

[7] http://dbpedia.org/

[8] http://www.dbpedia-spotlight.org/

[9] Natural Language Toolkit (NLTK) https://www.nltk.org/

As an example, Figure 4 shows the results of the annotation and summarization for the medical emergency event described above.



| | |
|---|---|
| places | Zurich(4), Palma_de_Mallorca_Airport(2), Palma,_Majorca(2), Marseille(1) |
| airports | Palma_de_Mallorca_Airport(2), MRS(1) |
| airlines | Vueling(2), Vueling(8) |
| aircrafts | |
| flightno(?) | VY7985(11), A320(1), PF7700(1) |
| planefinder | Squawking 7700: EC-LVU https://t.co/C9Eiky7OkR #PF7700 |
| squawk76_77 | VY7985 from Palma de Mallorca to Zurich II https://t.co/1qZpjC9wgv II II ＃スコーク ＃squawk7700 ＃VY7985 #VLG7985 https://t.co/URSXJqPYKf |
| SquawkAlert | VY7985 / VLG7985 / A320 (EC-LVU) squawked 7700 https://t.co/GAAUa00ycJ https://t.co/LopxZqhpAI #Squawk7700 |
| OpenSkyAlerts | 2017-08-10T09:39:52Z - VLG7985 is declaring an emergency. Follow on https://t.co/sJFam0VYtL #avgeek #squawk7700 |
| [redacted] | #VY7985 looks like it's heading to #Nice airport. |
| [redacted] | VY7985 from Palma de Mallorca to Zurich https://t.co/hUJLgXI5RQ |
| [redacted] | Flight VY7985 from Palma de Mallorca to Zurich II https://t.co/e4uJGgNxAS https://t.co/ZkOd1jAOrJ |
| [redacted] | @[redacted] @flightradar24 @SydneyAirport WHATS HAPPENING VY7985 ?? |
| [redacted] | @[redacted] What happened to VY7985, Vueling airbus from Palma to Zürich? Landed on MRS |
| [redacted] | Vueling #VY7985 from Palma to Zurich has diverted to Marseille after declaring a general emergency https://t.co/T6xiaXB2GK |
| [redacted] | 🤔 alerta de un a320 de #vueling EC-LVU VY7985 https://t.co/SkhzgHCmCT |
| [redacted] | After a Marseille medical diversion earlier @vueling #VY7985 PMI-ZRH (A320 EC-LVU) just into Zurich 2hrs later. -… https://t.co/0p6A8abBjk |

**Figure 4: Detection of selected terms of interest (places, airports, airlines, aircrafts and flight numbers – count of term is given in parenthesis) and ranked selection of tweets with annotation highlights**

## 4.2 Anomalous Flight Path

During our research, a mid-air collision of a small airplane and a helicopter happened near Aylesbury, UK[10]. In this case, no squawk alert happened as the collision happened without warning. We investigated the available data manually, looking at flight data and tweets (see Figure 5). Quickly within the first tweets after the collision, the registrations of the two involved aircrafts are mentioned. Results of running our entity recognition and summarisation of tweets are shown in Figure 6. After resolving the registration numbers to icao24 codes via online background data, flight data from OpenSky show the collision in the altitude data while both aircraft seem to descend to a nearby airport (see Figure 7).

[10] "Aylesbury mid-air crash: Four dead as plane and helicopter wreckage lands near Rothschild manor house" https://www.telegraph.co.uk/news/2017/11/17/aylesbury-mid-air-crash-fatalities-feared-afteraircraft-helicopter/

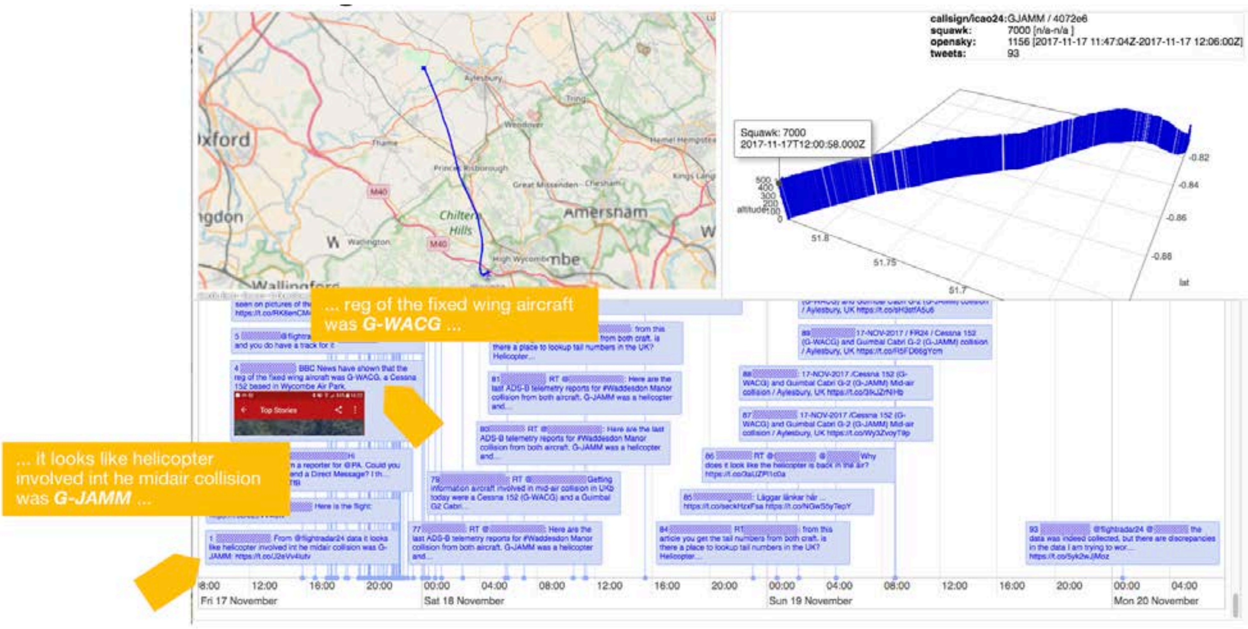
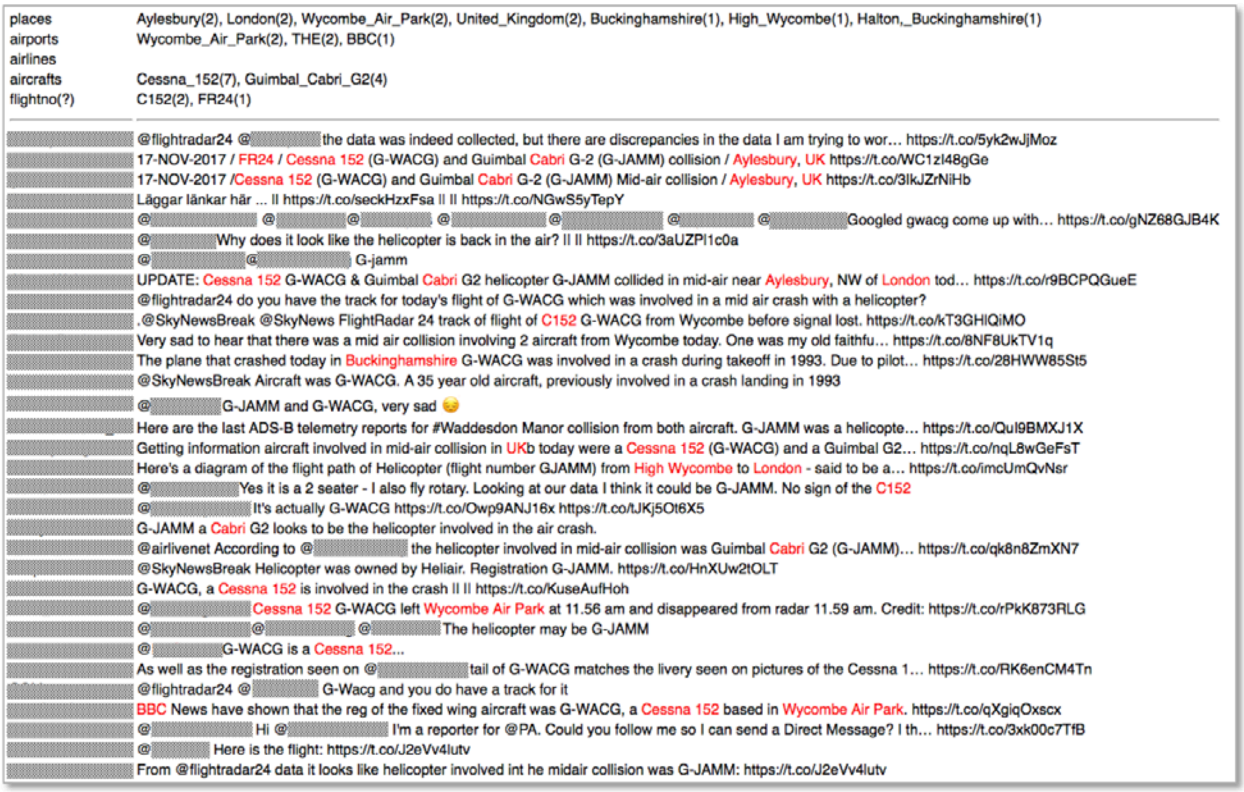


**Figure 5: Collision of two aircrafts near Aylesbury on 2017-11-17**



| | |
|---|---|
| places | Aylesbury(2), London(2), Wycombe_Air_Park(2), United_Kingdom(2), Buckinghamshire(1), High_Wycombe(1), Halton,_Buckinghamshire(1) |
| airports | Wycombe_Air_Park(2), THE(2), BBC(1) |
| airlines | |
| aircrafts | Cessna_152(7), Guimbal_Cabri_G2(4) |
| flightno(?) | C152(2), FR24(1) |

[redacted] @flightradar24 @[redacted] the data was indeed collected, but there are discrepancies in the data I am trying to wor… https://t.co/5yk2wJjMoz
[redacted] 17-NOV-2017 / FR24 / Cessna 152 (G-WACG) and Guimbal Cabri G-2 (G-JAMM) collision / Aylesbury, UK https://t.co/WC1zl48gGe
[redacted] 17-NOV-2017 /Cessna 152 (G-WACG) and Guimbal Cabri G-2 (G-JAMM) Mid-air collision / Aylesbury, UK https://t.co/3lkJZrNiHb
[redacted] Läggar länkar här ... II https://t.co/seckHzxFsa II II https://t.co/NGwS5yTepY
[redacted] @[redacted] @[redacted] @[redacted] @[redacted] @[redacted] @[redacted] @[redacted] Googled gwacg come up with… https://t.co/gNZ68GJB4K
[redacted] @[redacted] Why does it look like the helicopter is back in the air? II II https://t.co/3aUZPl1c0a
[redacted] @[redacted] @[redacted] G-jamm
[redacted] UPDATE: Cessna 152 G-WACG & Guimbal Cabri G2 helicopter G-JAMM collided in mid-air near Aylesbury, NW of London tod… https://t.co/r9BCPQGueE
[redacted] @flightradar24 do you have the track for today's flight of G-WACG which was involved in a mid air crash with a helicopter?
[redacted] .@SkyNewsBreak @SkyNews FlightRadar 24 track of flight of C152 G-WACG from Wycombe before signal lost. https://t.co/kT3GHIQiMO
[redacted] Very sad to hear that there was a mid air collision involving 2 aircraft from Wycombe today. One was my old faithfu… https://t.co/8NF8UkTV1q
[redacted] The plane that crashed today in Buckinghamshire G-WACG was involved in a crash during takeoff in 1993. Due to pilot… https://t.co/28HWW85St5
[redacted] @SkyNewsBreak Aircraft was G-WACG. A 35 year old aircraft, previously involved in a crash landing in 1993
[redacted] @[redacted] G-JAMM and G-WACG, very sad 😔
[redacted] Here are the last ADS-B telemetry reports for #Waddesdon Manor collision from both aircraft. G-JAMM was a helicopte… https://t.co/QuI9BMXJ1X
[redacted] Getting information aircraft involved in mid-air collision in UKb today were a Cessna 152 (G-WACG) and a Guimbal G2… https://t.co/nqL8wGeFsT
[redacted] Here's a diagram of the flight path of Helicopter (flight number GJAMM) from High Wycombe to London - said to be a… https://t.co/imcUmQvNsr
[redacted] @[redacted] Yes it is a 2 seater - I also fly rotary. Looking at our data I think it could be G-JAMM. No sign of the C152
[redacted] @[redacted] It's actually G-WACG https://t.co/Owp9ANJ16x https://t.co/tJKj5Ot6X5
[redacted] G-JAMM a Cabri G2 looks to be the helicopter involved in the air crash.
[redacted] @airlivenet According to @[redacted] the helicopter involved in mid-air collision was Guimbal Cabri G2 (G-JAMM)… https://t.co/qk8n8ZmXN7
[redacted] @SkyNewsBreak Helicopter was owned by Heliair. Registration G-JAMM. https://t.co/HnXUw2tOLT
[redacted] G-WACG, a Cessna 152 is involved in the crash II II https://t.co/KuseAufHoh
[redacted] @[redacted] Cessna 152 G-WACG left Wycombe Air Park at 11.56 am and disappeared from radar 11.59 am. Credit: https://t.co/rPkK873RLG
[redacted] @[redacted] @[redacted] @[redacted] The helicopter may be G-JAMM
[redacted] @[redacted] G-WACG is a Cessna 152...
[redacted] As well as the registration seen on @[redacted] tail of G-WACG matches the livery seen on pictures of the Cessna 1… https://t.co/RK6enCM4Tn
[redacted] @flightradar24 @[redacted] G-Wacg and you do have a track for it
[redacted] BBC News have shown that the reg of the fixed wing aircraft was G-WACG, a Cessna 152 based in Wycombe Air Park. https://t.co/qXgiqOxscx
[redacted] @[redacted] Hi @[redacted] I'm a reporter for @PA. Could you follow me so I can send a Direct Message? I th… https://t.co/3xk00c7TfB
[redacted] @[redacted] Here is the flight: https://t.co/J2eVv4lutv
[redacted] From @flightradar24 data it looks like helicopter involved int he midair collision was G-JAMM: https://t.co/J2eVv4lutv

**Figure 6: Extract data found in tweets related to collision**

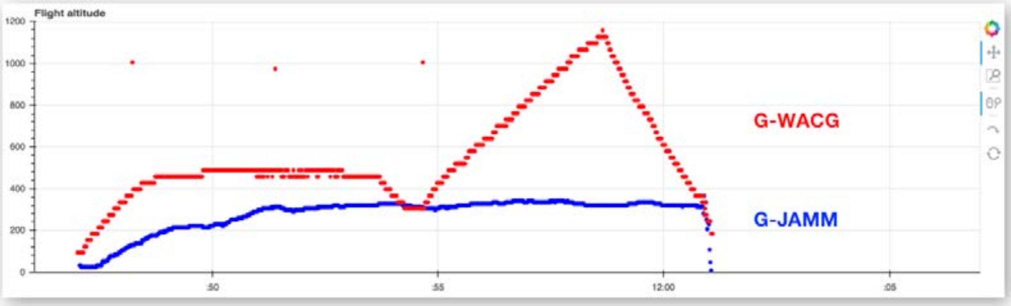


**Figure 7: Flight altitude data from OpenSky for the aircrafts involved in the collision**

As in this case the crash can be seen in the flight data of G-JAMM itself, we investigated other known accidents from the main list of the Aviation Safety Network[11]. From the base list of 335 events (2016-01-01 to 2018-01-15), we could find the icao24 identity for 151 cases. For 52 of these, there is data available in OpenSky. A manual evaluation showed that these events mostly consist of collisions with other vehicles on the ground (16), technical problems and bird strikes with normal landing (14), 'hard' landings with tail touching ground during landing (13) and runway excursions (5), all cases that would not lead to abnormal flight patterns. Only 4 cases involved fatal crashes that potentially could be seen in the flight data (2 of these being the two aircrafts involved in the Aylesbury collision).

This kind of crashes are luckily seldom, but still very relevant to be watched for in the available flight data. On the other hand, technical difficulties and bird strikes might lead to abnormal flight patterns because they deviate from planned (or historically seen) flight routes that we will evaluate in future work.

## 4.3 Alerts via Twitter content

Using Twitter content as trigger for flight related incidents is more challenging, as related information is 'hidden' in a discussion of all other topics and due to its free format. In our investigations we picked some specific combinations of hashtags and keywords for testing, like '#planecrash', '(flight OR plane) delayed', '(flight OR plane) diverted' etc.

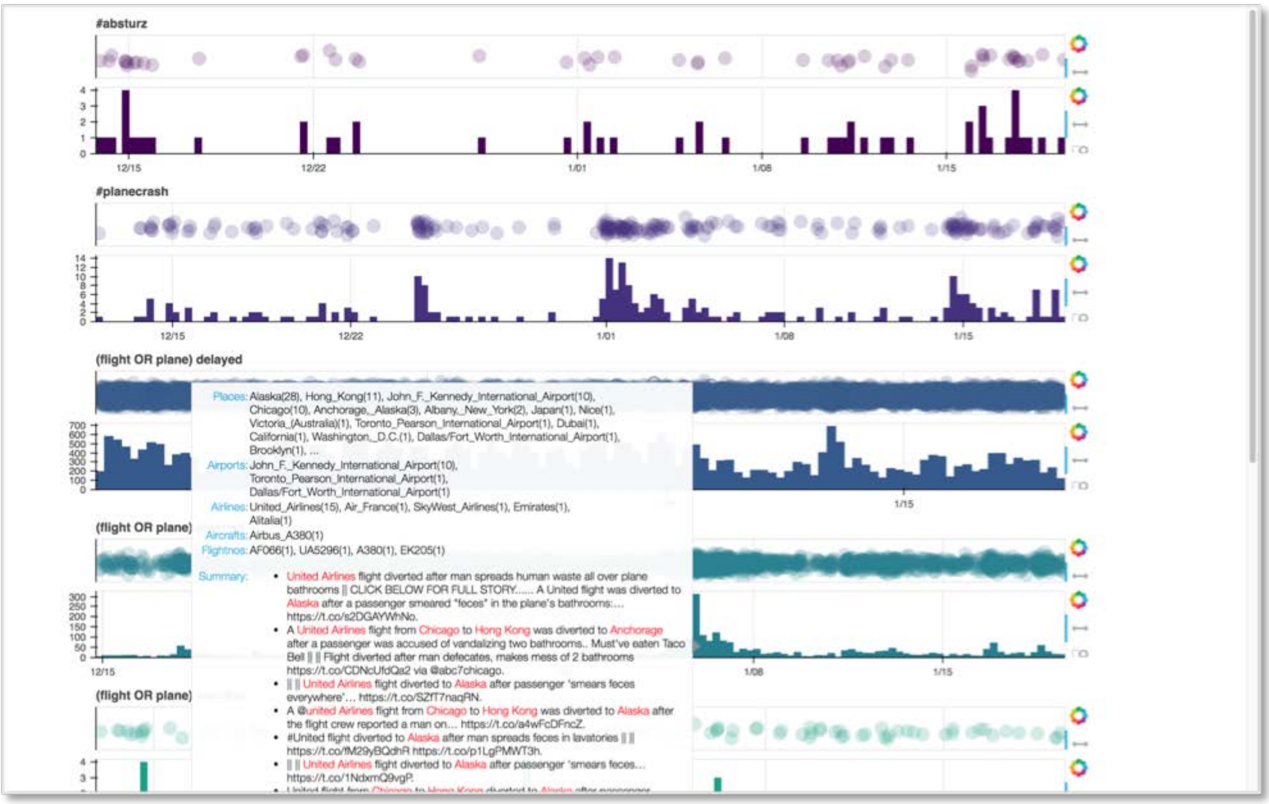
Figure 8 shows histograms of tweets with these phrases occurring over time – and for some phrases, like 'plane diverted', there are hundreds of tweets per day for different, overlapping events. This can be seen in the 'summarized event' for a specific time period in that figure – many different airports and airplanes are mentioned, though the tweets themselves seem to be dominated by one diversion of a plane on the route Chicago-Hong Kong to Anchorage due to a vandalized bathroom that seemed to have piqued the attention (and imagination) of many Twitter users.

In a different case that drew interest of Twitter users over many days, a plane on route Doha-Denpassar was doing an unplanned stop-over in Chennai due to a couple fighting after the woman found evidence for a liaison of her husband[12]. Triggered by tweets containing 'flight diverted' together with a flight number, Figure 9 shows tweets related to this event, allowing to fetch also flight data confirming the stop-over in Chennai.

More work is required to systematically and reliably extract incidents from the fuzzy Twitter data.

[11] Aviation Safety Network https://aviation-safety.net/

[12] "Plane bound for island paradise diverted after woman discovers husband's affair mid-flight" https://www.telegraph.co.uk/travel/news/flight-diverted-woman-discovered-affair/



Places: Alaska(28), Hong_Kong(11), John_F._Kennedy_International_Airport(10), Chicago(10), Anchorage,_Alaska(3), Albany,_New_York(2), Japan(1), Nice(1), Victoria_(Australia)(1), Toronto_Pearson_International_Airport(1), Dubai(1), California(1), Washington,_D.C.(1), Dallas/Fort_Worth_International_Airport(1), Brooklyn(1), ...

Airports: John_F._Kennedy_International_Airport(10), Toronto_Pearson_International_Airport(1), Dallas/Fort_Worth_International_Airport(1)

Airlines: United_Airlines(15), Air_France(1), SkyWest_Airlines(1), Emirates(1), Alitalia(1)

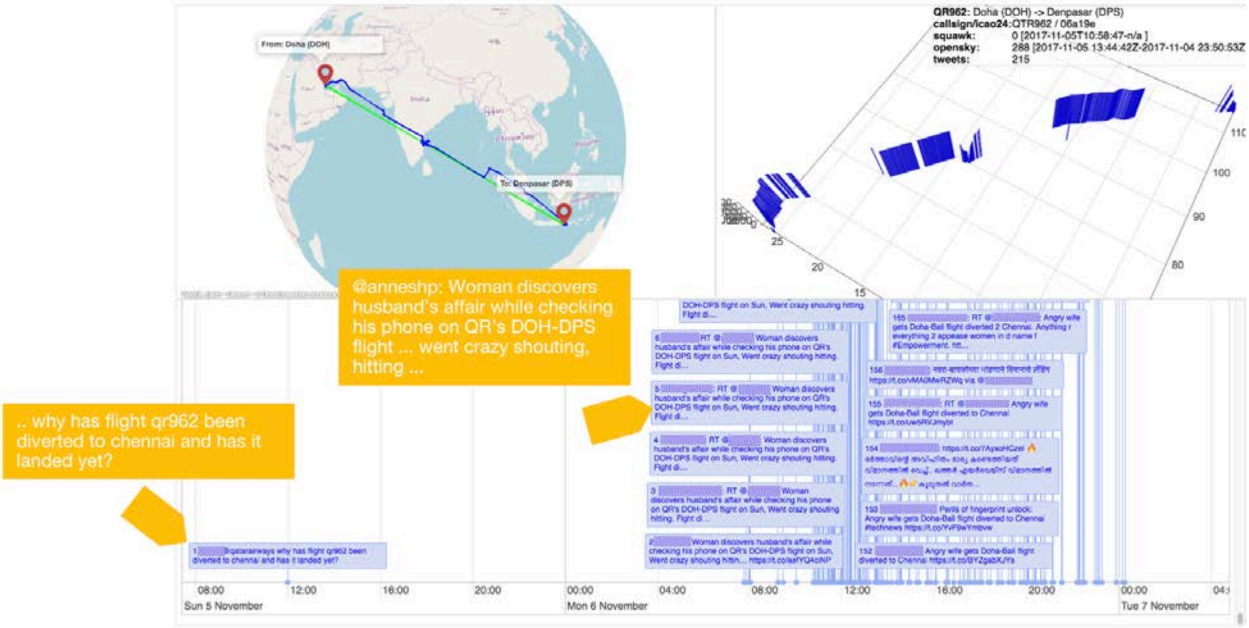
Aircrafts: Airbus_A380(1)

Flightnos: AF066(1), UA5296(1), A380(1), EK205(1)

Summary:

- United Airlines flight diverted after man spreads human waste all over plane bathrooms || CLICK BELOW FOR FULL STORY...... A United flight was diverted to Alaska after a passenger smeared "feces" in the plane's bathrooms:... https://t.co/s2DGAYWhNo.
- A United Airlines flight from Chicago to Hong Kong was diverted to Anchorage after a passenger was accused of vandalizing two bathrooms.. Must've eaten Taco Bell || || Flight diverted after man defecates, makes mess of 2 bathrooms https://t.co/CDNcUfdQa2 via @abc7chicago.
- || || United Airlines flight diverted to Alaska after passenger 'smears feces everywhere'... https://t.co/SZfT7naqRN.
- A @united Airlines flight from Chicago to Hong Kong was diverted to Alaska after the flight crew reported a man on... https://t.co/a4wFcDFncZ.
- #United flight diverted to Alaska after man spreads feces in lavatories || || https://t.co/fM29yBQdhR https://t.co/p1LgPMWT3h.
- || || United Airlines flight diverted to Alaska after passenger 'smears feces... https://t.co/1NdxmQ9vgP.

**Figure 8: Frequency of selected phrases in tweets over time**



**QR962**: Doha (DOH) -> Denpasar (DPS)
**callsign/icao24:** QTR962 / 06a19e
**squawk:** 0 [2017-11-05T10:58:47-n/a ]
**opensky:** 288 [2017-11-05 13:44:42Z-2017-11-04 23:50:53Z
**tweets:** 215

@anneshp: Woman discovers husband's affair while checking his phone on QR's DOH-DPS flight ... went crazy shouting, hitting ...

.. why has flight qr962 been diverted to chennai and has it landed yet?

1 [redacted] @qatarairways why has flight qr962 been diverted to chennai and has it landed yet?

DOH-DPS flight on Sun, Went crazy shouting hitting. Flight di...

6 [redacted] RT @[redacted] Woman discovers husband's affair while checking his phone on QR's DOH-DPS flight on Sun, Went crazy shouting hitting. Flight di...

5 [redacted]: RT @[redacted] Woman discovers husband's affair while checking his phone on QR's DOH-DPS flight on Sun, Went crazy shouting hitting. Flight di...

4 [redacted] RT @[redacted] Woman discovers husband's affair while checking his phone on QR's DOH-DPS flight on Sun, Went crazy shouting hitting. Flight di...

3 [redacted] RT @[redacted] Woman discovers husband's affair while checking his phone on QR's DOH-DPS flight on Sun, Went crazy shouting hitting. Flight di...

2 [redacted] Woman discovers husband's affair while checking his phone on QR's DOH-DPS flight on Sun, Went crazy shouting hittin... https://t.co/aafYQ4oINP

165 [redacted]: RT @[redacted]: Angry wife gets Doha-Bali flight diverted 2 Chennai. Anything r everything 2 appease women in d name f #Empowerment. htt...

156 [redacted] [illegible] https://t.co/vMA0MwRZWq via @[redacted]

155 [redacted]: RT @[redacted] Angry wife gets Doha-Bali flight diverted to Chennai https://t.co/uw5RVJmybt

154 [redacted] https://t.co/7AyxoHCzel 🔥 [illegible] 🔥 [illegible]

153 [redacted] Perils of fingerprint unlock: Angry wife gets Doha-Bali flight diverted to Chennai #technews https://t.co/YvF9wYmbvw

152 [redacted] Angry wife gets Doha-Bali flight diverted to Chennai https://t.co/BYZgabXJYs

**Figure 9: On 2017-11-05, a plane was diverted to Chennai airport on route from Doha to Denpassar**

## 5. DISCUSSION

Our experiments show that it is in principle possible to bridge data sources of even very different characteristics and that enriched information can be gained, though due its nature (data aimed for specialists versus colloquial free text by and for human consumption) there is a semantic gap. Having multiple data sources allows to identify arising situations from multiple triggers that observe different viewpoints. Triggered from either data source, a more detailed data capture can be started that would otherwise not be possible due to the volume of the data. The examples analyzed show that both data sources enrich each other for arising situations, providing a more complete picture of the event, although finding and evaluating data in Twitter remains a challenge due to its fuzzy nature of free natural language (often in sloppy and abbreviated terms) and the semantic ambiguity of many terms of interest.

In future work we are interested in a more sophisticated Twitter event analysis and understanding using more general machine learning techniques, a better analysis of flight path anomalies and a possible further enrichment with additional data sources.

## 6. REFERENCES

[1] F. Castanedo, “A review of data fusion techniques,” *ScientificWorldJournal*, vol. 2013, pp. 704504–704504, 2013.

[2] V. Lenders, A. Tanner, and A. Blarer, “Gaining an Edge in Cyberspace with Advanced Situational Awareness,” *Security Privacy, IEEE*, vol. 13, no. 2, pp. 65–74, Mar. 2015.

[3] L. Snidaro, J. Garcia, J. Llinas, and Blasch, Erik, Eds., *Context-enhanced information fusion*. New York, NY: Springer Berlin Heidelberg, 2016.

[4] A. Akbar *et al.*, “Real-Time Probabilistic Data Fusion for Large-Scale IoT Applications,” *IEEE Access*, vol. 6, pp. 10015–10027, 2018.

[5] Z. Zhang, “Fusing Social Media and Traditional Traffic Data for Advanced Traveler Information and Travel Behavior Analysis.”