

## DELIVERABLE SUBMISSION SHEET

**To:** Susan Fraser *(Project Officer)*  
EUROPEAN COMMISSION  
Directorate-General Information Society and Media  
EUFO 1165A  
L-2920 Luxembourg

**From:**  
Project acronym: PHEME Project number: 611233  
Project manager: Kalina Bontcheva  
Project coordinator The University of Sheffield (USFD)

### The following deliverable:

Deliverable title: Digital Journalism Prototype  
Deliverable number: D8.3  
Deliverable date: 30 June 2016  
Partners responsible: i-Hub Ltd.  
Status:  Public  Restricted  Confidential

is now complete.  It is available for your inspection.  
 Relevant descriptive documents are attached.

### The deliverable is:

- a document
- a Website (URL: .....)
- software (.....)
- an event
- other (Prototype.....)

Sent to Project Officer: <a href="mailto:Susan.Fraser@ec.europa.eu">Susan.Fraser@ec.europa.eu</a>	Sent to functional mail box: CNECT-ICT-611233 @ec.europa.eu	On date: 18 July 2016
--	--	--------------------------



---

## D8.3 Digital Journalism Prototype

---

David Losada (iHub Ltd.)

Sebastian Mitchell (iHub Ltd.)

Peter Tolmie (University of Warwick)

Christian Burger (SWI swissinfo.ch)

Geraldine Wong Sak Hoi (SWI swissinfo.ch)

### Abstract

FP7-ICT Strategic Targeted Research Project PHEME (No. 611233)

Deliverable D8.3 (WP 8)

The digital journalism dashboard prototype is now available online for perusal. This document contains the URL for the prototype, along with a short manual that guides users through the dashboard in its current state of development. The document also describes the technical approach to creating the prototype, plus relevant background information, chiefly on the user evaluations that have informed the development of the current prototype, the primary use cases, and a list of the key dashboard features that have been prioritised.

---

**Keyword list:** journalism dashboard, prototype, rumour verification, rumours

---

Nature: **Prototype and Report**

Dissemination: **PU**

Contractual date of delivery: **30 June 2016**  
2016

Actual date of delivery: **18 July**

Reviewed By:

Web links: <http://pheme-jd.usahidi.com>



## **PHEME Consortium**

This document is part of the PHEME research project (No. 611233), partially funded by the FP7-ICT Programme.

### **University of Sheffield**

Department of Computer Science  
Regent Court, 211 Portobello St.  
Sheffield S1 4DP, UK  
Tel: +44 114 222 1930  
Fax: +44 114 222 1810  
Contact person: Kalina Bontcheva  
E-mail: [K.Bontcheva@dcs.shef.ac.uk](mailto:K.Bontcheva@dcs.shef.ac.uk)

### **Universitaet des Saarlandes**

Language Technology Lab  
Campus  
D-66041 Saarbrücken  
Germany  
Contact person: Thierry Declerck  
E-mail: [declerck@dfki.de](mailto:declerck@dfki.de)

### **MODUL University Vienna GMBH**

Am Kahlenberg 1  
1190 Wien  
Austria  
Contact person: Arno Scharl  
E-mail: [scharl@modul.ac.at](mailto:scharl@modul.ac.at)

### **Ototext AD**

Polygraphia Office Center fl.4,  
47A Tsarigradsko Shosse,  
Sofia 1504, Bulgaria  
Contact person: Georgi Georgiev  
E-mail: [georgiev@ototext.com](mailto:georgiev@ototext.com)

### **ATOS Spain SA**

Calle de Albarracin 25  
28037 Madrid  
Spain  
Contact person: Tomás Pariente Lobo  
E-mail: [tomas.parientalobo@atos.net](mailto:tomas.parientalobo@atos.net)

### **King's College London**

Strand  
WC2R 2LS London  
United Kingdom  
Contact person: Robert Stewart  
E-mail: [robert.stewart@kcl.ac.uk](mailto:robert.stewart@kcl.ac.uk)

### **iHub Ltd.**

NGONG, Road Bishop Magua Building  
4th floor  
00200 Nairobi  
Kenya  
Contact person: Rob Baker  
E-mail: [robbaker@ushahidi.com](mailto:robbaker@ushahidi.com)

### **SwissInfo.ch**

Giacomettistrasse 3  
3000 Bern  
Switzerland  
Contact person: Peter Schibli  
E-mail: [Peter.Schibli@swissinfo.ch](mailto:Peter.Schibli@swissinfo.ch)

### **The University of Warwick**

Kirby Corner Road  
University House  
CV4 8UW Coventry  
United Kingdom  
Contact person: Rob Procter  
E-mail: [Rob.Procter@warwick.ac.uk](mailto:Rob.Procter@warwick.ac.uk)

## Contents

- [1 Introduction](#)
- [2 Relevance to PHEME](#)
  - [2.1 Relevance to project objectives](#)
- [3 The digital journalism prototype: User's guide](#)
  - [3.1 Listing and Creating Events](#)
  - [3.2 Browsing Themes detected from Events](#)
- [4 The digital journalism dashboard: Technical approach](#)
  - [4.1 Components](#)
- [5 Use cases, user evaluations and key prototype features](#)
  - [5.1 Journalism use cases](#)
  - [5.2 User evaluations](#)
    - [5.2.1 First Formative Evaluation](#)
    - [5.2.2 Second Formative Evaluation](#)
  - [5.3 Priorities for the prototype](#)
  - [5.4 Pending short-term work items](#)
- [6 Conclusion](#)
- [Bibliography](#)

## **1 Introduction**

The open-source digital journalism prototype aims to harness the systems being developed within PHEME and present them in a dashboard geared specifically at journalists looking to quickly locate and verify information online.

The development of the dashboard is being led by iHub Ltd., which is extending the open-source Ushahidi platform with PHEME technology for veracity and socio-semantic intelligence. Indeed iHub is collaborating closely with, and depends greatly on the work of, technical partners (ATOS, ONTO, USFD, UWAR and USAAR) and the use case partner (SWI) to achieve its task objective.

Taking the user requirements and interface mock-ups from T8.1 as its point of departure, iHub is developing the dashboard in an iterative process involving regular discussions with project partners, adjustments based on the status of various PHEME technologies, and outcomes of user evaluations completed so far.

Following this introduction and a brief look at the relevance of D8.3 to other work packages, Section 3 of this document provides a short user manual designed to guide the user through the current prototype. Section 4 describes the technical elements supporting the dashboard development. Section 5 is dedicated to background information, namely the outcomes of the two user evaluations that have informed the design and functionalities of the dashboard, the primary use cases that have served as the basis for the evaluations, and the dashboard features that have been prioritised for the prototype based on in-depth discussions at the PHEME plenary meetings.

## **2 Relevance to PHEME**

### **2.1 Relevance to project objectives**

Task 8.3 is at the core of the project use case, as it seeks to transpose the project results into a real world setting by creating a digital journalism showcase for the modelling, identification and verification of rumourous stories spreading on social media networks.

### **2.2 Relation to other work packages**

Task 8.3 is closely linked to, and highly dependent on, numerous other work packages that are concerned with the development of specific technologies being harnessed by the journalism dashboard. These are in particular: WP2 (Ontological modelling, multilinguality, spatio-temporal grounding), WP3 (Contextual interpretation), WP4 (Detecting rumours and veracity), WP5 (Interactive visual analytics dashboard) and WP6 (Scalability, integration and evaluation).

## **3 The digital journalism prototype: User's guide**

The prototype is available at <http://pheme-jd.ushahidi.com> . In order to perform any operations at this point, the user must be logged in with the following credentials:

email address: [demo@ushahidi.com](mailto:demo@ushahidi.com)

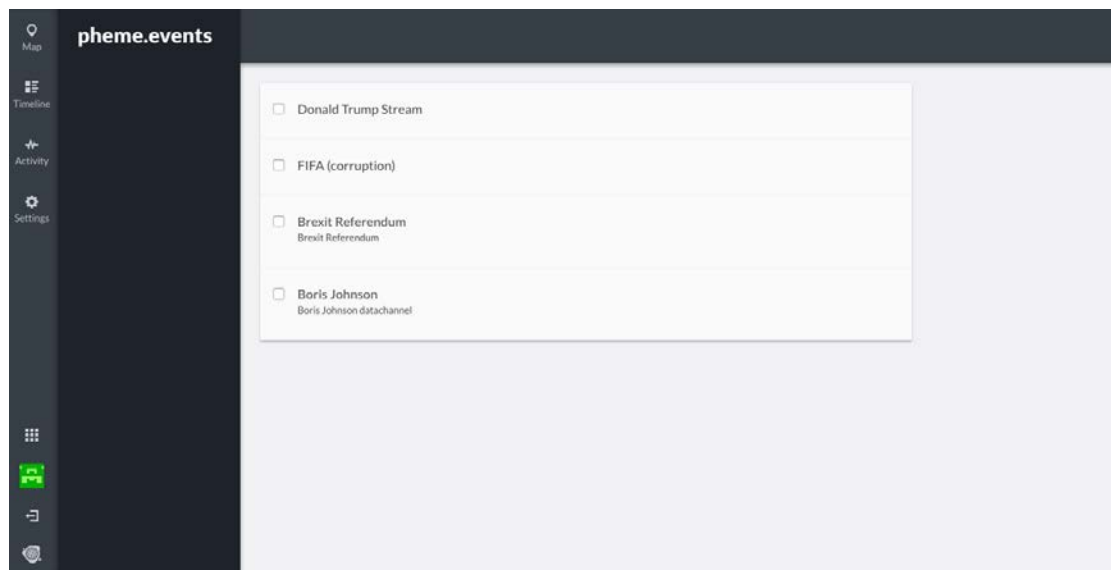
password: fp7Pheme

In its current state, the prototype is more a display of data integration capabilities than actual functionality. The general user experience in terms of information display and navigation is, objectively speaking, far from ideal. However, there are a couple actions that can be performed. These are described below.

### 3.1 Listing and Creating Events

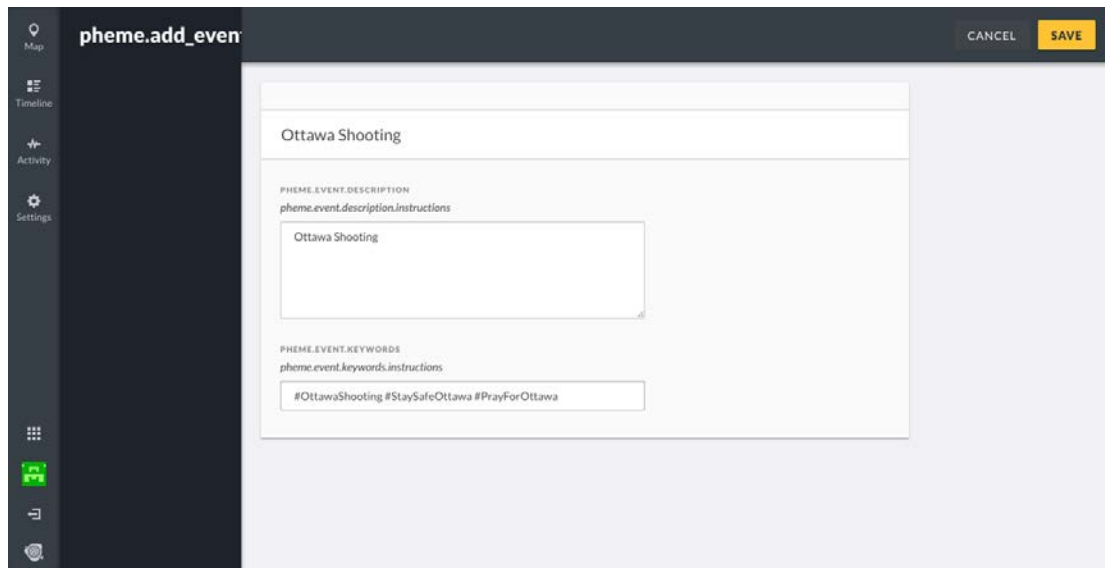
An Event refers to a persistently running search over social media platforms. When a journalist decides to find more information about a real-world event that is unfolding, he/she will create an Event in the dashboard and provide keywords associated with the search in question.

To see the list of Events currently being streamed into the system, the user can visit the following URL: <http://pheme-jd.ushahidi.com/pheme/events>



In the above screen shot, four Events are registered in the system.

In order to create an Event, the user will need to visit this URL: <http://pheme-jd.ushahidi.com/pheme/events/create>



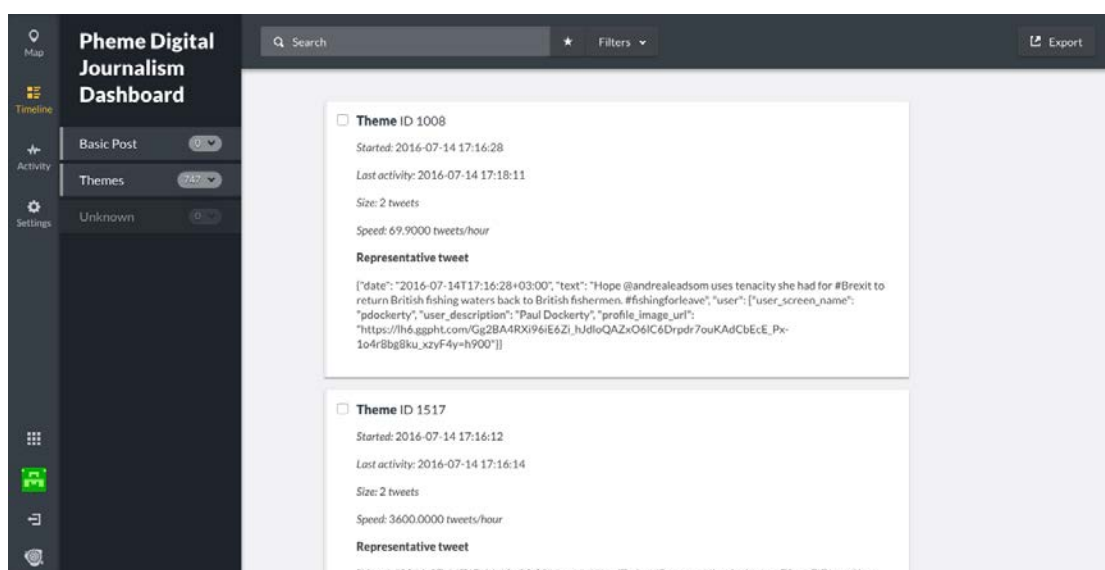
Please note that all fields as shown in the screen shot above must be completed.

After clicking “Save”, the search parameters will be sent to the system and queries across social media will be initiated. Any messages from that moment on will be captured by the system.

### 3.2 Browsing Themes detected from Events

Themes are groups of messages that the system has identified to be related to each other. This grouping can be useful in identifying trends and allowing the user to “cut through the noise” and find the more relevant and rumorous conversations more quickly.

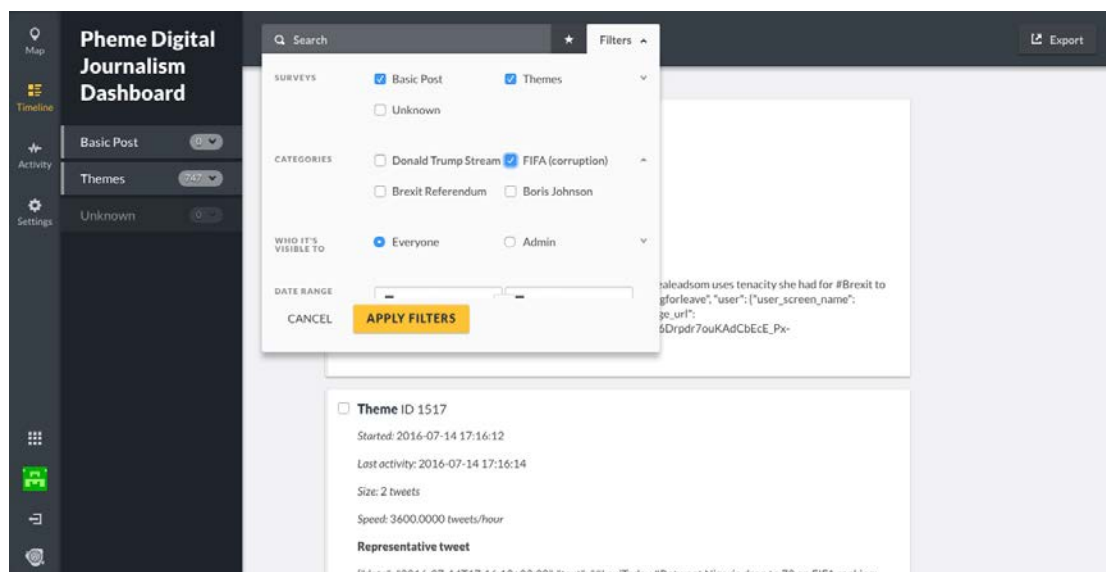
To see the list of Themes, the user can visit this URL: <http://pheme-jd.usahidi.com/pheme/posts/themes/list>



The screen displays a list of the Themes that have been detected by the system and the summary information about them that is available to the journalism dashboard. This information includes:

- Theme ID number: to be used as a persistent means of identification (other information, such as size, authors, speed and so on, can be updated as messages are added to the Theme). In a future version of the dashboard, this ID will be replaced by a more meaningful, descriptive summary (e.g. most relevant keywords).
- Started: date and time of the oldest message in the Theme
- Last activity: date and time of the most recent message in the Theme
- Size: number of messages in the Theme
- Speed: rate at which messages have been added to the Theme
- Representative tweet: a message selected from among the group of messages as a representative account of the content of the Theme. In a future version of the dashboard, this tweet will be formatted in a more comprehensible manner.

It is possible at this point to do some filtering over the list, in order to see the Themes that belong to one (or a subset) of the Events registered in the system. Some of the displayed element (e.g. “Basic Post”, “Unknown”) are yet artifacts of the standard Ushahidi implementation and will be removed or repurposed in future versions.



After clicking “Apply Filters” the user will see only the Themes belonging to the selected Event.

## 4 The digital journalism dashboard: Technical approach

### 4.1 Components

The components that support the operation of the digital journalism dashboard can be divided into three categories:



- **Components for conversation collection and analysis.** As delivered and integrated through other work packages in the project.
- **The digital journalism dashboard web application.** This is a modified version of the latest Ushahidi Platform. It consists of the following components:
  - The database server. This is a standard relational database, more specifically MySQL. It stores the data that is most frequently displayed in the dashboard as well as the profiles and information associated with users of the system.
  - The server side component (or API). This is web request processing code, coded in PHP following the MVC pattern. Modifications have been introduced to communicate with the rest of the system through the interface component detailed below.

Code can be found at <https://github.com/ushahidi/project-pheme-platform>.

- The client side application, which loads in the user browser. This is a fully dynamic web application running in the browser, coded in Javascript, HTML and CSS with AngularJS, and following the MVC pattern. It is pre-designed to work well in desktop, laptop and tablet computing devices.

This component is undergoing more extensive customisation, in order to adapt the presentation of information and interaction model of the application to the necessities of the collection and analysis package, as well as those of the journalists potentially using the system.

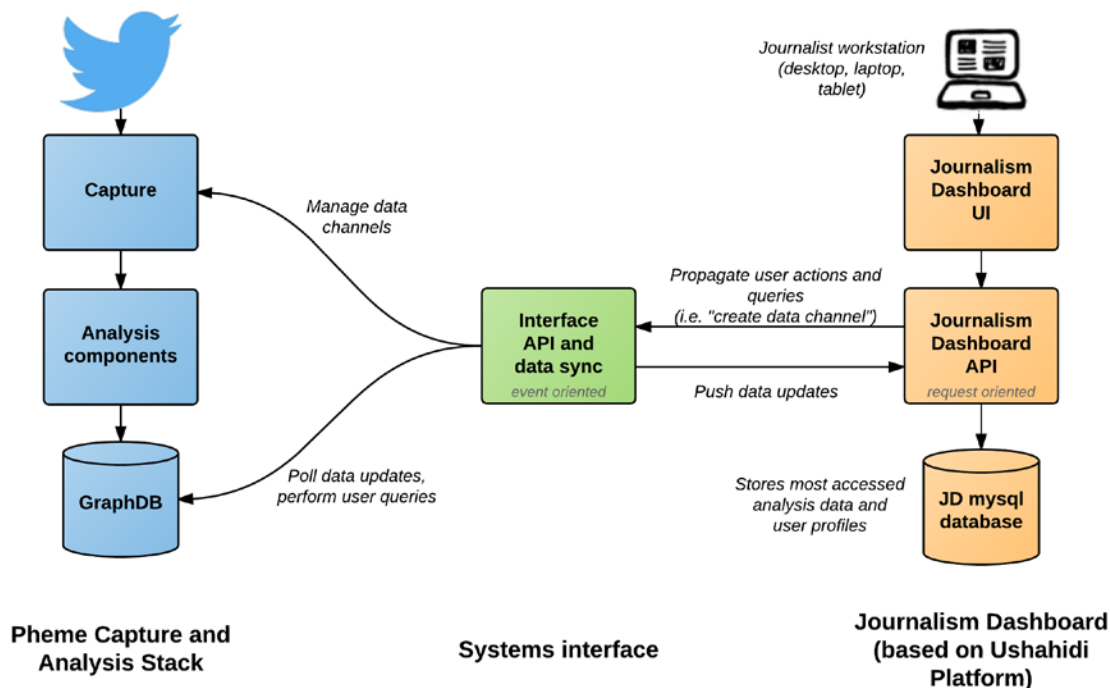
Code can be found at <https://github.com/ushahidi/project-pheme-platform-client>.

- **Interface components.** This component handles the communication with capture and analysis stack elements, in order to track the evolution of the events, threads and rumours. Thanks to this approach, no significant architectural changes need to be introduced into the Ushahidi platform code.

This component is a continuously running application, built around an event loop, using Python and the Tornado framework.

Code can be found at <https://github.com/ushahidi/project-pheme-data-interface>.

The following diagram illustrates the components and their relationships.



## 5 Use cases, user evaluations and key prototype features

### 5.1 Journalism use cases

The journalism dashboard needs to be able to support two principal use cases:

- Working on the newsdesk, where journalists are less interested in conversation history, origins, and false veracity;
- Writing features, where the conversation history, origins, and false veracity *are* of interest.

#### 1. Primary case: On the newsdesk – searching for story support

- A potential story is identified using the wires or other additional sources.
- The journalist uses the dashboard to locate further information
- This is done by searching, using relevant terms
  - The default results show the latest related tweets at top
- The journalist clicks on the tweets to see the associated conversational threads
- The tweets are able to be sorted in various ways, e.g.:
  - By veracity – the most likely to be true at the top
    - *A graph shows how confidence in veracity has changed over time*
  - By author proximity to the event
- The journalist is also able to use the map view
  - This locates where tweets were generated
  - A mouse-over of markers on the map shows the tweet content and author details
- The tweets can also be filtered, e.g.:
  - To remove single tweet threads from un-trusted authors

- To display only tweets mentioning a location or an event
- Tweets of interest can be copied to a Word document
- The journalist can also view associated news reports
  - Mouse-over for these shows the initial text
  - Content from these can also be copied and pasted to a Word document
- It is also possible to view pictures and videos coming out of the body of tweets
  - Here too a mouse-over shows author details and their location
  - The images can be saved for possible later use

### *1a. Related Case – On the newsdesk, looking for possible stories:*

- The journalist is looking for newsworthy stories
- In addition to the features presented in Use Case 1, the journalist is able to:
  - Focus on what tweets and conversations are trending
    - Here a graphical display shows how the popularity of the tweet has changed over time
  - Have the tweets pre-filtered to match pre-specified terms
  - Identify trustworthy tweets from trusted news organizations
  - Display author details relating to twitter handles embedded in the tweets
  - Filter tweets with factual content and highlight facts within them
  - Expose the actual annotations underlying the veracity assessments
  - Mark up specific tweets of interest and share saved sets with other journalists and editors

### *2. Subsidiary Case: Writing features*

- In this use case, in addition to the features presented above, the journalist is able to:
  - Specifically unpack the evolution of stories and rumours by seeing the original tweet and mapping how popularity, veracity assessments, etc. changed over time
  - Engage in a detailed review of conversational threads
  - Display the controversy associated with particular rumours (the extent to which opposing views were presented) and how it changed over time

## **5.2 User evaluations**

### *5.2.1 First Formative Evaluation*

The first formative evaluation took place in January 2016. It involved 2 journalists from within SWI, with a separate session for each of them. They both received the same standardized instructions before beginning to solve the assessed tasks. All Evaluations were recorded with a screen capturing software and an external video camera.

The evaluation was conducted using canned data from an annotated dataset that predominantly related to the Germanwings crash in March 2015. It pursued an instructed process mimicking the basic journalistic workflow (as specified in the defined use cases) and working between the wires, the dashboard, associated reports and images, and a Word document. They were asked to use the dashboard as though they were looking for a story to prepare.

### *Findings*

#### *Relation to Journalistic Workflow and Resources:*

An overriding feature of work on the newsdesk is pressure of time. As a consequence journalists need to know the exact state of affairs right now. This has implications for the liveness of data and the frequency of updates.

With regard to the placement of the dashboard within the workflow, it was considered to be a useful adjunct to the wires and a potential quick route to associated reports. It was also seen to offer potential support for story justification when proposing stories to editors.

In terms of its interoperability with other tools and resources, the use of the dashboard in tandem with the wires, MS Word, and a browser all worked well. It was also considered to be complementary to, and distinct from, other social media platforms or tools.

#### *Features:*

A key point picked up on was the presence of a map. This was seen as being potentially of high value if it can rapidly display where tweets are coming from. Journalists would like to be able to use the map to select a subset of tweets and drill down. Generally there is a strong interest amongst journalists in being able to see tweets according to event proximity. Journalists would also like to see author information attached to tweet labels on the map. In general terms, for the localization of stories journalists want to be able to sort according to event proximity more broadly within the tool.

Another feature that was deemed to be of definite value was the access the dashboard can give to associated reports. Journalists would like to be able to localize these as well. They would also like to see indicative information regarding the reports prior to actually opening them so as to assess through relevance, perhaps via mouse-over in a similar way to the Twitter interface.

Related pictures and videos captured from the tweets were also considered useful: as a possible source of additional detail; and as a possible source of actual images to use. Images will need labelling with source and author information to help to assess their reliability, perhaps once again via mouse-over.

The potential value of having a conversation history was less clear-cut. It was felt to be of most use when there is uncertainty about a topic. It was also seen to be potentially useful for identifying the source of a rumour. It may also help with the identification of potential witnesses and experts. Time constraints, however, make its use more evident for features.

This version of the dashboard also presented general author details, though not always to the satisfaction of the journalists. Journalists need to be able to see all available information in the tool itself. There would be an overhead if they had to go to Twitter get and, once in Twitter, they might as well stay in Twitter. It was also noted that they need to be able to see information about people who are mentioned *within* tweets, perhaps yet again by using a mouse-over display in a similar fashion to Twitter.

Being able to add user comments was the least well-received feature. Journalists are more likely to interact with others about the content using existing preferred channels, e.g.: face-to-face; email; or messenger/chat.

### *Noted Requirements:*

Some requirements that were not currently available were mentioned. One of these was being able to embed part or all of a tweet within a story. They would ideally like interoperability with their existing publishing tools. They would also like to be able to click on tweets copied to Word for compilation and, by doing so, be able to return to the thread in the dashboard, for instance as a means of checking the latest status of an unfolding thread.

Being able to sort and filter was particularly emphasized. This is a key requirement, with two parts: 1) the shaping of the landing page with pre-specified filters, much as they can already do with the wires; 2) being able to apply additional filters. Sort functionality should cover the following elements: Time Order; Veracity; Tweeter Location in Relation to an Event; Number of Followers; Language; Number of Contributions to the Tweet Thread. The preferred defaults for sorting are time order or veracity. They would also like to be able to sort within already filtered lists.

Having search functionality is the most important requirement of all. Journalists need to be able to search on compound as well as single terms. Searching also needs to return results quickly if journalists are working on the newsdesk. Here too they need to be able to search within filtered lists.

For the purposes of veracity assessment journalists would like to have general confidence measures together with a display of changes in confidence over time. A graphical display is preferred for this. They also need to be able to organize tweets in relation to trusted sources, i.e. other trustworthy news organizations. Additionally, foregrounding of factual content (names, numbers, specific facts, etc) would be useful, together with a capacity to sort accordingly. They would also like to be able to inspect the source of veracity assessments upon request, though this is a particularly challenging requirement.

With regard to the popularity of various tweets, they would like to have an indication of tweets retweeted, favourited, etc. and trends in tweet/thread popularity over time. Once again, a graphical display would be preferred.

One other noted requirement was being able to preserve a use trail. For this they would like to be able to mark-up specific tweets in the dashboard (for instance, by starring them). They would also like to be able to build collections of specific sets of tweets that can be returned to later for justification, verification, etc. Furthermore, they would like to be able to share links to these sets with other people, such as editors.

### *5.2.2 Second Formative Evaluation*

The second formative evaluation took place in April 2016, using the same 2 journalists from within SWI. Some technical issues and delays in components required us to adjust the goals and instructions such that the construction of a full story was no longer possible. Nonetheless, a range of important issues surfaced that had not been visible in the first evaluation (but it also confirmed previous requirements). As a consequence the second formative evaluation led to significant technological refinement of the planned prototype.

### *Findings*

#### *Liveliness:*

During the evaluation issues within the pipeline meant that refresh of the data was patchy. As news is, by definition, trading in the latest information, the currency of the tweets visible via the dashboard is of critical importance. A system that doesn't get updated on a regular basis and renders these updated visible quickly loses its value in a journalistic context.

### *Search:*

The absence of search or filtering functionality meant that the stream was very generic and undifferentiated. The automatically provided keywords to identify tweet clusters (topics) were unhelpful and the journalists couldn't see the connection between them. As a consequence a lot of time was wasted sifting through the clusters and trying to make sense of the contained tweets. In a future version, the data will have to become better organized and more descriptive.

### *Coherence:*

The tweets were grouped together according to a relationship algorithm but the grounds of the relationship were opaque to the users and the groupings therefore confusing. There needs to be a more coherent grouping strategy or the operation of the algorithm needs to be made more transparent.

### *Temporal cohesion:*

As a shortcoming of the Twitter Search API the grouping of tweets to conversations is not happening properly and it seems to be the case that tweets falling outside of a certain time window are very hard to be grouped as part of the same conversation. As being able to see conversations is core to being able to properly identify unfolding rumours and their spread this issue will need to be solved by the consortium as good as possible.

### *5.2.3 Future evaluation*

A third formative evaluation will take place in September 2016 once the core elements discussed above (liveness; coherence; cohesion and searchability) are operational.

## **5.3 Priorities for the prototype**

At the project plenary meeting which took place in Vienna in May 2016, PHEME partners discussed the outcomes of the formative evaluations and the short-term capabilities for implementing certain key dashboard features.

Partners agreed to prioritise the following features:

- A search function where users can enter keywords to pull up results about a particular news event (so-called data channel), and filter these results by entering additional search terms
- Ability to sort results by size of conversation thread/story (based on number of re-tweets), veracity/controversy score, and by chronology
- Map-based location-relevant conversation threads

- Ability to filter tweets by verified accounts, number of followers, oldest/newest accounts, tweets with the most re-tweets (eventually also filter by trusted users, based on list provided by journalists)
- Display information on the tweet author, to be taken from Twitter user profile.

### 5.3.1 *Re-evaluation of priorities*

After the meeting in Vienna, it became clear that the implementation of the information system that is necessary to support the prioritised features was going to take longer than initially expected. The development of the dashboard is highly dependent on components being developed in other work packages, and some of these have gone through recent revisions for improvements.

Also, from the side of iHub, a decision was made to invest time in re-evaluating the implementation plan for the journalism dashboard. As a result, the implementation is now based on the latest version of the Ushahidi Platform (v3.4 and higher) instead of the older one (v2.x). This carries a number of benefits:

1. A code architecture that is significantly cleaner and faster to develop on. Adding features and modifying user interface screens and components can be done much faster on the new version of the platform.
2. Cleaner separation between the code written for the project and the base code, allowing the resulting dashboard to enjoy updates and bug fixes made on the Ushahidi Platform itself.
3. Faster prototyping of changes, in order to accommodate changing demands and realities among partners and prospective users.
4. Responsive user interface design, with support out of the box for a higher number of computing devices, such as tablets and smartphones.

As a result, we now have a more modern codebase than initially envisioned, but only the first priority among those discussed in Vienna has been met so far (search function so users can enter keywords about a particular news event). The other priorities will be met in the short term.

## 5.4 Pending short-term work items

In order to fulfill the priorities for the prototype mentioned above, we are planning the following next steps:

- Work with consortium partners to obtain access to each of the metadata items deemed important by the journalists.
- Finish properly lining up the layout and navigation of the site, to make it first minimally functional and, in a second instance, adapted to an intuitive and optimal workflow for the journalists.
- Cosmetic changes on the presentation of data, making it dramatically easier to interpret the information displayed in the screen and follow relationships between data items.
- Enhance the sorting and filtering capabilities of the dashboard. This will cover filtering by size, veracity/controversy score, chronology, verified accounts, number of followers etc, just as outlined in the priority list above.

### D8.3 / Digital Journalism Prototype

- Work with partners to make location information available, and connect it to the map feature that comes out of the box with the Ushahidi platform.
- Integrate the future conversation thread insight capabilities, to be provided by the capture and analysis system.

To this effect the partners involved in developing the journalism dashboard will be meeting for a two-day hackathon in late July 2016.

## **6 Conclusion**

The development of the digital journalism dashboard has been and continues to be dependent on many factors: the primary use cases identified in the early stages of the project; the availability and status of various PHEME technologies being developed by technical partners; the outcomes of formative evaluations undertaken at SWI and outcomes of regular discussions with partners.

At the time of delivering this initial prototype, partners had agreed on the key features that need to be prioritised if the consortium is to offer a dashboard that meets journalists' main requirements at the end of the project. A system is now in place where users can view a basic dashboard containing a set of Events and Themes, and enter keywords into a search box to display additional Events. All partners are working hard to make sure all pending components are functioning as needed in order for iHub to implement additional features. Partners continue to have regular discussions both by email and via conference calls, and plan to meet in person for a two-day hackathon in late July.



## **Bibliography**

Burger, C. et al. (2015). *Requirements gathering, use case design and interface mock-ups V2*. PHEME Deliverable D8.1a.