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

Computing Veracity Across Media, Languages, and Social Networks

D8.1 Requirements gathering, use case design and interface mock-ups

Rob Procter (University of Warwik)

Peter Tolmie (University of Warwik)

Christophe Bruttin (SWI swissinfo.ch)

Christian Burger (SWI swissinfo.ch)

Abstract

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

Deliverable D8.1 (WP 8)

This document describes the applied method for requirements gathering, the results of the first iteration, the mock-ups derived from these requirements and the journalistic use case design.

Keyword list: work requirements, journalistic use case design, interface mock-ups

Nature:Report

Dissemination:CO

Contractual date of delivery: 30 April 2014
Actual date of delivery:7 May 2014

Research partially supported by the European Union under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme, grant PHEME (611233)

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

Universitaet des Saarlandes
Language Technology Lab
Campus
D-66041 Saarbrücken
Germany
Contact person: Thierry Declerck
E-mail: declerck@dfki.de

MODUL University Vienna GMBH
Am Kahlenberg 1
1190 Wien
Austria
Contact person: Arno Scharl
E-mail: scharl@modul.ac.at

Ontotext AD
Polygraphia Office Center fl.4,
47A Tsarigradsko Shosse,
Sofia 1504, Bulgaria
Contact person: Georgi Georgiev
E-mail: georgiev@ontotext.com

ATOS Spain SA
Calle de Albarracin 25
28037 Madrid
Spain
Contact person: Tomás Pariente Lobo
E-mail: tomas.parientelobo@atos.net

King's College London
Strand
WC2R 2LS London
United Kingdom
Contact person: Robert Stewart
E-mail: 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

SwissInfo.ch
Giacomettistrasse 3
3000 Bern
Switzerland
Contact person: Peter Schibli
E-mail: Peter.Schibli@swissinfo.ch

The University ofWarwick
Kirby Corner Road
University House
CV4 8UW Coventry
United Kingdom
Contact person: Rob Procter
E-mail: Rob.Procter@warwick.ac.uk

Executive Summary

This report provides the journalistic use case scenarios as well as the basic set of requirements of newroom journalists for the final (journalistic) Pheme product.

The requirements are structured around the different tasks in the journalistic work process:

- Uncovering stories
- Selecting stories
- Verification
- Handling supposition, speculation and rumour
- Writing up stories
- Revision and re-verification

From the requirements, the interface mock-ups were derived, with emphasis on the two speed newsroom work that is common in todays work practice.

Contents

1 Introduction

The development of the World Wide Web offered journalists and news organisationsa new source of content. While better search engines helped tremendously in finding relevant content, Social Media later appeared as a magical tool to point to what mattered, even if it got published in the dark corners of the internet. Social Media quickly took on a life of its own, not merely pointing to content but also creating content of importance and shaping the discussion around it. From a solution to better sort information and quickly get confirmation and expertise, Social Media turned into a major provider of data, a voluminous new source to take into account. In a world where more and more content becomes available, it has become virtually impossible for humans to process this vast stream of information. For the journalists to do their work, solutions like PHEME are needed to support the newsrooms in navigating the content and quickly identifying what matters. They become virtual assistants replacing tedious parts of journalistic work needed to assess the importance or the credibility of content. While robots may never be able to identify that an emitter has been hacked, or to call the adequate office for confirmation, they can definitely help in making sense of large amounts of data.

In the Digital Journalism Use Case, PHEME is aiming specifically at helping the newsroom identify rumours spreading on Social Media, using the Ushahidi platform to visualize the relevant information. The tool will be able to track a large number of twitter sources linked by geographical position and/or common topics and give insights into the provenance of crowd sourced information. This will provide the journalist with a better view of what may or may not constitute a legitimate piece of information, and to understand how rumours develop.

Asserting credibility and curating relevant information is at the core of a journalist's work. PHEME may increase substantially his or her productivity in everyday work.

2 Intended audience

This document details internal procedures to ensure the project is being governed according to plan and meeting its own targets. With respect to self-assessment, internal procedures are being put in place, to ensure that deliverables are peer reviewed and verified.

As such, this document is intended for all actors involved in the PHEMEproject.

3 Work Package 8 Overview

Task 8.1 Requirements gathering, use case design and interface mock-ups

Duration: January 2014 – June 2014 (1st report)

Task 8.2 Journalism Corpus Collection and Annotation

D8.1 Requirements gathering, use case design and interface mock-ups

Duration: January 2014 – June 2015

Task 8.3 Open-source digital journalism showcase

Duration January 2014 – December 2016

Task 8.4 Iterative Evaluation

Duration August 2014 – December 2016

3.1 8.1 Deliverables: Requirements and Use Case Design document

3.1.1 Use case scenarios

The content of this deliverable is to develop 2-4 possible **use case scenarios** for PHEME rumor detection and veracity assessment tool (N=2-3).

The principle approach that will be adopted to investigate how journalists go about doing their daily work will be ethnographic observations. This will involve initial short preliminary interviews with key individuals to identify appropriate targets for specific observations, observation of the news room, followed by more structured interviews will be conducted in order to drill down into the exact characteristics of significant and relevant features of the work that the observations have identified.

Data gathered during the ethnographic work will be written up and anonymised prior to careful explication of the work's salient organisational characteristics. These organisational characteristics will be translated into use case scenarios that reflect the real world accomplishment of the work.

Data gathered will take the form of field notes and, where appropriate and permitted, audio-video recordings and/or photographs. All of the ethnographic data gathered will be stored in accordance with the UK Data Protection Act 1998, on a password protected drive in a secure facility and only for the duration for which it is required. It will only be directly accessible by the ethnographer who has collected the data and anonymisation will take place prior to any of the data being shared with other members of the project.

3.1.2 Requirements gathering / interface mock-up

Establishing the requirements of newsroom journalists consists mainly of looking into what (and how) journalists are doing in their daily work (i.e. discovering potential stories (focus on rumours), checking up on facts etc.). In order to do this, work contents have to be analysed in as detailed a way as possible. For this reason, the approach followed will be based on ethnographic observation.

1. Preliminary literature analysis with focus on task centered social work analysis in order to benefit from existing work.

2. Observations will be conducted by watching and recording in detail the ordinary everyday work practices of journalists in order to uncover the exact methods required to accomplish their work and the kinds of reasoning underlying that accomplishment. To diversify the insights into journalistic workpractices, target subjects will include journalists with different core tasks from different work settings (swissinfo.ch, BBC, Guardian)

3. Subsequent to these observations, more structured interviews will be conducted in order to drill down into the exact characteristics of significant and relevant features of the work that the observations have identified. The structure of these interviews would include how rumours are detected and judged in respect to their credibility, how they are the incorporated into articles and where tools or software functions might support these processes.

   Issues of specific interest will include: duties and associated work tasks (complete list) of journalistic work with rumours. Within these, try to get individual work processes ("typical cases") and heuristics in face of problems (e.g. if a doesn't work, do b), information and collaboration needs as well as (directly) requirements and scope for a hypothetical journalistic newsroom tool (in accordance with identified use case scenarios). Discuss / validate structured Interview and apply internally (N=2-3) and externally (i.e. SWI partners, BBC, Guardian, N=1-2 each).

4. Analysis of data from literature, observations and interviews to get lists of different tasks, processes, information & collaboration needs, as well as their relative importance as a set of **use case scenarios.**

5. Draft **mock-ups of interfaces** from requirements to capture salient features of the work.

3.2 8.2 Deliverables: Journalism Corpus Collection and Annotation

3.2.1 Journalism Corpus Collection

The rumour corpus consists of a database of content (one for each rumour) corresponding to a specific rumour (including mostly twitter feeds (https://dev.twitter.com/docs/api/1/get/search) but also corresponding media publications, blog articles or other social media content).

This corpus will mainly have to be collected in real time, since the Twitter search API goes back only 5-6 days (https://dev.twitter.com/discussions/20513).

The corpus of rumours will be collected by SWI in collaboration with internal journalists. The target is to come up with 8 – 10 rumours by month 18 (June 2015). For the collection, SWI will inform its journalists of the task and have them report candidate rumours to the PHEME work package leader. Upon identifying a rumour as fitting our specific need, it will then be tracked and stored accordingly.

Since a number of rumours are to be analyzed already in a very early stage by UWAR in Work Package 2.1, in the best case there will be already for June 2014 (month 6) one rumour

database for each of the four types of rumours: misinformation, disinformation, unverified information and disputed information[1].

On the basis of these four early rumours, UWAR will develop an "actor code frame" and annotation schema for the annotation of future rumours and provide these to swissinfo at month 6 (June / July 2014).

3.2.2 Annotation

For analysis of the corpus, every single corpus data point will have to compared and classified according to corrections, clarifications and retractions in mainstream media, to made claims and their truthfulness, contained hashtags and to the presence of specific words or word patterns (e.g., conspiracy, hoax, a claims b etc.).

Given the annotation framework from UWAR, the annotations task in phase 2 (after month 6, i.e. for the remaining 4 – 6 rumours, see timeline) will partly be executed by SWI, partly by crowdsourcing (organized by UWAR) and partly by volunteer annotators (organized by USFD).

4 Methods and Results

4.1 8.1: Requirements gathering, use case design and interface mock-ups

An important part of the work in Pheme is to establish and develop the relevance of the research as a whole (and associated aspects of design) to particular application domains where the passage of rumours and associated concerns with veracity assessment are especially important. The two identified application domains within the project are healthcare and journalism. Work Package 8 is focused upon the journalistic domain.

Initial work in Work Package 8 is formulated around a loose pipeline that is open to iterative development over the course of the research being undertaken. Central to the Work Package is a foundational concern with providing tools that can simultaneously deliver the insights growing out of the research being undertaken in other Work Package whilst ensuring that these tools are scoped as well as possible for their use within the proposed application domain. The loose pipeline is therefore organised around: detailed engagement with the real world practices of journalism in order to identify how those practices are organized, the tools and reasoning they depend upon, and the concomitant requirements arising from those practices that can be specified; extraction from this overall body of requirements of the requirements most particularly aligned with the objectives of Pheme and its design objectives

[1] **uncertain information** or **speculation** (e.g. an analyst claiming the Bank of England will raise interest rates at their next meeting)
disputed information or **controversy** (e.g. aluminium may or may not cause Alzheimer's)
misinformation (e.g. misrepresentation and quoting out of context)
disinformation (e.g. Obama is a Muslim)

and further refinement of these requirements through deeper investigation of journalistic practice in dedicated interviews; articulation of the identified requirements through a series of use case scenarios in order to aid their transmission to design; the iterative development of prototypes that integrate both these requirements and the other elements being developed within Pheme; and the evaluation of those prototypes.

The pipeline is described as being loose because it is anticipated that the requirements gathering and testing processes will go through a number of iterations before suitable levels of adequacy are accomplished. The means to achieve this is through proven and established instruments of social and work analysis. Most importantly and at the centre of these instruments, ethnographic observations are being conducted by an experienced ethnographer. This is accompanied by a thorough preliminary literature review and analysis in order to profit from already existing work and, where additional information is necessary, in depth interviews with selected journalist.

This section of the report provides the first body of input into what is expected to be a three-phase exercise in requirements gathering. Here we provide a set of requirements relevant to journalistic work that have been gleaned from the existing corpus of studies of such work and assessed for their relevance to Pheme. A second report will assemble new findings and developed requirements that are based upon studies specifically undertaken for the purposes of Pheme, largely in one setting. A third report will bring together a tightly specified set of requirements informed by further studies of journalistic work across other settings so as to ensure their portability and relevance across the domain.

In this report we make use of several preceding studies of journalistic work we have at our disposal where the reported data is available in sufficient depth for us to be able to assess its relevance for the formulation of requirements within Pheme. The primary resources being used are:

- An as yet unpublished study of day-to-day journalistic work for the Guardian covering two Euro 2012 football championship matches, that was undertaken to provide requirements for the design of a workbench for the Guardian's Olympics sports desk. The study particularly focused on the use of social media in the context of both through the day Live Blogging and Minute-By-Minute coverage of the matches.
- A series of interviews with members of the Guardian Sports Desk that examine their working practices across a range of themes.
- A recording of a prototype assessment exercise that took place at the Guardian after the workbench had been designed.

In these cases access to original data has enabled us to explore its relevance in detail. We have, additionally, drawn upon a couple of resources that offered particularly rich specificity regarding journalistic work such that requirements could be quite clearly identified:

- Various articles discussing the use of User-Generated Content published by the BBC College of Journalism.
- A blog-based article on Digital First journalism that explores in detail the workflows associated with focusing on the digital production of journalistic work.

Beyond this we have drawn upon a much larger body of previously published studies of journalism and guidelines where there are elements that are also indicative in various ways of requirements relevant to Pheme.

As will be seen in the subsequent sections, we have sought to organize the uncovered requirements around a series of primary considerations within the Pheme project, namely:

- The work of uncovering stories
- The work of establishing the usability of stories
- The work of verifying facts and associated suppositions, speculations and rumours
- The work of actually writing stories
- The work of revision and re-verification
- The tools required to get the job done
- The practices central to accomplishing journalistic work
- The aspects of how journalistic work is organized in situ that are impactful for how the work unfolds as a whole

4.1.1 Use case scenarios

The use of Social Media monitoring as a provider of unverified content occurs in very different contexts. An early analysis from swissinfo.ch proposes to look into 3 types of journalistic scenarios which lend themselves to intensive Social Media fact-checking:

- Geographic analysis of messages around an localized event. This covers the full spectrum of unexpected events (natural disaster) to expected events (vote), including local events that are newsworthy or become newsworthy because of their scope (protests). The common element is the location of the event, not limited to a topic in the hope of finding new bits of relevant information. The more unexpected an event, the more disorganised the discourse (no hashtag in place, no common vocabulary, etc.), which explains the importance to look at everything that is being said in a defined location. The most important source of information in this case is a map of a defined region, in order to help the journalist identifying where the action is taking place and where potentially interesting reports are coming from and an assessment of their veracity to quickly sort out relevant information.
- Analysis of a specific rumour. This covers the analysis of a specific bit of information and the way it was shaped in Social Media. Starting with a specific bit of Social Media content, this allows the journalists to analyse each and every iteration of the message, identify the influencers that helped convey or modify the message, get back to all non-Social Media sources that are quoted in the context of the message, the crowd reaction to that message and evolution of topics when new information becomes available.
- Topic-based analysis of a massive amount of live messages around a non-localized news event. The geographic zone might also be too big, or the news not big enough to represent most of that region's Social Media chatter. This covers the live comments that stack up during news coverage and may or may not be true statements. The goal is to allow the journalists to quickly sift through the incoming messages with the possibility to faster assess the truthfulness of a statement ("The party had the exact same score in 1967, drop it!") or correct it if it gains popularity and needs to be invalidated.

The analysis suggests that all other cases discussed so far can be dealt with within these scenarios. The specificity of the information to obtain and the type of result to display

may end up with several Ushahidi solutions or a single one. The ideal solution has all three scenarios in a single package, allowing the display, classification and analysisof information in the same way independent of a specific type of event. This will have to be challenged once all final requirements are defined.

Cases	Scenario
Haïti Earthquake, Ukraine crisis, Syria's war, Arab Spring uprisings in specific cities, London Riots, Boston Marathon, Afghan election, etc.	Geographic analysis of messages around an localized event.
ET702 hijacked, Ben Laden raid, London Eye on fire, Marathon bombers identified, etc.	Analysis of a specific rumour.
Olympic medals, death of Bin Laden, US election, geopolitical issues, etc.	Topic-based analysis of a massive amount of live messages around a non-localized news event.

4.1.2 Established Requirements

As indicated above the following materials will indicate a series of requirements, visible in prior studies that are available to us and the extant literature, that are likely to be relevant to the development work within Pheme. In each case we provide at the end of the section a table that summarises the various requirements, the kinds of design considerations they implicate, and the likely significance they will have moving forward.

4.1.2.1 Uncovering stories

A primary consideration for the project is the way in which journalistic work is organised such that stories may be uncovered in the first place. Existing work indicates that this can happen in a number of ways, including: the use of news wires and news feeds which assemble breaking stories in a number of ways and can be set with filters by journalists to ensure they receive the most relevant notifications; RSS feeds and email notifications from a variety of pre-selected sources and known individuals; other newspapers and journals; social media, especially Twitter; the radio and television media; phone calls and tips from other journalists; and so on. Twitter and news wires tend to be the most regularly consulted of these. With regard to this note the following:

- An increasingly important resource used by journalists for uncovering potential stories is Twitter. Observations of journalists working on the sports desk at the Guardian revealed a strong preference for using the web-based Twitter service rather than Tweetdeck or the mobile application. The rationale provided for this was that the web-based service requires manual updating rather than giving a live feed. In this way journalists can hold potentially interesting material in place at the top of the list without it vanishing off the screen. Other observations of the routine use of news wire services, where live feed is automatic, revealed journalists either putting a marker against stories of interest, printing them out, or copying them and pasting them into a word document. The common concern here is with keeping possible leads on stories in view and to hand rather than having to hunt them down later. This indicates a need for future tools to provide effective mechanisms for holding salient information in view.
- Other observations of similar kinds of work suggest that there is quite a large variety of user input that can be drawn upon, extending beyond specific tweets or other direct UGC. This can include things like aggregated trends, the most re-tweeted tweets, lists of

prolific twitter users and so on. Additionally, where journalists are making regular use of their personal Twitter accounts it is evident that they are quite assiduously crafting the lists of people they follow in order to maximize the chances of them encountering relevant tweets in the future. Similar cultivation of contacts occurs on Facebook and LinkedIn, as well as through more traditional lists such as email and mobile phone contacts. All of this suggests two distinct core requirements. One of these relates to providing support for the identification of what kinds of things are capturing people's interest at the moment. The other relates to supporting activities of identifying and retaining prospectively useful future sources which, in the case of social media in particular, is not a static enterprise but rather one that requires ongoing management to ensure ongoing relevance. Note that these kinds of sources may currently be spread across a number of different applications and stores.

- Assessment of previous prototypes of dashboard-type systems suggest that mechanisms for assembling flexible user-relevant feeds such as Twitter monitoring lists will need to incorporate ways of updating the lists on a frequent basis, at least daily, and possibly more often than that.

Requirement	**Design Implication**	**Importance**
Keeping potential leads on stories to hand for rapid recall when required.	Information sources that are turned to for uncovering stories should be as easy to preserve and bring back into view as possible.	High
Provide support for uncovering current trends	Integrate into the system mechanisms for uncovering and/or collating trending information across user-specified resources. Incorporate flexible ways of displaying this information so that it can be viewed when required.	High
Provide ways of keeping feeds relevant and fresh	Build into the system the scope for regular updating of targeted sources and feeds to reflect current trends and interests	High
Provide support for identifying, retaining, updating and managing prospectively useful future sources.	Build into the system ways of making current sources and source feeds visible without drilling down into multiple applications or changing browsers.	Moderate

4.1.2.2 Selecting stories

The selection of stories is tightly related to the preceding consideration of how stories are uncovered. However, existing observations indicate that journalists often have a number of potential stories under consideration at any one time, but not all of them get worked up. There is therefore a further element within the work where actual choices get made. As mentioned above, one resource that is increasingly being turned to for content is UGC (User-Generated

Content). UGC can be collected from a number of sources but the most commonly witnessed ones are direct comments to news sites and blogs, email, and Twitter:

- Materials gathered from journalists working on the Guardian sports desk confirmed a regular use of UGC. A particular challenge uncovered here is the varied circumstances in which UGC may need to be engaged with by journalists when assessing whether to make use of it. When covering live events selection has to happen very rapidly because of the difficulties of staying on top of providing coverage of the event itself. Under these circumstances UGC has to be monitored as quickly as possible. In this respect Twitter was found to be more amenable to review than emails because of the 140-character restriction. Requirements here relate to a need to be able to support this varied kind of engagement with UGC. At some times of the day selection may be open to happening quite leisurely with lots of time for judgments of interest, relevance and veracity to be made. At other times these judgments may have to happen in minutes or seconds.
- Actual decisions regarding which stories to run with pass through a number of phases: the initial application of crude criteria regarding what might count as a suitable story in the first place, such as relevance of content to the interests of the specific journal, potential interest of the story itself, the presence of verifiable information, what else is already on the brew, and so on; the exercise of journalistic 'gut-feeling' as to what might be best to run with; discussion of the story with one's peers; discussion of the story with editors and at editorial and group meetings; the way the story shapes up as it is written; and how the story is or isn't overtaken by turns of events such that it gains or loses relevance. This work is embedded within the use of a range of different resources, mostly face-to-face interaction, but also phone calls, emails, text messages and instant messaging, and so on. Particularly key is the work of negotiation with editors through to the point where the pursuit of the story is ratified and part of the work is also preparing oneself for the work of explaining a story to the editor, often based upon understandings of what kinds of concerns editors may exhibit such as the importance of figures to back up claims and what kinds of stories they shot down just a week ago. Requirements here are diffuse and chiefly relate to providing surrounding context and the scope for verification because most stories that are not relevant or not interesting are dismissed at first encounter. This points to a complex calculus which involves the weighing up of things like how much work it will take to make a story viable against its relative pertinence or merit in view of what else is also to be considered for pursuit at the same time. Particularly relevant support of value here would be an indication of what the work of verification might look like before that work actually takes place.

Requirement	Design Implication	Importance
Support rapid assessment of UGC and its usability in terms of interest, relevance and veracity.	Provide (as is planned in Pheme) mechanisms for assessing, highlighting and making visible at a glance aspects such as veracity and relevance. Conjointly, being able to rapidly integrate usable UGC into the ongoing stream.	High
Provide ways of rapidly seeing what elements of a story might require verification and indicate what resources are available for the verification to take place	Build in mechanisms for recognizing and highlighting elements in initial sources that will require some form of verification. Provide further indicators of the available resources for doing that work of verification or, where available, indicate where veracity and required detail has already been established.	High / Moderate
Supporting tracking of several stories at the same time	Build in capacity to context switch between different dashboard instances	High / Moderate

4.1.2.3 Verification

The work of verification itself is often embedded within the preceding activities associated with first uncovering possible stories and then choosing between them, but it is also a feature of the ongoing composition of journalistic text (and, indeed, other media as well):

- A fundamental part of journalistic practice, visible in observations and emphasized across numerous sets of guidelines, is the checking of facts and sources. The actual work of checking to happen in a variety of different ways but always turns upon a) the recognition of features that require verification, and b) the location of means whereby verification can take place. A core requirement is therefore the support of these two features, regardless of the specific ways in which this is actualized. It should be noted that this is something that can happen when a possible story is first uncovered, when candidates for selection are being considered, once the story has been decided upon, and even afterwards as the story is revised, subbed, translated, submitted for editorial approval, or even once published.
- Discussions of the use of UGC (User-Generated Content) at the BBC indicate that an increasingly important feature of how journalists verify the trustworthiness of UGC is by checking locations against maps and geo-located images, This suggests that verification support should incorporate easy ways of moving between UGC and these kinds of resources, ideally through these kinds of associations being already ready-to-hand should they be required.
- Another resource that has been pointed to is the use of the expertise of others within the same organisation to assess certain aspects of the content such as language use. This

points to a classic topic in knowledge management where the requirement is to somehow make visible the range of expertise in one's organisation such that it can be drawn upon as and when it proves relevant.

- More technical treatments of UGC can include things like verification of date by tracking down the upload provenance of content or comparing imagery to discoverable weather and light conditions in the claimed location. Here the requirement is to be able to facilitate smooth and timely interaction with the technical resources that can expose this kind of information.
- BBC materials relating to the use of UGC also indicate the importance of preserving and keeping up to date lists of verified materials and providing mechanisms for ensuring that relevant materials of this order are made visible to people as they work on particular stories.
- Other resources also stress that, even in the realm of digital journalism based upon UGC, more basic techniques for verifying content continue to be relevant, such as personal inspection of user profiles to see whether or not they promote trust, making phone calls, sending emails, and even visiting people in person where this is feasible. This underscores an ongoing requirement to continue to make available basic information such as a source's profile, affiliations, and contact details.

Requirement	**Design Implication**	**Importance**
Provide support for recognition of features that require verification	The system should be able to recognise features that might require verification and should be able to highlight these to users	High
Provide support for locating ways in which particular features might be verified	The system should have components that are able to indicate potential verification strategies according to features identified and should be able to point users towards suitable resources for following those strategies	High

Provide to-hand map and image based verification resources for UGC	The system should be able to identify location relevant elements within UGC and be able to associate these with relevant maps and images. These should then be made available upon request (e.g. by clicking on a simple link or button)	Moderate
Provide up to date information about available expertise for the purposes of verification within the same organisation or one's personal network.	Build into the system ways of collating and making visible people's known expertise in various domains such that they can be proposed as potential experts in relation to the story in hand.	High / Moderate
Provide ways of bringing together quickly all available information regarding a source's profile, affiliations and contact details.	Build into the system a mechanism that identifies sources and assembles from all available online repositories profile and contact information for the sources so that these can be inspected and drilled into according to need.	High / Moderate
Provide ways of handing over materials requiring technical verification to relevant and available verification providers.	Provide a conduit that is able to: recognise and extract elements requiring technical validation; to then assess availability of appropriate sources of validation; to forward those elements to the relevant source; to capture the source's findings; and to then ensure return of the findings to the originator.	Moderate
Provide ways of keeping and updating a store of verified materials and making visible those materials as and when they are required.	Creating and maintaining a database of verified materials and ensuring that it is indexed in a way in which links to verified content relevant to specific stories can be made visible.	Moderate

4.1.2.4 Handling supposition, speculation and rumour

Journalistic work is often about more than just the simple verification of facts. A central feature of verification work, that is relevant to Pheme, is the way in which journalists handle materials they see as being currently grounded in supposition, speculation and rumour. This is central to the Pheme project and is an area that is currently under-researched. Some points of relevance are:

- Many of the existing ways in which supposition, speculation and rumour is handled are a part of the verification activities already discussed above. The specific requirement here is to a) be able to directly recognise content as amounting to one of these things, to b) be able to track the provenance of such content, and c) to be able to provide an assessment of the validity or otherwise of such claims. These are central components of the proposed Pheme platform. Additionally, note:
- Reflections upon the use of UGC at the BBC indicate ways in which it is used in tandem with other resources when potentially dubious claims are encountered, e.g. by using Twitter to broadcast requests for verification, use by other parties of other resources such as those already mentioned above to provide assessment, using Twitter to circulate revisions and correctives. This suggests a need to support fairly rich and diverse workflows for some kinds of verification, especially in the context of news journalism where speed is often of the essence.

Requirement	**Design Implication**	**Importance**
Provide a way of assessing the validity of suppositions, speculative claims and rumours as and when they are encountered.	Build a system that can recognise supposition, speculation and rumour. Provide mechanisms for uncovering the provenance of such content. Provide a mechanism for assessing the veracity of such content. Provide a mechanism for suitably feeding back the outcomes of such an assessment to users.	High
Provide support for workflows that can: exploit feeding social media requests for verification into external checking mechanisms, provide rapid feedback on the output of those mechanisms, exploit social media to provide confirmation, revision or correction of previously unverified social media content.	The system should be able to support an end to end process that covers the issuing of verification requests, the identification of appropriate verification protocols, the exercise of those protocols, the communication of the outcomes to the originator of a verification request, and the shaping of responses to the original content.	High / Moderate

4.1.2.5 Writing up stories

Of course, a central part of journalistic work is the actual writing up of stories and the actual work of writing in this context is under-explored in the literature. Examination of this will feature in the second requirements document but some initial observations here would include:

- Something noted in several sources is the way in which journalists will often seek to assemble their materials into a single text document. Relevant content from other sources is often copied directly into this document for ease of reference without having to switch between numerous different windows, and text is often composed around these materials, typically with a story lead that summarizes the following content at the top of the page and with the title being composed last.
- Around this key documents are often downloaded or even printed and annotated as a further source of compositional reference.

- Additional, enriching elements can include the conduct of interviews, the recording of those interviews, the editing of those interviews and the subsequent incorporation of either transcripts or audio / video clips into the journalistic product.
- Other materials such as scans of documents or UGC may also need to be embedded in the product.
- Finalisation of the story often also includes the location of suitable images, which may involve negotiation with in-house providers of such images or may involve the use of various online image libraries.

Requirement	**Design Implication**	**Importance**
Provide tools for easy compilation of materials to a single file from across multiple resources in a consistent format open to specification by the user.	Provide as a feature of the system a means of checking relevant sources and having them already collated to a single place in a pre-specified format.	Moderate
Provide automatic association of resources to a specific story as soon as compilation is undertaken.	Incorporate into the system a means of tracking all resources that are copied from within the construction of a story file such that the association is preserved.	Moderate
Provide for easy and flexible assembly of multimedia content and UGC as a story is being crafted.	Provide authoring tools as part of the system.	Low (news organisations have already invested heavily in this and a range of applications are already in existence)

4.1.2.6 Revision and re-verification

Another notable feature of journalistic work is the way in which text is ongoingly and iteratively revised and reformulated, both by its original author, and by other parties engaged in 'subbing', 'editing' or 'translation' activities. This frequently leads to a further round of verification work. A further element of this that is increasingly visible is the use and curation of UGC, especially in the context of things like liveblogs:

- Observations of use of UGC by the BBC indicate a growing role for journalists in the curation of UGC generated through social media such as Twitter into ongoing 'conversations' through artefacts such as liveblogs. Others also discuss the use of Storify to accomplish this. Curation involves the selection of UGC to publish to the blog but also involves commentary and elaboration around this content. As such the 'story' is continually being developed and revised and demands ongoing verification of relevant features of the content. There is therefore an evolving set of requirements here regarding the support of these kinds of activities such that: usable UGC is uncovered and made available, potentially from a variety of sources (e.g. direct comments to news sites, emails, tweets, facebook comments, and so on); verification support is enabled where necessary; selected content is easily copied and crafted into place; and where appropriate re-crafting, re-verification, and updating is enabled. This process currently involves the interleaving of a number of different resources. There is clearly space to simplify the workflow but, at the very least, tools designed by Pheme will need to integrate into this process in as seamless a fashion as possible.
- Various materials relating to the use of UGC in creating assemblies such as liveblogs also indicate that there are also points within the generation of these where summaries have to be provided. Resources that facilitate the summarization of selected UGC would therefore also have some efficacy here. The principal challenge here is quantity as major events can generate large amounts of UGC. Some sources also indicate that where coverage relates to live events filtering of these materials can cause significant trouble. Summarisation usually takes place at the end of events so time here is less critical.
- It is also worth noting that investigations of journalistic work suggest that, even with relatively traditional stories, there is also a sense of there being a developing story and journalists may continue to use a range of resources including Twitter feeds and searches, Facebook groups, wires, and government press releases to monitor a story's progress and incorporate changes along the way. In this way stories can sometimes remain open to revision for some time after they are first notionally 'completed'.

Requirement	Design Implication	Importance
Provide effective ways of verifying and incorporating UGC in the development of ongoing stories as an ongoing process with suitably straightforward mechanisms for update and revision of content	Provide a liveblog curation tool that incorporates the verification mechanisms developed above / Provide suitable APIs for integrating the verification mechanisms into the processes provided by other curation tools such as Storify	High

Provide resources for being able to summarise UGC over a selected period.	Building into the system mechanisms that can do text highlighting within UGC according to chosen features and, potentially, certain kinds of text summarization.	Moderate
Provide for the ongoing tracking of story-relevant content after it has gone through the initial publication cycle such that important new content can be identified.	Build into the system mechanisms for automatically associating relevant content to recently released materials with a concomitant mechanism for making this visible and, potentially, providing alerts.	Moderate

4.1.2.7 Tools

Underpinning all of the work journalists do are the tools they make use of in order to go about their daily business. Whilst there are a multiplicity of tools available for this kind of work some primary organizational features need to be considered:

- The study at the Guardian of work on football match coverage revealed that journalists have highly individual ways of organising their screen displays. This suggests that it is important that the display features of tools designed to support journalists are customizable and open to being organized by the journalists themselves in a variety of ways.
- Journalists frequently work in an interleaved fashion between a large number of different resources, including different kinds of media. This can lead to heavily over-populated screens, with journalists working between a wide range of different tabs and windows. Where video is a primary focus journalists often seek to work across multiple displays to ease the congestion and make it more straightforward to work seamlessly between video and other media. Tools supporting journalists therefore need to be careful to provide for something other than all-in-one displays. Instead it is important that features of most relevance to the task can be foregrounded whilst others can be easily set aside but recalled and switched to without difficulty.

Requirement	Design Implication	Importance
Customizability of display features	Avoiding a single block display and providing instead a collection of logical components that can be opened and closed and moved around the screen according to preference.	High
Display simplification so that: only features of relevance stay in the foreground; it is easy to bring other elements to the fore instead; cross-referral between different elements can be accomplished without continual switching	Providing modular displays such that different elements can be brought to the fore according to need, switched between, and organised for simultaneous access	High / Moderate

4.1.2.8 Practices

Ultimately all of the other concerns we have been addressing are embedded within an assembly of journalistic practices and bodies of journalistic reasoning that provide for the realization of any specific instance of journalistic work. Additionally, existing studies also make visible the following kinds of relevant concerns:

- Journalists working on the Guardian sports desk were noted to use a mixture of their own Twitter accounts and the organization's account. This mix has been seen amongst other journalists as well. In some cases it is a case of different journalists preferring one or the other. In some cases it is a case of journalists using both. Various mechanisms were seen to facilitate the use of both, such as having different accounts open in different browsers. Consoles are also used, such as Hootsuite. The background to this mix is that different people are followed and tweeted to according to which account is chosen and there are different reasons for following and tweeting, according to circumstance. A number of journalists have personally crafted a wide selection of trusted resources whose tweets may provide them with potential leads. At the same time, tweeting news was seen as needing to come from the officially sanctioned account. These are fundamentally distinct orientations and both require support.

Requirement	Design Implication	Importance
Equally good access needs to be provided to both personal and official resources for getting leads and communicating stories.	Build a system that is able to support selecting, collating, switching between and keeping in view relevant resources attached to both personal and organisational identities.	High / Moderate

4.1.2.9 The organisation of the work

Journalistic work is further organized around its own particular routines and rhythms and foundational organisational constraints. These, too, give rise to certain kinds of requirements:

- Observations of sports desk work, and also of broader forms of news coverage, have revealed an oscillation between periods of high pressure where it is hard to keep on top and deliver in a timely fashion and other periods where the pace is less intense. This results in journalists making use of quite distinct kinds of resources at different times of day. Short-form materials such as Twitter feeds, for instance, that are easily understood at a glance and that can be easily copied and pasted from one place to another are concentrated upon when pressures are high. By the same token, broader information resources that require more focused engagement, such as aggregated statistics and trends and web-links, are more commonly turned to when things are slower. In some cases these busy and quiet periods are predictable. For instance first thing in the morning when a lot has to be looked at and got underway can be a period of higher pressure. Providing cover of live events can be equally demanding. Afternoons and late at night may be quieter. At the same time, the breaking of stories for news is inherently unpredictable so pressure can arise at any point. What is clear is that resources that require equal levels of engagement at all times, especially where that engagement is more than momentary, would sometimes hinder effective journalistic practice. This resources need to be open to being organised so that they can be 'light' and rapidly intelligible in one mode, but open to being shifted to other modes of presentation where content be exposed in greater depth, with the shift in modality being itself quick and easy to accomplish.
- News is a 24-hour a day business and events do not go on hold when a journalist leaves their desk. Many news organisations therefore operate in shifts. Observations at the Guardian and elsewhere show that communication between members of different shifts can be limited or piecemeal. This can be problematic for a number of reasons. Processes may have been begun that need following through because to leave it until the next day may be too late. People on a new shift may unwittingly replicate work that has already been undertaken by people on the previous shift. Held off leads may get overlooked. Items that might have been pursued at other times may not get pursued close to the end of a shift and this needs communicating. All of this points to a need for mechanism that provides for effective cross shift communication and an easy way of inspecting potentially salient previous activity by colleagues.

Requirement	**Design Implication**	**Importance**
Alternative modes of presentation ranging from at-a-glance understandable and usable to comprehensive and expansive.	Provide a tool that has different collapsed and expanded views where collapsed views still provide coherent and usable summaries, but where movement to expanded forms or other kinds of content can be done with minimal effort	High
Cross-shift, cross personnel communication of incomplete, postponed, and salient prior action.	Effective tools need to support cross-personnel working that is not just organised around a division of labour but also around temporally encompassing more than one individual working on the same task	High

4.1.2.10 Requirements 1st round: overview

	Requirement	Design Implication	Importance
1	**Uncovering stories**		
1.1	Keeping potential leads on stories to hand for rapid recall when required.	Information sources that are turned to for uncovering stories should be as easy to preserve and bring back into view as possible.	High
1.2	Provide support for uncovering current trends	Integrate into the system mechanisms for uncovering and/or collating trending information across user-specified resources. Incorporate flexible ways of displaying this information so that it can be viewed when required.	High
1.3	Provide ways of keeping feeds relevant and fresh	Build into the system the scope for regular updating of targeted sources and feeds to reflect current trends and interests.	High
1.4	Provide support for identifying, retaining, updating and managing prospectively useful future sources.	Build into the system ways of making current sources and source feeds visible without drilling down into multiple applications or changing browsers.	Moderate
2	**Selecting stories**		
2.1	Support rapid assessment of UGC and its usability in terms of interest, relevance and veracity.	Provide (as is planned in Pheme) mechanisms for assessing, highlighting and making visible at a glance aspects such as veracity and relevance. Conjointly, being able to rapidly integrate usable UGC into the ongoing stream.	High
2.2	Provide ways of rapidly seeing what elements of a story might require verification and indicate what resources are available for the verification to take place	Build in mechanisms for recognizing and highlighting elements in initial sources that will require some form of verification. Provide further indicators of the available resources for doing that work of verification or, where available, indicate where veracity and required detail has already been established.	High / Moderate
2.3	Supporting tracking of several stories at the same time.	Build in capacity to context switch between different dashboard instances.	High

3	**Verification**		
3.1	Provide support for recognition of features that require verification	The system should be able to recognise features that might require verification and should be able to highlight these to users.	High
3.2	Provide support for locating ways in which particular features might be verified	The system should have components that are able to indicate potential verification strategies according to features identified and should be able to point users towards suitable resources for following those strategies.	High
3.3	Provide to-hand map and image based verification resources for UGC	The system should be able to identify location relevant elements within UGC and be able to associate these with relevant maps and images. These should then be made available upon request (e.g. by clicking on a simple link or button).	Moderate
3.4	Provide up to date information about available expertise for the purposes of verification within the same organisation or one's personal network.	Build into the system ways of collating and making visible people's known expertise in various domains such that they can be proposed as potential experts in relation to the story in hand.	High / Moderate
3.5	Provide ways of bringing together quickly all available information regarding a source's profile, affiliations and contact details.	Build into the system a mechanism that identifies sources and assembles from all available online repositories profile and contact information for the sources so that these can be inspected and drilled into according to need.	High / Moderate
3.6	Provide ways of handing over materials requiring technical verification to relevant and available verification providers.	Provide a conduit that is able to: recognise and extract elements requiring technical validation; to then assess availability of appropriate sources of validation; to forward those elements to the relevant source; to capture the source's findings; and to then ensure return of the findings to the originator.	Moderate
3.7	Provide ways of keeping and updating a store of verified materials and making visible those materials as and when they are required.	Creating and maintaining a database of verified materials and ensuring that it is indexed in a way in which links to verified content relevant to specific stories can be made visible.	Moderate

4	**Handling supposition, speculation and rumour**		
4.1	Provide a way of assessing the validity of suppositions, speculative claims and rumours as and when they are encountered.	Build a system that can recognise supposition, speculation and rumour. Provide mechanisms for uncovering the provenance of such content. Provide a mechanisms for assessing the veracity of such content. Provide a mechanism for suitably feeding back the outcomes of such an assessment to users.	High
4.2	Provide support for workflows that can: exploit feeding social media requests for verification into external checking mechanisms, provide rapid feedback on the output of those mechanisms, exploit social media to provide confirmation, revision or correction of previously unverified social media content.	The system should be able to support an end to end process that covers the issuing of verification requests, the identification of appropriate verification protocols, the exercise of those protocols, the communication of the outcomes to the originator of a verification request, and the shaping of responses to the original content.	High / Moderate
5	**Writing up stories**		
5.1	Provide tools for easy compilation of materials to a single file from across multiple resources in a consistent format open to specification by the user.	Provide as a feature of the system a means of checking relevant sources and having them already collated to a single place in a pre-specified format.	Moderate
5.2	Provide automatic association of resources to a specific story as soon as compilation is undertaken.	Incorporate into the system a means of tracking all resources that are copied from within the construction of a story file such that the association is preserved.	Moderate

5.3	Provide for easy and flexible assembly of multimedia content and UGC as a story is being crafted.	Provide authoring tools as part of the system.	Low (news organisations have already invested heavily in this and a range of applications are already in existence)
6	**Revision and re-verification**		
6.1	Provide effective ways of verifying and incorporating UGC in the development of ongoing stories as an ongoing process with suitably straightforward mechanisms for update and revision of content	Provide a liveblog curation tool that incorporates the verification mechanisms developed above / Provide suitable APIs for integrating the verification mechanisms into the processes provided by other curation tools such as Storify.	High
6.2	Provide resources for being able to summarise UGC over a selected period.	Building into the system mechanisms that can do text highlighting within UGC according to chosen features and, potentially, certain kinds of text summarization.	Moderate
6.3	Provide for the ongoing tracking of story-relevant content after it has gone through the initial publication cycle such that important new content can be identified.	Build into the system mechanisms for automatically associating relevant content to recently released materials with a concomitant mechanism for making this visible and, potentially, providing alerts.	Moderate
7	**Practices**		
7.1	Equally good access needs to be provided to both personal and official resources for getting leads and communicating stories.	Build a system that is able to support selecting, collating, switching between and keeping in view relevant resources attached to both personal and organisational identities.	High / Moderate

8	**The organization of the work**		
8.1	Alternative modes of presentation ranging from at-a-glance understandable and usable to comprehensive and expansive.	Provide a tool that has different collapsed and expanded views where collapsed views still provide coherent and usable summaries, but where movement to expanded forms or other kinds of content can be done with minimal effort.	High
8.2	Cross-shift, cross personnel communication of incomplete, postponed, and salient prior action.	Effective tools need to support cross-personnel working that is not just organised around a division of labour but also around temporally encompassing more than one individual working on the same task.	High

4.1.3 Interface mock-up

The design of the interface mock-ups relies to a wide extent on the requirements gathered in WP 8.1. At this stage, the process is conceptuallyrather than visuallydriven. Accordingly, the available requirements have to be broken down in information and interaction needs and then collected around the identified journalistic work processes. The results will have to evolve with the new insights around the ongoing analysis.

4.1.3.1 Uncovering and selecting stories (as candidates)

Information need	**Interaction need**
Uncovering stories	
Trending information / topics	Save / mark for later
Last update (or real time).	Read saved / marked information (how long back?)
Selecting stories	
Varacity indicator	Mark for verification
Verified information (sources, entities)	

4.1.3.2 Handling supposition, speculation, rumour and verification

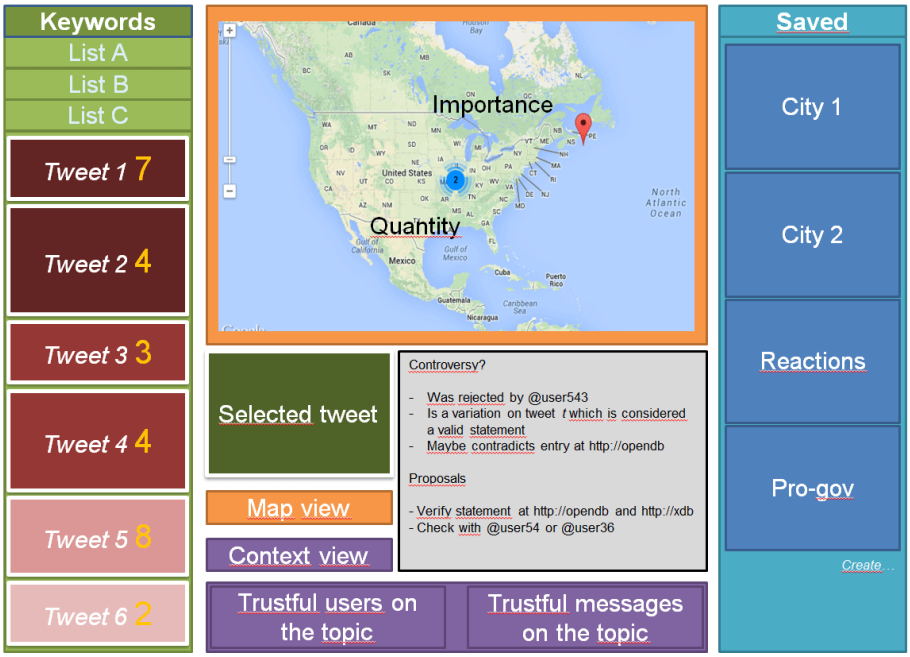
Information need	**Interaction need**
Handling supposition, speculation and rumour	
See if information is supposition, speculation or rumour.	Review steps that have been taken in Verification process
Level of Veracity	
Verification	
Marked items or information for Verification	
Verification hints: - Location - History (if available) - Etc.	Annotate Verification Process to specific item / information
Verification sources?? (e.g. contacts in house / external)	

4.1.3.3 Writing up stories and revision / re-verification

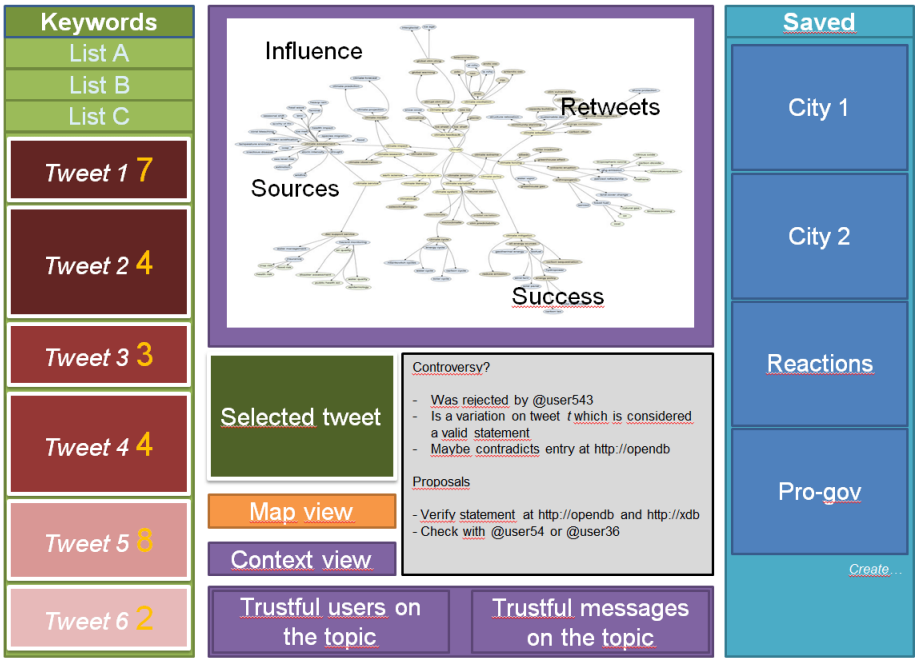
Information need	Interaction need
Writing up stories	
Revision and re-verification	
	Integrate personal ugc / feeds into tool

4.1.3.4 Practices and organization of the work

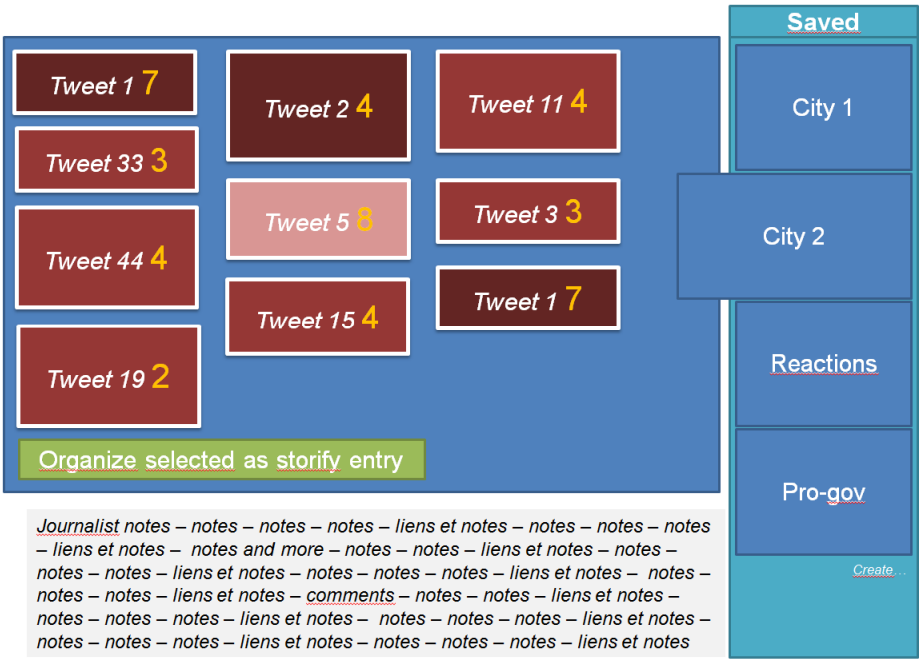
Information need	Interaction need
Practices	
Keeping in view relevant resources attached to both personal and organisational identities (e.g. Twitteraccounts)	Selecting, collating, switching betweenrelevant resources attached to both personal and organisational identities
Revision and re-verification	
Provide views for quick overviews (summaries) and detail-views for in-depth understanding	Allow cross-personnel working: sharing of all information on specific tasks.



This wireframe shows the trending tweets on the left, below the options to sort only specific emitters and editing the keywords. The map positions all messages. The selected tweet displays information about the credibility of the tweet and proposals for the fact-checking. The right displays the collections created by the journalist to sort out his or her content.



The context view displays the selected tweet in relation with the conversation around its content.



The collections on the right allow for easy sharing on other platforms, keeping track of relevant messages and keeping notes.

5 Further tasks

Further tasks include the development of an open source digital journalism use case (main responsibility USH, month 13 - 36) and the iterative evaluation (SWI, month 20 - 36). These tasks will be elaborated in further detail in due time. The following definition was made by PHEME project lead:

5.1.1 Task 8.3. Open-source digital journalism showcase (M13 - M36) - USH, ATOS, ONTO

This task will create a digital journalism showcase by extending USH's open source SwiftRiver platform with the PHEME integrated tools for veracity and socio-semantic intelligence (D6.3). At present SwiftRiver supports users working collaboratively to curate and filter real-time content from multiple channels, including Twitter, SMS, email and RSS feeds. The plugin architecture of SwiftRiver will enable its configuration and customisation, to the specific requirements of this use case.

The PHEME content analytics tools will significantly enhance SwiftRiver, resulting in an open-source digital journalism dashboard, supporting the cross-linking, verification, aggregation, and visualisation of multilingual media streams. Moreover, SwiftRiver's built-in support for crowdsourcing and collaborative data curation will enable journalists to correct text processing mistakes, which can then be fed back as new training data to the PHEME learning algorithms for veracity detection.

5.1.2 Task 8.4: Iterative Evaluation (M20-M36) – SWI, ONTO

This task will involve the deployment of the application with a set of newsroom journalists from SWI, who can provide feedback in an iterative fashion to aid in the development of the algorithms and to test effectiveness in a real-world situation. The evaluation will also involve the user group, described in T8.1. The steps of the evaluation will be: (i) Definition of the qualitative evaluation framework; (ii) SWI-based evaluation (M20 and M32); (iii) User-group evaluation (M25 and M35) (iv) Quantifying the improved efficiency and time saved by newsroom journalists by using PHEME's automated tools.

6 Literature

1. BBC College of Journalism, http://www.bbc.co.uk/academy/journalism/skills/social-media/article/art20130702112133524
2. Phillip Brooker (2012) A Study of Using Social Media in Journalism: Requirements Engineering Report, Unpublished Report, University of Warwick
3. Steve Buttry (2011) How a Digital First approach guides a journalist's work, http://stevebuttry.wordpress.com/2011/12/19/how-a-digital-first-approach-guides-a-journailsts-work/
4. Rob Procter (2012) Unpublished Interviews with members of the Guardian Sports Desk, University of Warwick
5. Rob Procter et al (2012) Unpublished transcript of Journalist's Workbench prototype demo and review at with the Guardian Sports Desk, University of Warwick