D7.1- Lessons from existing platforms

Lukasz Jonak ▸ UNIWARSAW ▸ 7/5/2016

Dissemination level	Public
Contractual date of delivery	Month 5 | May 2016
Actual date of delivery	Month 7 | July 2016
Work package	WP7 | Evaluation
Deliverable number	D7.1 | Lessons from existing platforms
Type	Report
Approval status	Approved
Version	1.0
Number of pages	51
File name	D7_1-20160705_1_DELAB.doc

Abtract
This report presents an overview of Collective Awareness Platforms which are relevant to ChainReact, with a focus on civilian reporting and the verification of such reports. A selection of relevant projects were identified and categorized according to their practices and other characteristics. Implications for the design and positioning of The Whistle were then considered.

Co-funded by the European Union

History

Version	Date	Reason	Revised by
0.7	27/10/15	Release of report on The Digital Information Verification field	Rebekah Larsen
0.8	03/06/16	First draft of supplementary material	Lukasz Jonak
0.9	27/06/16	Full draft, incorporating material from DIV report and new sections about Blockchain	Lukasz Jonak
0.95	30/06/16	Minor edits/re-structuring and formatting	Richard Mills
1.0	05/07/16	Referencing and hyper-link fixes	Lukasz Jonak

Author list

Organization	Name	Contact information
Cambridge University	Rebekah Larsen	bekahlarsen@gmail.com
Cambridge University	Matthew Mahmoudi	matt.3pehr@gmail.com
DELAB	Lukasz Jonak	lukasz@jonak.info
DELAB	Justyna Pokojska	justyna.pokojska@gmail.com
Cambridge University	Richard Mills	rm747@cam.ac.uk

Executive Summary

This report presents an overview of Collective Awareness Platforms which are relevant to ChainReact, with a focus on civilian reporting and the verification of such reports. A selection of relevant projects were identified and categorized according to their practices and other characteristics. Implications for the design and positioning of The Whistle were then considered.

Table of Contents

1 Introduction

This report presents an overview of Collective Awareness Platforms which are relevant to ChainReact, with a focus on civilian reporting and the verification of such reports. Some aspects of this report are taken from an existing report produced as part of the ESRC-funded "The Whistle" research project[1]. The authors of this report on the digital information verification field now work on the ChainReact project, where "The Whistle" is the name being used for the platform provisionally titled "TalkFree" in the project proposal. It is therefore expedient to re-use some of this existing material directly in this deliverable report.

This report aims to map the field of initiatives involved in the verification of digital information. We examined the actors, their aims, and their processes in order to identify best practices, find potential partners in the field, and determine if there were any gaps The Whistle could fill. Throughout this research, two aspects of verification were given special attention:

1. Pluralism: how can as many voices as possible be given attention in this space? More specifically, how can verification initiatives make it more likely that voices with historically less power are heard? One approach might be education—the Whistle aims to arm civilian witnesses with knowledge about digital information verification, e.g., the kinds of metadata that can make a claim easier to verify and then disseminate.
2. Speed: how can journalists and human rights defenders verify more quickly? The less time verification takes, the greater number of civilian witness voices can be heard. One approach might be the collation of current, dispersed online verification tools into one platform.

2 Digital Information - general notes

The Internet, especially after the "Web 2.0 revolution", can be seen as leveling the playing field when it comes to information creation and diffusion. The radical simplification of publishing information online (through the use of user-friendly blogging platforms, social media, consecutive generations of internet messaging tools) can be interpreted as democratizing communication. These technologies have lowered the barrier to becoming an information producer, and offer a variety of communication channels through which this information can be disseminated and consumed. These developments expand the range of sources from which citizens can obtain information, and increase the chances that an individual can obtain information related to issues they care about from sources they trust to provide accurate information. On the level of governments and institutions, this unprecedented information supply could potentially aid in evidence-based policy making.

The accuracy and relevance of digital information consumed by users is influenced by systemic mechanisms such as social network effects. Users' online experiences are greatly dependent on the actors in which they place trust - peers and organizations who either provide information or direct the user to relevant sources.

[1] Available at: http://thewhistle.soc.srcf.net/research/TheDigitalInformationVerificationField.pdf

This trust can be used as a proxy for the quality of the source and information itself. This basic trust mechanism, employing extended network neighborhoods of users as social filters, has been amplified multiple times by information technologies such as recommendation engines, tagging systems and news feed sorting algorithms.

The reality of information production, dissemination and consumption is far from perfect when it comes to the quality and relevance of the process and its products. There are two groups of problems that affect the quality of digital information production and dissemination. The first class of problems is systemic. The same mechanism that ensures that information filtered by network neighborhoods is "trustworthy" is also responsible for the decrease of novelty occurrence. The tightly knitted networks promote social control and trust, but create an "echo effect" (Burt, 2005) where the same information is repeated and reinforced within the network. In this kind of social structure, communication promotes social cohesion. Inclusion of outside, novel information, especially those kinds that could lead to questioning or verification of ingroup knowledge, could thus be inhibited. In this case validity and relevance of information is contextual, defined by in-group pragmatics.

The same mechanisms can be observed in the case of social filtering solutions and the collaborative creation of knowledge. These mechanisms can be beneficial in terms of general verification and refinement of knowledge, which is evident in many applications, especially technical (such as stackoverflow.com). However, the same mechanisms can also result in "group-think" and produce "more of the same" kind of knowledge, regardless of its validity. Similar mechanisms are evident at scales all the way up to global technologies such as Google's search algorithms, or Facebook's news feed filtering, which provide its users with answers and information based on their previous activities and interactions, keeping them in conservative information bubbles (Pariser, 2012). All these issues, wired into the fabric of popular communications tools, limit the internet's potential to become a source of objective, verifiable knowledge informing the actions of citizens and institutions.

Perhaps even more troublesome than systemic problems are the political reasons for imperfections in knowledge creation and communication processes – the intentional tampering with the generation and validation of digital information. This second class of problems includes examples such as the Russian troll army, a "web brigade" (Sindelar, 2014) supposedly acting on behalf of the government of the Russian Federation, with the aim of influencing internet discourse concerning the conflict in Ukraine. In an example from a more Western democratic sphere, Facebook (a social media company based in the USA) was recently accused of tampering with users' news feeds to present information skewed towards one side of general political arguments in US (Nunez, 2016). However, the agencies introducing misinformation to digital communication channels do not have to be global political or commercial players. Any militant group with basic knowledge of the internet can "invade" the knowledge generation platform of their choosing and abuse its mechanisms (be it searching, voting, tagging, commenting, etc.), with the intention of disrupting its functioning and skewing its message[2]. The openness and accessibility of the internet makes it easy to wage information wars on various levels.

[2] Perhaps the most spectacular example being numerous cases of "Google bombing" https://en.wikipedia.org/wiki/Google_bomb

Feedback loops that serve people more of what they appear to want, and the possibility of intentional introduction of misinformation to digital communication channels, make it desirable to develop processes and platforms of digital information validation. Such processes and platforms must provide for curation beyond crowdsourcing augmented by technologies and algorithms, for the reasons discussed above. There is a role for organisations that are trusted to ensure secure information acquisition and validation, lending their credibility to disseminated narratives. NGOs and specialized journalist entities (such as investigative journalism consortia) seem to be potentially the ideal institutions to fill this role. Their domain-specific expertise make them qualified to manage sources (verify the reports, communicate with informants and whistleblowers), identify tampering with crowdsourced verification processes or manage the verification themselves, and backup the validity of the results with their experience and position in the field. What they might need to be provided with, however, is the technical knowledge required to manage the validation/verification process, especially if this process is augmented by the use of online tools. These organisations are also limited by constraints on their available resources, and therefore tools that can increase the efficiency of this process have the potential to increase the quantity of information available that has been subjected to such validation.

3 Digital Information Verification: What is it?

Digital information verification is a burgeoning field, with entrants whose aims range from activist to scholarly to profit. It is a focus for manifold actors because of its potential to overcome some of the serious consequences of misinformation—social, political, and economic—in the Information Age. In terms of the social impact of misinformation, it can generate a climate of mistrust that complicates the work of those, like human rights defenders and journalists, who deploy fact-finding for accountability (McPherson, 2015a). Evidence of the political impacts of misinformation can be found in a variety of campaigns from the local to the global level. Such campaigns are increasingly focused on controlling the narrative via social media, as in the information war surrounding the Russian militarization of East Ukraine (Czuperski, Herbst, Higgins, Palyakova, & Wilson, 2015, p. 40). Economically, misinformation has significant costs. In 2014, the World Economic Forum ranked the "the rapid spread of misinformation online" as one of the top ten main issues the world faces. Misinformation in the financial realm can wreak havoc on the close knit global economy (Vis, 2014).

Thus, as detailed in this report, significant efforts are underway toward improved digital information verification methods and approaches, coming from many corners: information technology corporations, university labs, journalistic and media organizations, local NGOs, government-sponsored research groups, startups – the list goes on.

Digital information verification refers to the use of various tools and techniques to verify user-generated content (UGC), often created and shared on global platforms and networks. Verification is part of information evaluation and takes place after the collection of information (McPherson, 2015b). A plethora of initiatives exist that are engaged in this first step of collection, but this report is concerned with those activities that primarily focus on verification.

Verification is a process that the majority approach armed with both traditional journalistic investigation techniques and, increasingly, technology-based tools. The path, as described by Silverman et al.[3], "can vary with each fact", but there are recognized fundamentals of verification that hold true online and offline:

- "Identify and verify both original source and the content (including metadata location, date and approximate time). This step includes cross-referencing and corroboration with a variety of sources and methods.
- "Triangulate and challenge the source"
- "Obtain permission from the author/originator to use the content" (Silverman & Tsubaki, 2014, p. 122)

Thus, digital information verification is widely recognized as a human-based activity, utilizing traditional as well as new journalistic tools and techniques to verify civilian witness claims. However, with advances in technology, these human-based efforts have the potential to be optimized and improved.

As this report will show, there are a number of forms these tools and techniques can take, and uses to which they can be put.

- The forms include, variously, single-purpose tools used to verify aspects of a claim (e.g., locational search, metadata analysis, satellite imagery access), workflow platforms (e.g., allowing users to collate multiple pieces of evidence surrounding a story, building up to a determination of veracity or falsehood), algorithmic approaches (e.g., calculating 'credibility scores' of Twitter accounts or using big data analytics to identify misinformation), tool collation platforms (e.g., pulling together third party services, such as those related to image verification, via APIs), the creation and maintenance of active crowdsourcing networks (e.g., the newsroom Grasswire or the humanitarian focused Verily), etc.
- These tools and techniques can be employed in a number of contexts: the creation of new revenue models around verification for media, bolstering existing campaign and legal efforts in activism, increasing response times to humanitarian disasters, facilitating better engagement between local organizations and resource-rich global entities, etc.

This report presents a framework to assist in organizing and digesting the variety of tools, techniques, aims, and actors in the digital information verification field. In addition to the framework, we have attempted to organize the current initiatives inhabiting this field within this framework and have pulled insights around our key concerns with pluralism and speed from the categorization.

3.1 Report Aims

This report is an attempt to provide a snapshot of the current state of digital information verification field, as one of the first steps taken in developing the concept of The Whistle. We were able to collect abundant new information on a variety of creative practices—this undertaking has allowed us to more insight into questions such as: Are commercial entities pouring their resources into algorithmic verification or human-focused

[3] The Verification Handbook is an online resource, a collaborative effort managed by the European Journalism Centre, to collate "best practice advice on how to verify and use" social media.

crowdsourcing? What approaches are preferred for NGOs looking to verify local reports? More broadly, are there any knowledge gaps between these actors and their practices that The Whistle might be able to fill?

Furthermore, in researching the field, we were delighted to uncover that it is a space full of creativity and ripe for collaboration. Given its budding nature, this field's channels and spaces of communication between actors are being developed now. There are a number of coalitions and collaborations (particularly with a journalistic bent) arising in this space, and we hope this report will supplement these efforts at knowledge exchange.

In addition, this report aims to highlight some of the more interesting and creative practices in the digital information verification field.

Appendix A contains notes on some of the more interesting projects that were considered.

4 Methodology

4.1 Identifying Digital Information Verification Initiatives

As a first step, we compiled a list of initiatives that were engaged in practices directly related to digital information verification. This consisted of polling existing listservs, engaging with existing contacts, conducting online research (reviewing commercial offerings and publicized academic efforts), and contacting established experts. Specifically, we polled online communities such as the Association of Internet Researchers (AoIR); spoke to well-seasoned activists and journalists; reviewed relevant projects emerging from journalism, communication, and engineering departments of various universities; and circulated a growing list of initiatives to individuals heading coalitions to ascertain its comprehensiveness. Research was conducted predominantly in the English language, but we aim to expand on this to cover projects in other languages going forward.

This list was adjusted and expanded over the course of several weeks; at the end, we had identified a non-exhaustive list of 46 initiatives of varied aims, backgrounds, and practices. Some are directly tackling the problem of digital information verification; others have produced research or products that facilitate this process, including improved algorithmic approaches, social media analytics, online workflow platforms, and improved methods of visualization.

4.2 Gathering Information on Initiatives

The sources of publicly available information on the initiatives included consumer and client-facing websites, third-party descriptions in the media, academic papers, and materials disseminated online.

We reached out to the majority of these initiatives for further clarification of their practices, using existing contacts and publicly available information. For those willing initiatives of particular interest, we conducted semi-structured interviews over videoconferencing and the phone. This gave us insight into the specifics of their practices and aims, their main obstacles and knowledge gaps. Reaching out to organizations and the people behind them not only gave us information on current approaches (i.e., funding, technical details,

unforeseen costs, partnerships) but also insight into practical issues that would not have been readily apparent.

5 Categorizing and Analysis: Practices

During the information gathering process, we identified four main areas of practice by which to categorize the initiatives: Input, Processing, Output, and Education. High-level explanations for each of these categories and their subcategories are found below. Categorizations of the projects which have been reviewed are presented in Marimekko charts. The 'marimekko chart' (or mekko chart) is a two-dimensional stacked chart with varying column widths based on the data occupying each column in relation to the other columns; these varying widths allow for an additional dimension of data to be presented in the visualization. In the case of this project, given the weight assigned to each initiative and the number of initiatives assigned to each category, the resulting varying widths of the categorical columns denote how much developed effort is being made toward each category. For example, in the case of the Method of input Mekko chart (figure 1) one can see that Web apps are the most frequent type of input, with 20 projects utilizing this data type (resulting in the widest column). These categorizations are not always exclusive—for some categorizations it is possible for a project to fall into multiple categories.

5.1 Input

These categories distinguish the methods and sources by which initiatives acquire information; initiatives are additionally categorized by the types of data they accept as input.

Method of Input	This category refers to the tools by which information is acquired, e.g., web application forms, scraping of social media, basic email, etc. The choice of method often reflects the organization's aims and priorities. For example, the CitizenDesk workflow is built primarily for direct submissions from specific actors, meant to facilitate greater community participation in an area where reliable Internet connections are relatively sparse. On the other hand, commercial endeavors such as Ban.jo and EchoSec mainly monitor and ingest streams of public social media data with large software systems.
Source of Input	This category refers to the origin of acquired information, e.g., an actor (often journalist, activist, civilian) deliberately submitting information or automated collection of public posts by a software system. The source often reflects the initiative's aims and priorities. I.e., those initiatives whose main source of data is civilian witnesses will most likely be of an activist or journalist bent rather than commercial; those initiatives whose main source of data is automated scraping of Twitter or Instagram data might be more likely engaged in data monetization. However, these categorizations are not mutually exclusive; some journalist organizations also monitor social media for breaking stories, as do some academic projects.

Form of Input	This category refers to the types of data accepted as input on each platform. Some initiatives focus on a single type of data. E.g., Verified Pixel only accepts images as input, CitizenDesk primarily accepts data in the form of SMS (mobile texts), and some locational search engines (Yomapic) mainly accept social media posts. Often, these are initiatives with a narrower focus, be it solely image verification or a limited geographical area. Other initiatives accept a broad range of data types—images, audio, text, social media posts—and often these are initiatives with a wider aim, such as verifying a series of events with mappings of reports or broadcasting ongoing verified content. Examples of such initiatives include the Gaza Platform (Amnesty) and both established and new media organizations (Storyful, BBC, The Guardian).

5.1.1 Method of Input

- The most common method of collecting input is via web application (often a submission form on an official website—23 initiatives collect data in this manner). The second most common mode is directed collection of data from existing public platforms with varying degrees of openness—22 initiatives engage in this kind of practice. Several initiatives also accept submissions via email or mobile applications.
- The least common methods of collecting input include browser extensions, SMS ('short message service', also known as texting, via mobile communication systems), and high security anonymous dropboxes (for whistleblowing organizations in particular).
- There are arguably a number of tradeoffs between different submission tactics; e.g., providing a simple email address might encourage more submissions, but these submissions will likely not be as contextualized when compared to information collected via a form that prompts the submitter for various corroborating or explanatory data. Another tradeoff might occur depending on how well the submission process is tailored to its users. For example, in the case of low-resource communities without reliable access to the Internet but with access to SMS, a submission process based on web forms or emails might exclude many potential contributors. But as a result, such a system might not be as easily used in other contexts. In another example, a submission process might have a technological as well as expertise barriers, such as necessary knowledge of TOR for whistleblower submissions.

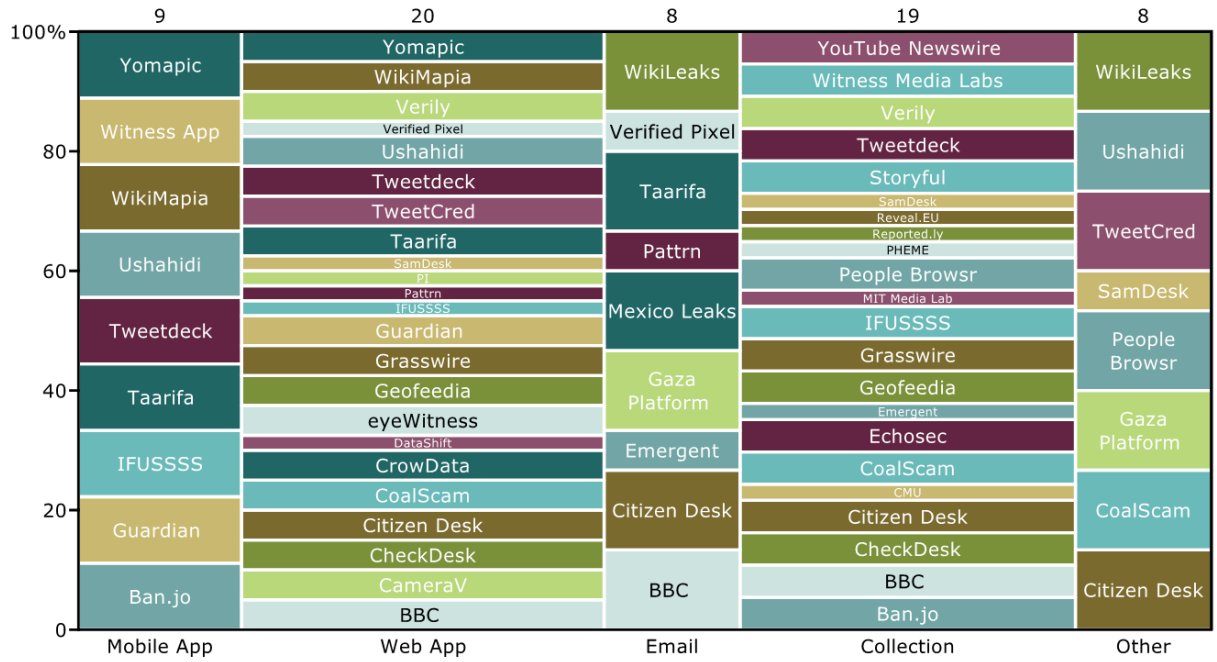

Figure 1 - Mekko chart showing categorizations of projects with respect to **Method of Input**

5.1.2 Source of Input

- Civilians (26) and journalists (24) were the two most common sources of direct input. This shows that there is a significant amount of attention being paid to the value of receiving data directly from users—while several initiatives' main operations consisted of scraping from already established platforms, many are attempting to engage more with sources of information, whether they are one-time submitters or ongoing front-end users of a platform, in the collection process.
- Other sources, not grouped into a specific category, included academics (e.g., engaging with each other on topics such as the spread of misinformation or aspects of algorithmic verification), whistleblowers (e.g., those engaging with MexicoLeaks or WikiLeaks, whose submissions are used to publicise leaked information on corrupt institutional practices), and internal sources (e.g., databases of sensitive information shared between trusted partners, as in the case of the Gaza Platform or CoalScam).

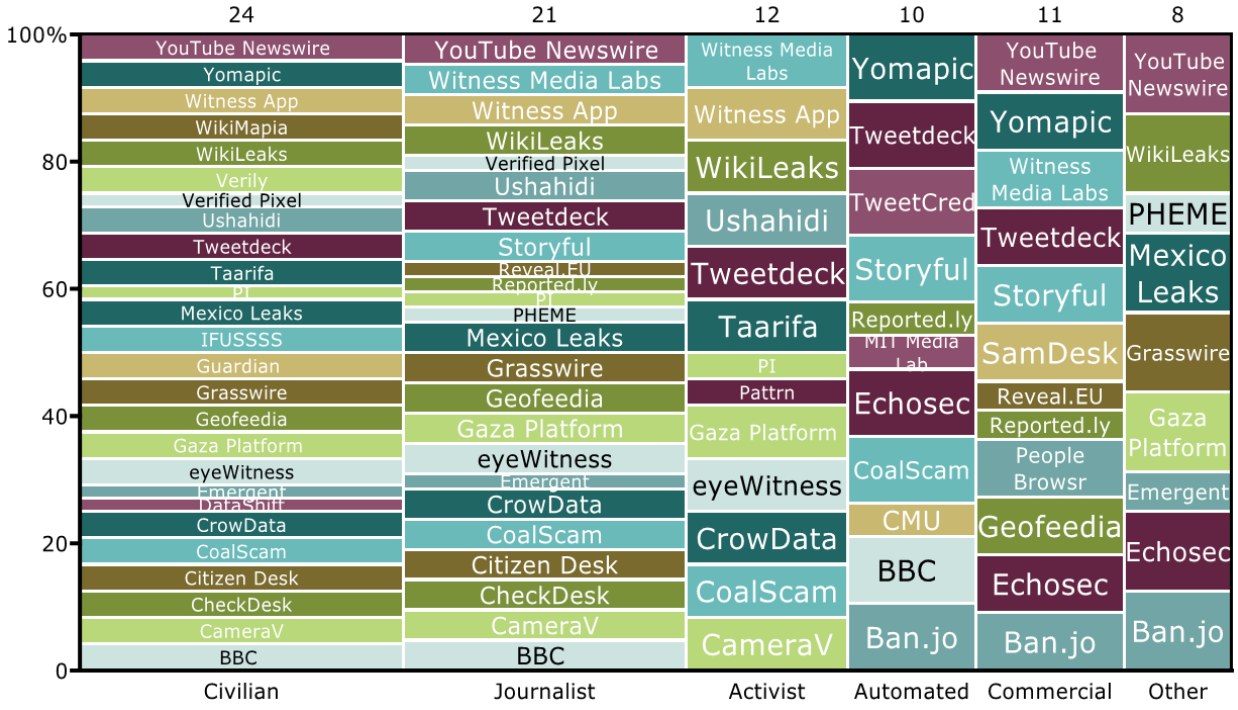

Figure 2 - Mekko chart showing categorizations of projects with respect to **Source of Input**

5.1.3 Form of Input

The most commonly accepted input is social media (24). This is mostly helpful in monitoring and verification via corroboration. This is a main focus of initiatives aiming to uncover and report breaking news, but a side focus for those organizations that use more contextualized stories around which to build long-term campaigns.

- When content is uploaded to many popular social media sites, often metadata is stripped or altered; for human rights defenders, retrieving this metadata can mean another step in the verification process. For example, there are third party services that can be employed, which tell if media has likely been altered or can retrieve metadata on the same. Given this value, a unique selling point for multiple initiatives is thus maintenance of the chain of custody – a kind of guarantee that content is original (e.g., IFUSSSS and eyewitness to Atrocities).
- The least common types of input include audio and SMS. The two initiatives that accept these input mediums are reporting applications built with specific users and contexts in mind.
    - CitizenDesk was built specifically for a local journalism endeavor – the community does not have widespread access to smartphones or consistent Internet connection, and so the platform was built with SMS as the main method of information delivery.
    - Witness is one of a variety of "panic-button" apps that allow users to live stream location data, in addition to audio and video, to preset emergency contacts. It is meant for users in high-risk situations; in addition, it can be used to document rights violations for further campaigns or possibly lawsuits.

- Most initiatives accept several kinds of data, especially those whose submission process is a simple email address or webform that accepts manifold data types. Initiatives have to make careful decisions in this arena, as there are a number of tradeoffs that exist between how inclusive a platform is of many data types and growing cost that comes with an increasingly flexible system.

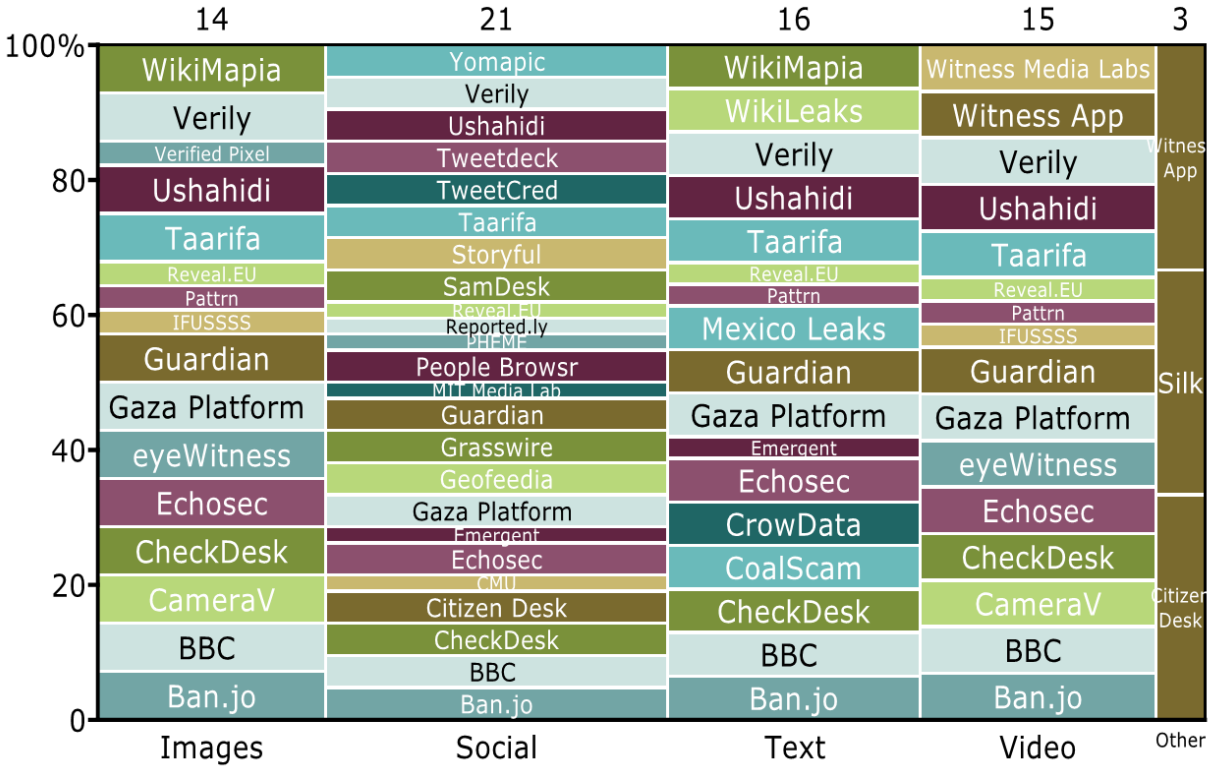

Figure 3 - Mekko chart showing categorizations of projects with respect to **Form of Input**

5.1.4 Input - Implications

In studying this step in the verification process, we have realized there are a number of tradeoffs that must be taken into consideration. These include lower developmental and operational costs v. more flexibility, and the ease of submission v. ease of verification. We have also realized that the tradeoffs are not the same in every context. E.g., those initiatives looking to build a trusted and easy communication channel with a particular community might prioritize a specialized, less flexible system.

One of The Whistle's main aims is the empowerment of the civilian, the unlikely and uninformed (in verification) witness, in the human rights context. Empowerment of the civilian who is submitting a claim can happen at the input stage by including as much corroborating information as possible. In this way, much of the verification onus is taken off the reputation of the source. Including more corroborating information (metadata, other links to similar claims, etc.) also quickens the verification process for the human rights defender, or the person analyzing the claim; they do not need to spend as many limited resources in finding and retrieving the corroborating information. Thus, The Whistle team is focusing energy on making

submission as low cost as possible for civilians but also prompting them for all kinds of important corroborating information.

5.2 Process

Process refers to how the various initiatives manage and derive insight from the data received. We categorized initiatives based on the nature of this processing (whether human-focused or automated) and method of processing (looking at the various ways in which these initiatives compared, analyzed, and manipulated data).

Human vs. Machine	This category refers to the place where most of the processing takes place, and makes the distinction between two types of platform: human-centered workflow and automated 'black box' analysis. These categories are in practice not mutually exclusive, e.g., some workflow platforms pull in third-party automated services for processing. However, each initiative prioritizes one form of processing over the other.
Method of Processing	This category refers to the various methods of processing in which an initiative engages, e.g., cross-referencing with other data banks or third party analyses, or crowdsourcing for uncovering corroborating evidence. Many initiatives focus exclusively on one kind of processing while others try to incorporate a range of different processing methods.

This category includes those initiatives that manage, organize, and/or analyze data submitted. We categorized based on the preference of initiatives for human-centered verification, and the methods of processing preferred or proffered by the initiative.

5.2.1 Human vs. Machine

- Between user-focused workflows and automated processing, the most common practice was the former—24 initiatives use workflows of some sort to facilitate collation of information around claims. Those users of the platform adding to the collation could be whoever is given access to the platform—any user online (Grasswire) to specific communities grounded in a geographic location (CitizenDesk) to the internal workflows used by newsrooms (BBC). On the other hand, 16 of the initiatives reviewed were classified as relying mainly on automated means of processing.
- Those initiatives that focused mainly on automated processing included a large proportion of those classified as academic (exploring new ways to monitor and verify social media) and commercial (using large systems for widespread social media monitoring).
- The spread and popularity of these two methods supports the current accepted view that verification remains a human-centered activity; not only was this reflected in practice, but multiple interviewees for this report, from commercial, activist, and academic backgrounds, explicitly expressed this view.

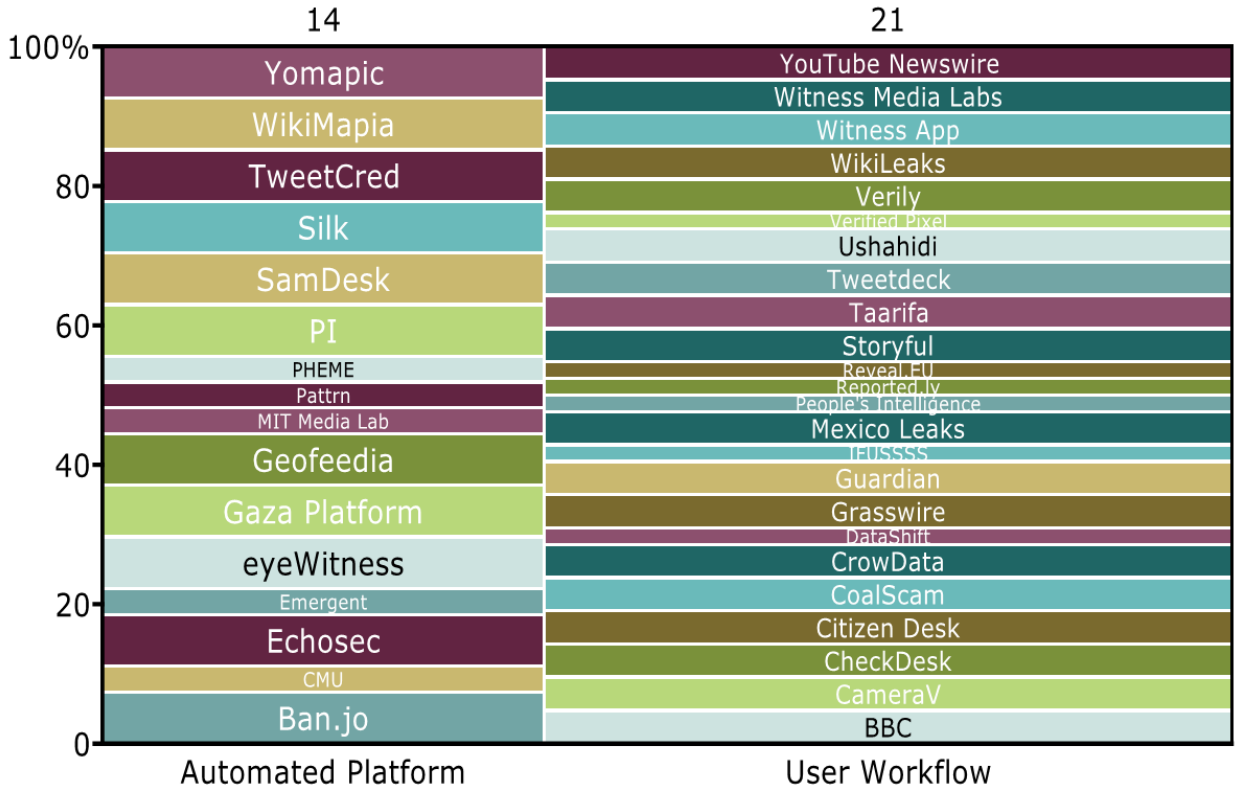

Figure 4 - Mekko chart showing categorizations of whether each project uses primarily automated **machine** analysis or relies on a **human** user workflow.

5.2.2 Method of Processing

- Of the six different subcategories, the most common processing practices were cross-referencing—both third party services (16 initiatives) and databases (17 initiatives). Cross-referencing could be done either by human (with the help of collaborative workflows) or by machine (e.g., data is pulled from separate databases for comparison).
- Several initiatives used multiple approaches to verification: cross-referencing databases as well as the results from third party services, appealing to communities for insight and collection of supporting information, running information through automated analysis, etc. (Examples include Verily and Grasswire). Others initiatives were included in this report mainly because of their products' use in facilitating verification; these focused on fewer or just one approach. One example is TweetCred, an academic project focused on a browser plugin that provides an algorithmically based rating of the credibility of Tweets.
- Those initiatives that employed crowdsourcing techniques fell into either journalistic or NGO categories; the NGOs that used crowdsourcing made it their main focus, though their communities often used other tools to corroborate any claims. This approach often involved proffering a select claim for analysis, and the community would be encouraged to produce evidence in a structured manner, e.g., posting corroboration or debunking evidence in a kind of contributory narrative under

each claim. The media organizations involved in crowdsourcing (BBC, The Guardian, Grasswire) maintain online spaces in which users can interact, post content, and contribute content to any specific queries the organization might have about a certain event.

- Within the 'Other' category, two initiatives (CameraV and eyeWitness to Atrocities) provide a unique treatment of data – one of their main purposes is to create a secure repository and also an unbroken chain of custody for sensitive data.

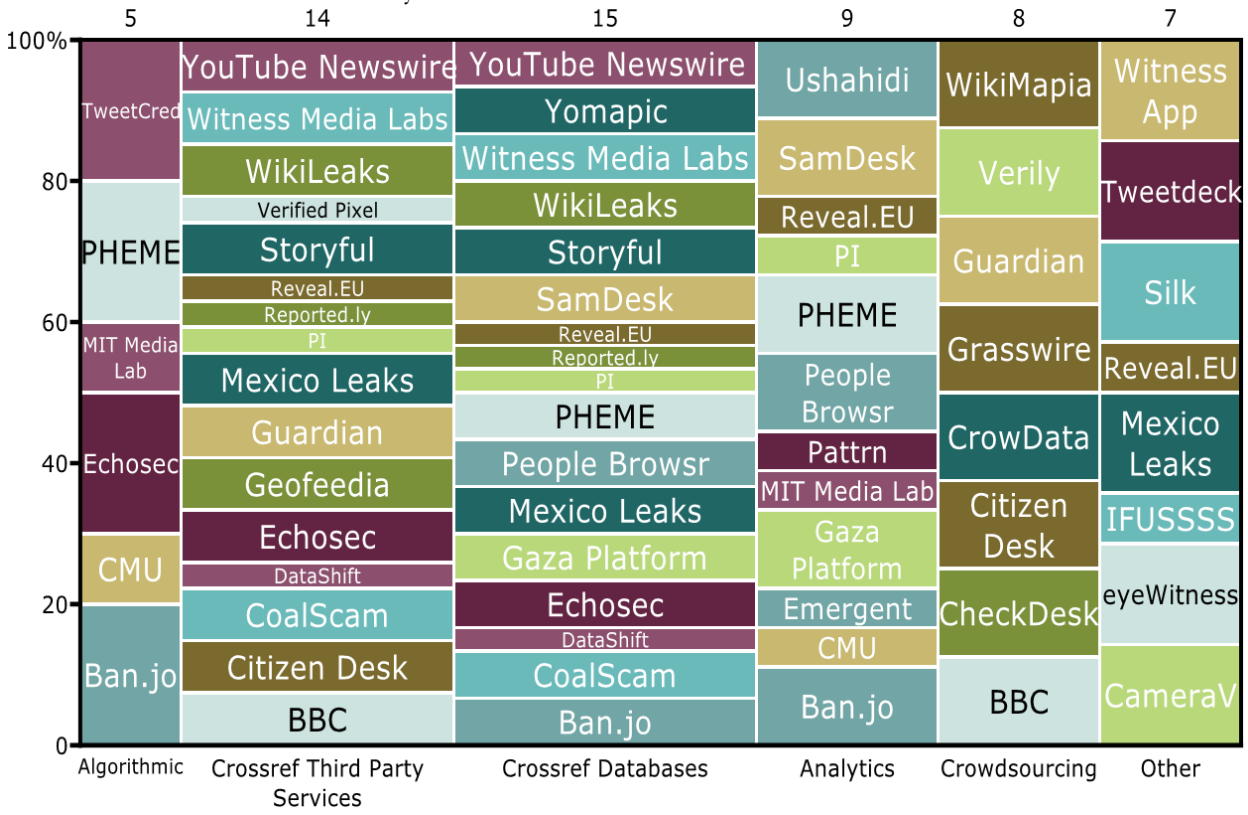

Figure 5 – Mekko chart showing categorizations of projects with respect to their **method of processing** data

5.2.3 Process - Implications

The Whistle aims to make a reporting platform flexible in that it can be tailored to diverse data – but keeping in mind the tradeoffs with cost. Realizing that each human rights organization might have its own set of verification issues and systems has led us to consider a platform that prioritizes interoperability. This will allow for a flatter learning curve for human rights defenders who are integrating The Whistle into their work, giving them more time and resources to verify more claims.

In keeping with most of the initiatives, The Whistle will focus on human-centred processing. More specifically, The Whistle aims to provide a workflow environment tailored to the organization, but in addition, pulling in several commonly used third party services via APIs. In aggregate, this will save human

rights defenders large amounts of time in verifying content, and allow them to process more claims—also contributing to the growing plurality of the field.

Another piece of insight pulled from this research is the careful attention that must be paid to the security and privacy of processing sensitive data. The Whistle will be publishing blog pieces on this topic (as it is not a focus in this report), and prioritizing this aspect in platform development.

5.3 Output

Output refers to the main product at the end of processing, and how the various initiatives package and disseminate these findings. We categorized initiatives based on the form that delivery of output (if any) comes in, the extent to which such output is disseminated, and whether or not they made a 'verification determination'.

Delivery	This category refers to the manner in which results are conveyed, closely tied to the goals of the initiative. Examples of different delivery systems include dashboards, databases, publications, visualizations, etc. The Gaza Platform produces a publicly viewable and interactive mapping of reported events while Eyewitness to Atrocities produces a closed report on each incident that goes into a protected database. The former aims to facilitate further research into accountability and human rights violations; the latter is meant to produce a space for protected information and maintain the chain of custody.
Dissemination	Dissemination refers to how widely and to which audiences initiatives publicize their reports. Some commercial initiatives tend to produce private reports (or products that create the same) for their clients; human rights initiatives collecting submissions comprised of sensitive information will also tend to keep such information private. Other initiatives aim at creating mappings of public data to aggregate location information in a more publicly accessible format (e.g., nonprofit Wikimapia and commercial Yomapic).
Determination	Some initiatives provide a determination of verification, e.g., the Grasswire or CheckDesk platforms, whose communities collate evidence into stories to the point a consensus is reached. Other initiatives provide pieces of information that can facilitate this determination, e.g., Verified Pixel (providing access to several pieces of such information on one platform) or locational search engines such as Geofeedia.

This category includes those initiatives that create output of some sort from the data received and processed. The subcategories are based on method of delivery, the degree to which initiatives disseminate processed information, and whether they make a verification determination.

5.3.1 Delivery

- The most common type of delivery of output among the initiatives was via a dashboard (22). A dashboard could also be described as the frontend of some systems, often accessed via web application or app interface. A dashboard could be used to present a variety of 'stories' or 'claims' with any associated contextualization and corroboration (e.g., Guardian, Grasswire, or CheckDesk); alternatively, it can be used to further analyze and explore collected and processed data (e.g., Gaza Platform).
    - The popularity of the dashboard supports the general trend toward the human-centric approach to verification, where evidence is gathered and presented almost in a narrative.
    - The dashboard approach is particularly popular with crowd-sourced, citizen journalism, and activism initiatives. It can provide public access and participation to the act of verification. Given the preferences of the initiative, it can allow the user to contribute to the veracity of the information. Users are defined again by whoever is given access to the platform—anyone wishing to join online (Grasswire) to specific communities grounded in a geographic location (CitizenDesk) to the internal workflows used by newsrooms (BBC). Rather than present information decontextualized as truth or fact (as a published story on the BBC), these initiatives can also present the pathway to that determination, and maintain an ongoing conversation or dialogue surrounding the verified state of a claim.
- Delivery can also take form as a derivative product of the verification system: a publication (14), which can take the form of reports or articles. This is a popular form of delivery for those organizations that have made a verification determination internally and disseminate the results as truth or fact (e.g., traditional and new media organizations BBC, Guardian, Storyful). Additionally, this is a means employed by advocacy organizations that sometimes also use dashboards, when they arrive to a consensus on a certain claim or wish to incite action on a particular issue (e.g., CheckDesk, WikiLeaks).
- Other forms of delivery include output that takes the form of visualizations—geotagged maps, interactive graphics and charts, etc. Both commercial and activist organizations use this form, especially if they are focused on one issue or product. Examples include CoalScam, an activist organization meant to map the impact of coal mining in a particular region, or Yomapic, location-focused social media monitoring effort.
- For a few organizations, delivery is not at an unspecified third-party user's request—rather, data, once processed, is housed in a secure location and only recalled for a specific purpose. E.g., eyeWitness to Atrocities records and embeds metadata in submitted content, then carefully preserves the chain of custody when the data is saved to their secure servers.

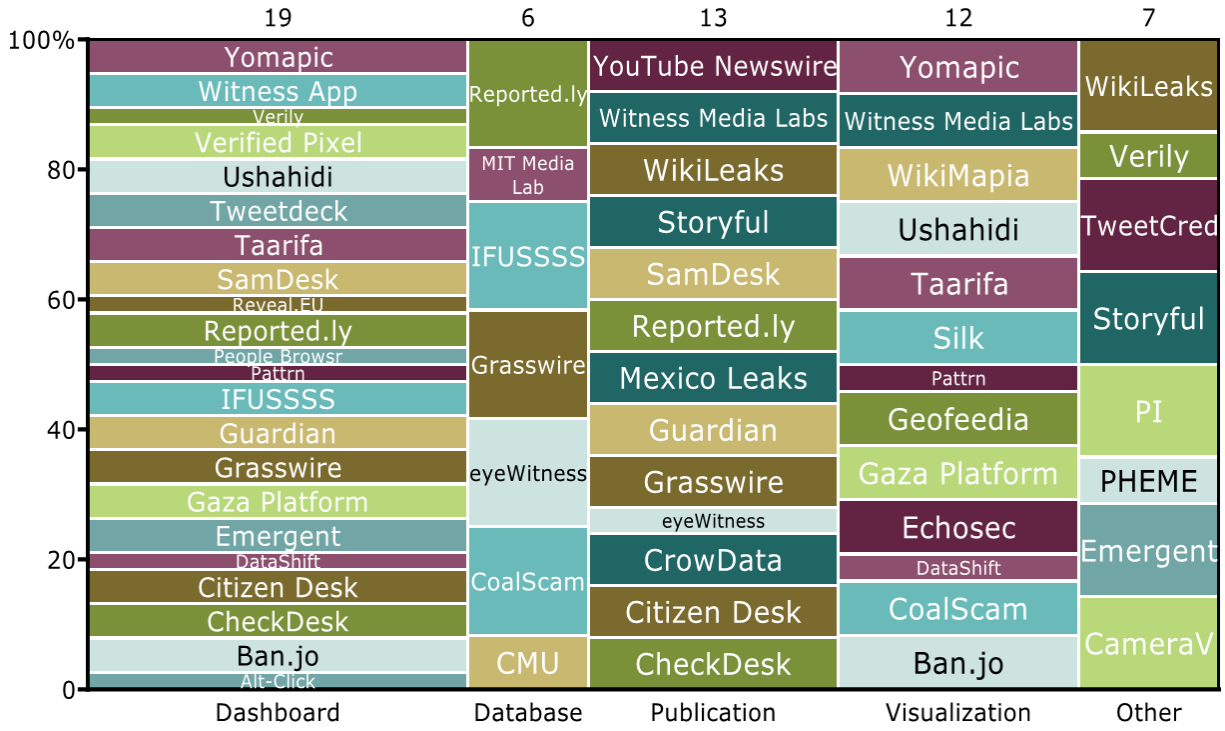

Figure 6 - Mekko chart showing categorizations of projects with respect to their method of **delivering output**

5.3.2 Dissemination

- Many initiatives reviewed do not make their output easily and readily available to the general public; of those reviewed, 22 restrict dissemination in some manner, often based on the user type. Commercial initiatives usually restrict their work to clients or behind paywalls, though they may publish reports showcasing the nature of their product (e.g., EchoSec, Geofeedia, SamDesk). Some civil society and activist initiatives produce data specifically for their partners or users—the motivation behind this practice is often tied to the sensitivity or specific, narrow use of the data (e.g., CrowData, eyeWitness App).
- Out of the initiatives reviewed, 13 generally publicize their data. The motivation for this practice is tied to the nature of the initiative; this category includes news organizations that want to attract the largest audiences (e.g., BBC, Reportedly) and activist organizations whose main goal is widespread advocacy or campaigns (e.g., CoalScam, Gaza Platform, Wikileaks). This category also includes those commercial ventures whose products become more attractive the larger their datasets and comprehensive coverage, and who don't monetize via subscription (e.g., Yomapic, YouTube Newswire).
- Comparatively few initiatives completely internalize their data and findings; out of the four that fall into this category, three do not publicize because they are academic projects in progress. However, there are those apps whose main purpose is not to disseminate but to securely preserve and contextualize data (e.g., eyeWitness to Atrocities and CameraV).

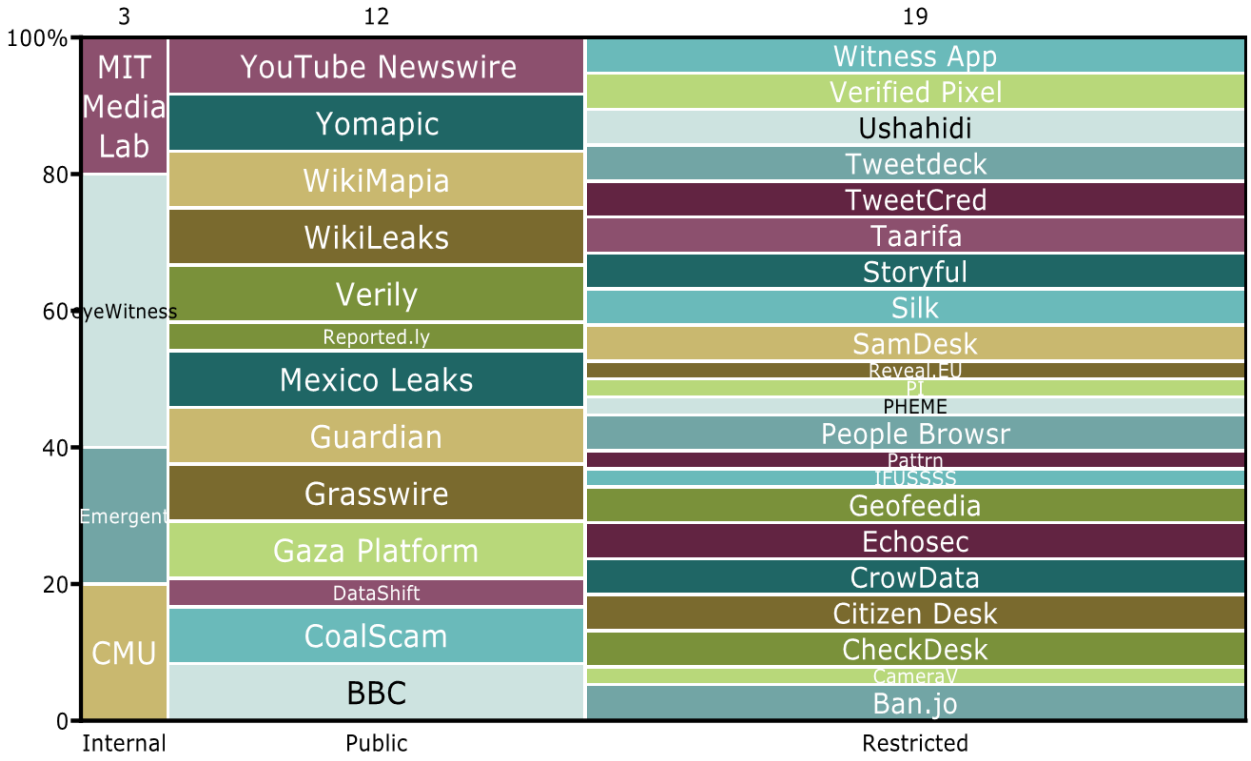

Figure 7 - Mekko chart showing categorizations of projects with respect to how they **disseminate** reports

5.3.3 Determination

- Not every organization processing content makes a definitive claim as to its veracity. In fact, 12 in this report do not make any claim at all; many simply provide general tools or information that can assist others in making determinations.
- Out of the initiatives reviewed, 14 give an indication as to the likelihood that a claim or story is true. This category includes many of the academic and commercial initiatives that provide technological tools to assist in verification. As in the case of Verified Pixel, theirs is a deliberate decision to refrain from marking content as 'verified' or 'not' after running it through various third party tests. This is because they recognize that the verification process in praxis is still very much human-centric. The nature of this indication can also take the form of credibility ratings (e.g., TweetCred) or workflows that collate all of the evidence, both supporting and otherwise (e.g. Verified Pixel).
- Out of the initiatives reviewed, 17 make explicit claims as to verification or offer content that is assumed by its consumers to be verified. This includes news outlets that report on claims (BBC, The Guardian), ventures whose product is digital information verification (Storyful, YouTube Newswire), and those organizations both activist and commercial that focus on the preservation of chain of custody (IFUSSSS, CameraV, eyeWitness). There are also those initiatives that publish stories as

'verified' when a story has crossed the consensual threshold of verified via crowdsourced corroboration.

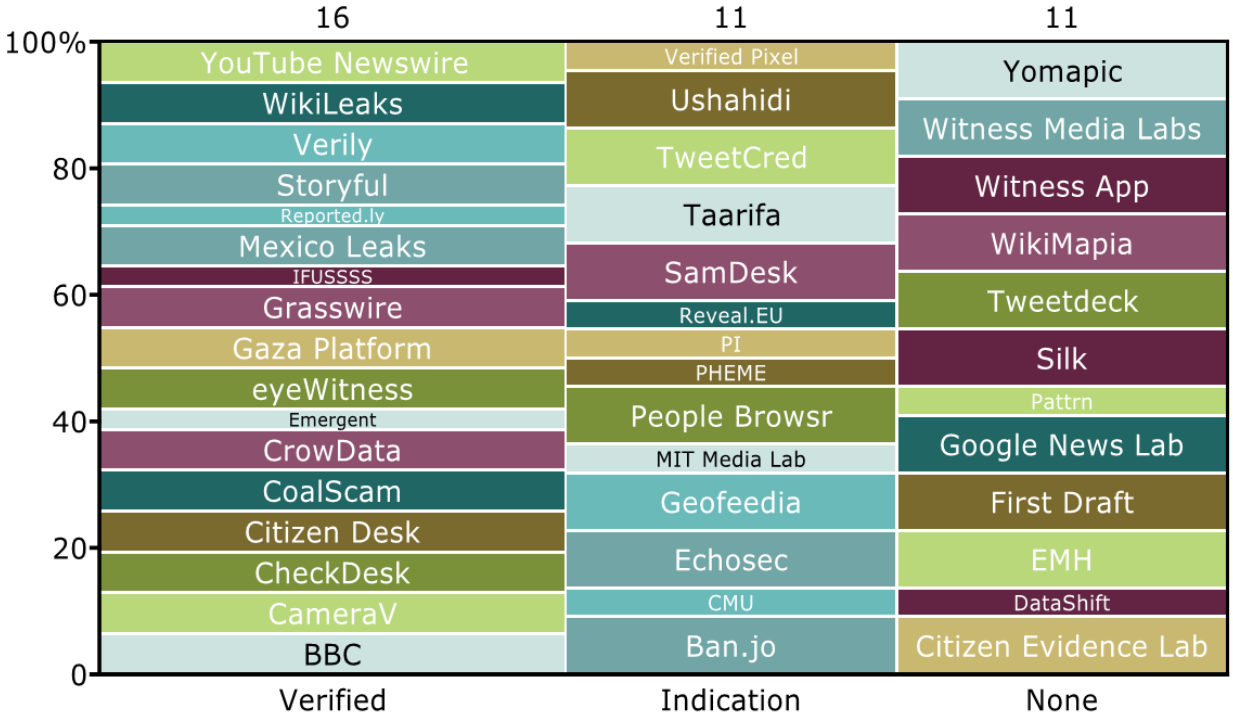

Figure 8 - Mekko chart showing categorizations of projects with respect to whether they make a **verification determination**.

5.3.4 Output - Implications

The method of delivery and degree of dissemination is one way to attract contributors to a platform; The Whistle will only achieve its aim to empower citizens if they are aware of its existence. In addition to inclusion in well-publicized reports and campaigns, one way to spread the word is via partnership with legacy NGOs or well-known human rights organizations. The Whistle is envisioned as a flexible system that could be folded into the work stream of more than one initiative. This flexibility is the main value-add in the output process, after the input and processing stages, where information is contextualized and presented to the human rights defender in a workflow. The Whistle will allow the human rights defender to easily sort and share (via a standardized, common format such as an excel spreadsheet) the metadata and contextualization collected on a claim. This feature will allow for more easily shared findings between collaborating initiatives, and a faster uptake integrating The Whistle into existing systems, both technological and otherwise. This in turn will lead to more processing of claims with limited resources.

In terms of determination of verification, The Whistle aims to provide tools to help human rights defenders make an assessment, but the platform will not make the verification determination itself.

5.4 Education

Education refers to any measures taken to inform users and contributors about verification or best use of the platform. One of the main problems the Whistle aims to ameliorate is the lack of literacy around digital information verification, particularly among civilian witnesses. Thus, we examined the different ways in which initiatives attempt to educate their users, the public in general, contributors, or key actors in the verification space. We categorized initiatives based on the nature of these measures and the key groups they are targeting.

Nature of Education	There are a variety of ways in which the initiatives in this category approach education. Some provide education for the use of their tool or platform; others make general education of practice a main focus—many do not process data but provide venues for discussion and practice sharing between actors (e.g., Eyewitness Media Hub, Witness Media Lab, First Draft Coalition). Other initiatives have a data processing component but also include educational tools for their users on general verification practices (e.g., Verily).
Target Group	This category refers to the main group at which any educational measures are aimed—different examples include civilian 'digital detectives' (Verily), journalists who are participating in coalitions (First Draft), or company employees who are using enterprise software for social media monitoring and analysis (EchoSec).

This category includes those initiatives whose aim includes knowledge transfer, whether for specific platform use or more general education on digital information verification practices and risks. This is key to The Whistle's aims as well, as enhancing the knowledge of civilian witnesses (who do not have an established reputation and are not embedded in well-known networks) might be one way to increase the pluralism in the human rights space (McPherson, 2015a). This report touches on two aspects of education: the groups at which these education efforts are targeted, and the ways in which they are targeted.

5.4.1 – Nature of Education

- The educational aims in the digital information verification space range from introducing users (whether at the frontend or backend) to the platform or initiative, providing more general background on digital information verification practices, expanding awareness of current tools and resources, and giving ethical, legal, and security guidance on collection and publication of content. The initiatives intended for more specialized participants often provide spaces or venues for discussion and knowledge transfer. Out of the initiatives examined, the most common approach was the use of online modules (instructional videos, presentations, or text online).
- The educational practices often match up with intended audiences, based on their existing knowledge and the intentions of the organizations:
    - Those initiatives focused on citizens and crowdsourcing often provide educational tools aimed at those new to the field, such as online modules, toolboxes, and community fora that allow for knowledge exchange (e.g., Verily). This is especially the case for those initiatives that rely on community consensus in analyzing a story or claim; users are exposed to different approaches and techniques as they engage with each other in the verification process (e.g., Grasswire).

- Those initiatives that are built up around a specific topic or partnerships (e.g., Taarifa, CheckDesk, CitizenDesk) often provide hands-on training to their target clients, partners, or communities.
- Those initiatives aimed at professionals or practitioners with some exposure to digital information verification and investigative techniques often focus on creating a knowledge exchange space or exploring promising new approaches (e.g., First Draft and Witness Media Lab.
- Those initiatives whose main project is a relatively narrow use-specific application or platform often focus in educating surrounding its function (e.g., Silk); they often also include information for contributing developers, as many of these projects are open source (e.g., CrowData, Ushahidi).

- Other practices and measures to inform users include presentations of products at industry conferences, various options to contact the initiative and developers directly, and blog posts that elaborate on specific issues or concepts.
- Several organizations often direct their users to resources collaborated and maintained by others, particularly in the activist/journalist/citizen sphere. E.g., Grasswire users often employ Citizen Evidence Lab's YouTube Data Viewer when verifying content; MexicoLeaks directs its potential submitters to Security In a Box and instructions on using TOR.

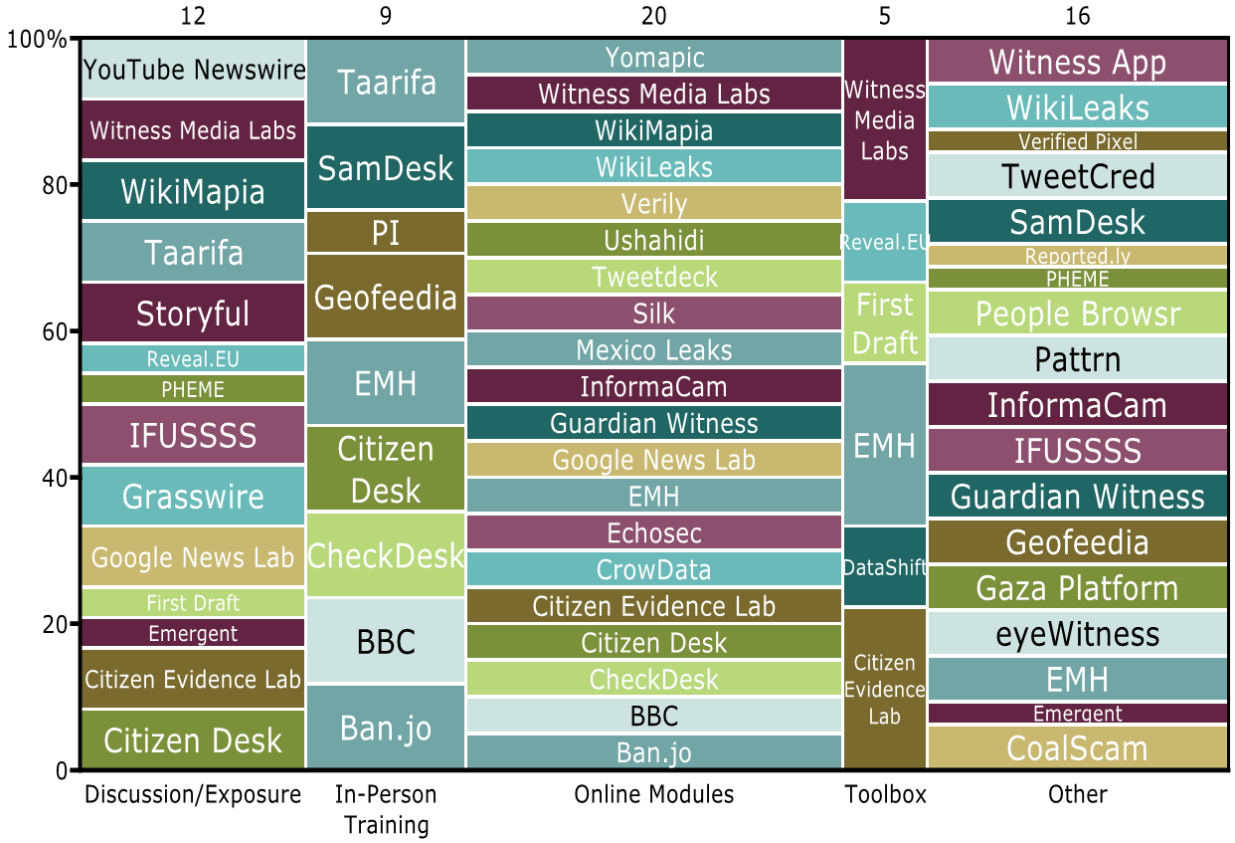

Figure 9 - Mekko chart showing categorizations of projects with respect to the **nature of education** offered to users and contributors

5.4.2 – Target Group

- Out of the groups targeted in the various knowledge-transfer efforts, those two most common were citizens (23 initiatives) and journalists (26 initiatives). Though the educational efforts take different form, depending on the intended group, several initiatives targeted citizens, journalists, and also activists (e.g., Witness App, eyeWitness, EMH, CoalScam, etc.). Not only is this in keeping with the general observation that people are the engine behind digital information verification, but it means that there is significant room for collaboration between many of these efforts.
- Educational efforts include those focused on introducing users to a platform or space (e.g., IFUSSSS), and those aimed at more general education surrounding digital information verification (e.g., Citizen Evidence Lab). This general education often takes the form of either creating spaces for experts to convene and trade insight/practices (e.g., First Draft), or the collation of tools and tutorials (e.g., Verily).
- Less common groups targeted, at least in terms of openness and publicity, include:
    - Whistleblowers (with education focused on the technical and logistical aspects of secure submission);
    - The clients of commercial outfits whose platforms and products are introduced behind closed doors or more generally at industry conferences;
    - Entrepreneurs or other innovators in the social media space, whose funding can be seen as a way to boost the economy (e.g., the EU-funded Reveal[4] project);
    - Intelligence professionals (including law enforcement), investigators embedded in think tanks, and agencies that are increasingly using such tools in open source intelligence (Czuperski, Herbst, Higgins, Palyakova, & Wilson, 2015, p. 40).

[4] http://revealproject.eu/about-reveal/

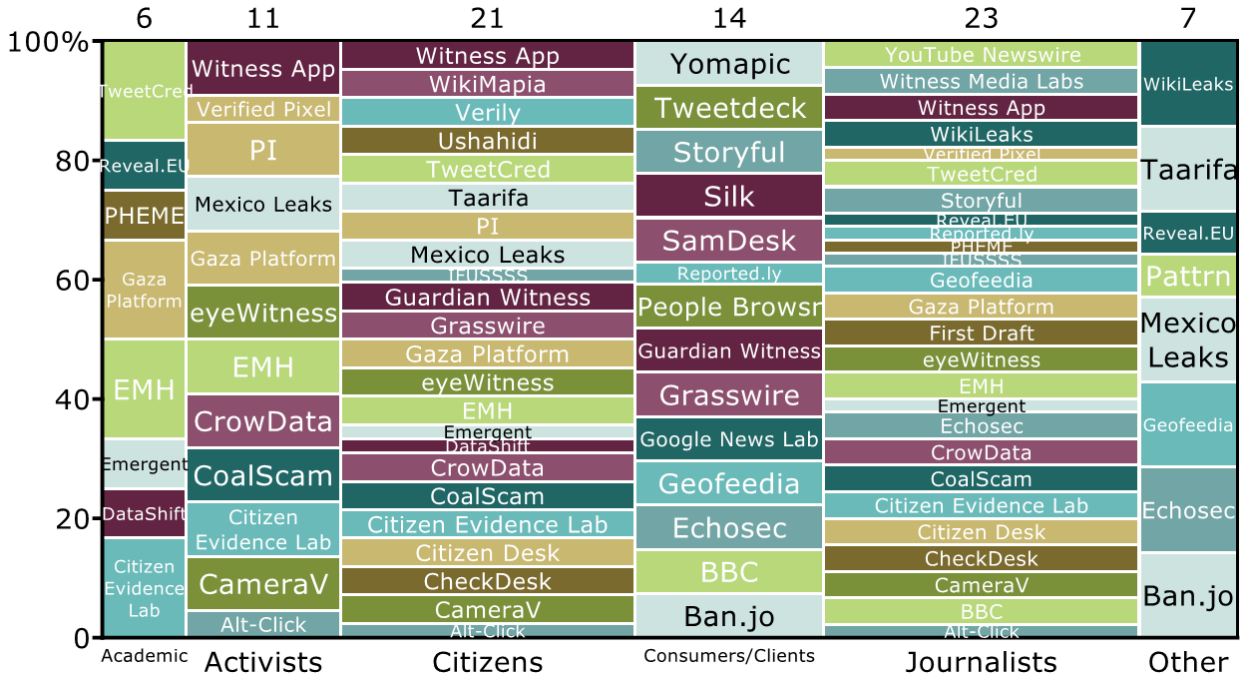

Figure 10 - Mekko chart showing categorizations of projects with respect to the **target group** for education measures

5.4.3 – Education - Implications

As noted, the education of civilians is key to increasing both the pluralism of voices in human rights and the speed at which claims are verified. Civilians must understand the state of the verification process—the high traffic of information, the time-consuming process to verify each claim—for them to understand the kinds of metadata and corroboration that will make their content more easily verified and thus heard.

The Whistle aims to not only provide for general education in verification—directing civilians to online fora and toolboxes—but will attempt to educate during the submission process. This will occur by prompting civilians for certain kinds of corroborating information and explaining the import of the same. This will entail a carefully designed submission process, balancing tradeoffs between complexity and amount of information received—a process that is not so time intensive or complicated that it is off-putting to submission, but also one that collects as much useful corroborating data as possible.

The education space in particular is ripe for collaboration—the knowledge built up in expert groups or active online communities, if packaged in easily digestible formats and widely disseminated, can go a long way in increasing the pluralism of the digital information verification space. Those initiatives with specialist insight into certain aspects of digital information verification (e.g., user interface design, how to best intersect with particular communities) can also contribute greatly to this space.

6 Categorizing and Analysis: Characteristics

In addition to classifying based on practice, we also categorized initiatives based on characteristics: maturity, funding, and aim of the organization. High-level explanations for each of these categories are in the table below.

Maturity	Within this category, we gave each of the initiatives a ranking in maturity. We considered various factors, based on publically available information. These factors included the number and types of users, how far along the project seemed to its stated goals, if the project (if it was seeking funding) had been described at a certain level of maturity on databases such as Crunchbase, and the degree to which the platform was being used for its intended purposes.
Funding	Within this category, we attempted classify sources of funding for each project, so as to better understand not only their aims in digital information verification but also the constraints. Though some projects or initiatives publicized very clearly the forms of their funding (especially those receiving well-known grants or support from brand name institutions), others we made an educated guess based on the nature of their funding based on the nature of the organization. E.g., those commercial, closed platforms that license and customize enterprise software are likely to receive revenue for their products.
Nature	Within this category, we attempted to sort the initiatives based on the nature of their goals—e.g., whether an initiative's ultimate aims were academic, journalistic, commercial, activist, etc. There is considerable overlap in this kind of grouping; several organizations possess manifold aims, e.g., those academic projects whose case studies relate to human rights and activism; those commercial and journalistic entities who comprise coalitions, etc.

6.1 Maturity

Of the 46 initiatives surveyed, we classified 28 as fully developed. Of the remaining initiatives, nine were classified in beta, two were in alpha, and seven were in pre-alpha. Fully developed, for our purposes, indicated whether or not the platform was being utilized as intended, though additional updates might be ongoing.

6.2 Funding

We classified different types of funding into nine different categories (full descriptions in the visualization online): revenue, academic, corporate, foundation grant, internal (if the initiative was the project of a larger entity), investors, donations, government, and other.

Of the 46 initiatives surveyed, the two most frequent forms of funding are revenue and internal 11; academic and government funding combined 12. The least common form of funding is direct donations – only three surveyed initiatives solicited donations directly from the public.

6.3 Nature

We classified initiatives based on seven different aims: academic, activist, coalition, commercial, journalist, humanitarian, and whistleblowing. (The descriptions for each category can be found in the online visualizations.) The category with the most initiatives was commercial (15), closely followed by both journalistic (13) and activist (13). The categories were not mutually exclusive – several initiatives fell into more than one category (e.g., Google News Lab fell into the Coalition/Journalist/Commercial categories; PHEME fell into Academic/Coalition/Commercial). The least common type we reviewed for verification practices was whistleblowing. Though it arguably could fall into a journalist or activist category, its aims are sufficiently specific that we thought it warranted demarcation.

6.4 Characteristics: Findings and Implications

Of those 18 initiatives not classified as fully developed, only four were commercial in nature: Reportedly, Pattrn, Ban.jo, and IFUSSSS. The rest were mainly academic, activist, and humanitarian.

- These findings hinge heavily on our working definition of 'fully developed' (cultivated from the degree of intended use). Many of the commercial and journalistic initiatives are integrated into well-developed practices or platforms. One example could include the BBC's User Generated Content hub, which has been in existence since 2005 and is an internal department facilitating monitoring and verification for the organization's main publishing platforms.
- Some types of initiatives are more easily built than others; developing a coalition does not require the same kind of deliberate, many-iterations approach that developing a specialized newsroom workflow platform might. Thus, these types of initiatives are more easily classified as fully developed though they might not have the same amount of infrastructure to develop.

There is significant crossover in terms of funding sources when it comes to grants from foundations, government schemes, and academic/journalist/activist projects. Two examples include the Knight Foundation (a US-based journalism foundation with well-publicized grants) and Horizon 2020.

There was not much publicly available evidence that commercial initiatives (developing actual platforms and software as products) collaborate extensively with other initiatives. However, the newsrooms with a commercial bent, who are monetizing expertise in the human aspect of digital information verification, seem to collaborate with other actors. This is exemplified in Storyful, a social media-focused news agency. It was acquired in 2013 by News Corps, and despite its proprietary nature, has since worked with and for corporations and NGOs to create spaces and platforms for quality citizen journalism and human rights reporting: YouTube's Human Rights Channel, YouTube Newswire, First Draft Coalition, and Witness Media Lab. Members also contributed to the Verification Handbook.

Of the 46 initiatives, 17 were categorized with more than one 'aim'. The biggest overlap could be found between activist and journalist categories, arguably because of the importance in civilian reporting and

participation for both. Interestingly, the only journalistic projects not falling into multiple categories were the hubs and internal projects of established conglomerates (e.g., The Guardian and BBC). Examples of initiatives that fell into this overlap include Eyewitness Media Hub, CameraV, eyeWitness to Atrocities, and the Gaza Platform. In another area of overlap, government funded projects were headed by or included commercial and academic partners. These projects were funded as part of research and innovation policy; the Reveal Project and the academic collaboration PHEME are both funded under the European Commission's research and innovation programme, Horizon 2020.

Academia is where ideas furthest from the field are developed—basic research is a public good funded by universities. This could be why we found that the academic projects were those mainly focusing on algorithmic verification and not operating fully developed platforms; there is a general consensus that verification of content is too nuanced a process to be completely automated. However, initiatives such as Verified Pixel, whose main function is the collation of several third party services that can assist in verification onto one platform, might also include algorithmic services in its collection.

7 Summary – Field Overview Implications for The Whistle

Undertaking this field overview has helped The Whistle team structure our aims and focuses; we have worked to identify best practices, gaps in offerings, potential partners, and those areas in which efforts should not be replicated. This was done using two loose metrics: increasing pluralism and increasing the speed of processing.

Based on these findings, The Whistle aims to:

- Assist fact finders, particularly in the human rights and citizen journalism realm, in speeding the verification process. For example, we will collate third party tools that these actors already use into one platform; this will save considerable time and effort in aggregate.
- Continually search for ways to place the onus of verification on data submitted, rather than on the resources and reputation of the source.
    - Prompt information producers to submit metadata.
    - Spend considerable resources on designing education, aspects of platform access, and tools that are most accessible to under-empowered users.
- In addition to creating services for human rights defenders, we hope to create a platform with enough flexibility so that it can be put to use by a plethora of practitioners, from journalists to academics.
    - On a lower level, the platform will be flexible and versatile enough for implementing organizations in manipulating features to best fit their purposes – e.g., choosing/naming the types of input accepted, altering the user-facing interface, and creating forms of output that will be easily integrated, manipulated, and shared.
- Explore and mitigate the risks incumbent in human rights reporting online, particularly in the arenas of privacy and security. We intend to implement flexible approaches that maximize tradeoffs in closing the feedback loop. This could take the form of pseudonymous, anonymous, or identifying sources and submissions as the case may be.

8 The Whistle and NGOs

As Goh and Guay stated, "The rising influence of NGOs is one of the most significant developments in international affairs over the past 20 years. Although social movements have been part of the political and economic landscape for centuries, we trace the emergence of NGO activism in the USA during the modern era to mid-1984, when a range of NGOs, including church and community groups, human rights organizations, and other anti-apartheid activists, built strong networks and pressed US cities and states to divest their public pension funds of companies doing business in South Africa" (Doh & Guay, 2006).

Today, NGOs may play a vital role as the verifiers of the information provided by civilians. They are responsible for choosing the pieces of information that hold the greatest potential to advocate on behalf of endangered individuals and marginalised communities.. Obviously, NGOs are an important actor in civilian activism, coordinating bottom-up initiatives and ensuring that the voice of the weakest groups will be heard; some of them become actively involved in direct interventions in cases when human rights are violated.

Furthermore, in countries where the practice of whistleblowing has a longer tradition, NGOs may assist other actors of civil society in choosing proper methods and channels of disclosure, as well as attracting a substantial audience via media. They may provide legal advice and enable cooperation with other whistleblowing organizations.

The credibility of the digital information verification process is determined in part by the credibility of the actors engaged in the verification. The NGOs, particularly those already engaged in whistleblowing, are natural candidates for the role of verifiers, but there appears the classical dilemma of legitimization and control ("quis custodiet ipsos custodes?", or who watches the watchmen?). In other words, the NGOs need to understand their roles and responsibilities, and they need to engage in the process of self-education in relation to the specificity of the digital verification process. Also, they have to take care of their social legitimacy and transparency if they are to be a reliable partner in the digital information verification process.

Some of the NGOs participating in the process of digital information verification openly declare their method and attitude towards professional standards of verification. For example, Stratfor declares that "Our narratives are not derived from any canon, but materialize from careful examination of what could feasibly transpire rather than what someone says will occur".[5] Generally, there is a need to establish standards of verification of the verifiers.

9 Emerging technologies – Blockchain

The Whistle, while ultimately a human-centered verification platform, will depend on various technologies and links with other tools in order to provide users with accurate information, a secure analytical environment and means of disseminating findings. One of those emerging technologies that should be considered in the context of The Whistle, and the ChainReact project more broadly, is blockchain.

[5] https://www.stratfor.com/weekly/geopolitical-intelligence-political-journalism-and-wants-vs-needs

Blockchain is a class of distributed databases, which means that they do not depend on a single storage location; instead, numerous nodes of the blockchain network store copies of the database. The copies are being constantly updated on a peer-to-peer basis and the validity of a particular state of the database is established as a result of a consensus building throughout the entire network. Blockchain databases, created as a basis for emerging distributed currencies, are ledger-like. They record incremental events – transactions, check-ins etc. – ultimately storing the history of whatever aspect of reality is represented in the database.

Blockchains are created with security in mind. Their distributed nature and consensual updating mechanism make it extremely difficult for a single entity to tamper with the records and introduce misinformation to the database. In principle, the blockchains are supposed to be immutable, and although changing the history of the records is possible, it requires a great deal of consensus building both on technological and social levels.

The blockchain databases are transparent, which allows for various parties to examine their integrity. At the same time there are numerous cryptographic mechanisms involved in blockchain technology in order to provide a security layer both to the users of the databases and the assets (currency, information, contracts, etc.) they encode in it.

Most of the features of the blockchain technology are interesting from the point of view of digital information verification. The distributed and transparent nature of databases make them resistant to attempts at misinformation introduction. Heavy encryption provides security option for potential witnesses, whistleblowers and analysts, and the ledger-like nature of records make it easy to track the history and origin of information.

There already exist initiatives involving blockchain technologies which are potentially relevant in the context of ChainReact both in relation to the verification of reports and its specific application to the transparency of supply chains. Below, two such initiatives are reviewed in this context.

9.1 Blockchain and Privacy (Enigma Project)

MIT's project Enigma[6] is a distributed computational engine, built with the empowerment of data providers (namely, individuals deciding to share the data they own) in mind. It is supposed to allow for massive scale analytical computations while allowing people who contribute the data to preserve their anonymity and even profiting from sharing their personal information.

Enigma promises to combine the decentralisation and accuracy of information characterised by blockchains, with computational efficiency and privacy of an off-chain network. This way data analysis (aggregations, modelling etc.) can be performed without ever disclosing the raw data to anyone other than the original source. Enigma is supposed to work as a functional analog of "homomorphic encryption" which ensures that the results of computations on encrypted data is the same as computations performed on raw, open data.

Potentially, this platform could be applied to the field of digital information verification. The witnesses and whistleblowers could use an Enigma-like solution to submit and control their information without the fear of their identities ever being discovered. The validity of information, once data is submitted, is assured by the

6 http://enigma.media.mit.edu/

blockchain mechanisms. The limitations of this use case are related to the specificity of data which can be treated this way. Enigma seems to be well suited to processing well-defined, quantitative, (presumably) tabular data. In the context of whistleblowing activities this would correspond to, for example, leaking financial information by a group of informants, exposing various aspects of fraudulent activity. However, in case of witnesses dealing with less structured information, the use of Enigma-like solution seems to require at least some sort of preliminary information encoding on the informant's side, which makes the whole use scenario less compelling.

9.2 Blockchain and Supply Chains (Provenance Project)

The Provenance initiative directly engages with issues central to the ChainReact project; however, from a significantly different perspective. While The Whistle 'strut' of the ChainReact project wants to make supply chains transparent by collecting, verifying and aggregating information acquired from witnesses and insiders involved in supply processes, Provenance hopes to build transparency into supply chains themselves.

In order to achieve supply chain transparency, Provenance proposes to represent them in the blockchain database. The users of this solution would enter information (such as origin, maker, current owner, raw materials used, etc.) about a product (or service) on every step of product's lifetime. The data, stored in a blockchain database, would benefit from features characterising blockchain technology. The supply chain representation would be decentralised (resistant to tampering), the incremental nature of database would allow users to trace the history of the product and its components all the way to their origins – again, provided that the relevant information had been introduced into the system. This kind of representation would allow for the detailed scrutinising of certain aspects of supply chains and even whole supply networks.

An important aspect of Provenance project is that it requires that the parties controlling supply chains or their elements provide relevant information about it. There are at least two factors that might encourage these entities to do so. One motivation could be the public image of a company benefiting from the openness and transparency of procurement and production processes. On the other hand, companies might also benefit directly from detailed mapping their own supply flows, by identifying inefficiencies or fraudulent activities. However, entities which are aware of illicit activities within their supply chains might be not inclined to participate in initiatives such as Provenance.

The relation of the Provenance project and ChainReact should be viewed as complementary. While Provenance is concentrating on exposing product and material flows within supply chains, ChainReact concentrates on making supply processes more transparent by sharing and aggregating witnesses' stories and associating them with enterprises' profiles and supply chains mapped on the level of companies (as opposed to products). Certainly, there is potential for both initiatives to thrive and cooperate. The Provenance project may produce outputs that could be integrated in the corporate network mapping aspect of ChainReact, and/or used directly as source material by WikiRate. WikiRate should consider initiating a dialogue with Provenance to determine whether there is potential for collaboration on the task of mapping supply chains itself, or for WikiRate to push for the cooperation of companies on behalf of Provenance.

10 Conclusion

There are continual developments in the digital information verification space, as new initiatives arise, institutions catch onto the opportunities embedded in verification, and the flood of information shows no sign of calming. This is a quickly growing space, and as such, we view this report as a snapshot; no doubt, within the coming years and even months, many more initiatives will arise, more approaches will be explored.

Given the relative youth of the field, it is ripe for collaboration on a number of fronts. Different initiatives hold expertise in varying aspects of verification. For those initiatives with compatible aims, especially those centered on journalists, activists, and civilians, the main barrier to collaboration might be a lack of knowledge about the field. We hope our research might serve to break down such barriers.

Our ongoing concern with this field is that the disempowered become as empowered as possible. Great rises in information mean that selectivity is necessary, and the mechanisms for selectivity translate into unequal power distribution—actors with certain aims control the main channels of communication, and the choice of which information is disseminated on those channels is influenced by politics as well as pragmatics. As put by M. Hindman, there is a difference between speaking and being heard—and "on the Internet, the link between the two is weaker than it is in almost any other area of political life" (Hindman, 2009, p. 17).

11 References

Bourne, W. (2015, April). The Most Important Social Media Company You've Never Heard Of. *Inc. Magazine*. Retrieved from http://www.inc.com/magazine/201504/will-bourne/banjo-the-gods-eye-view.html

Burt, R. S. (2005). *Brokerage and Closure. An Introduction to Social Capital*. New York: Oxford University Press.

Czuperski, M., Herbst, J., Higgins, E., Palyakova, A., & Wilson, D. (2015). *Hiding in plain sight: Putin's war in Ukraine*. Washington, DC: The Atlantic Council of the United States. Retrieved from http://www.atlanticcouncil.org/publications/reports/hiding-in-plain-sight-putin-s-war-in-ukraine-and-boris-nemtsov-s-putin-war

Doh, J. P., & Guay, T. R. (2006). Corporate Social Responsibility, Public Policy, and NGO Activism in Europe and the United States: An Institutional-Stakeholder Perspective. *Journal of Management Studies*, *43*(1), 47–73. http://doi.org/10.1111/j.1467-6486.2006.00582.x

Hindman, M. S. (2009). *The myth of digital democracy*. Princeton: Princeton University Press.

McPherson, E. (2015a). Digital Human Rights Reporting by Civilian Witnesses: Surmounting the Verification Barrier. In *Produsing Theory in a Digital World 2.0: The Intersection of Audiences and Production in Contemporary Theory* (Vol. 2, pp.

193–209). Peter Lang. Retrieved from http://www.academia.edu/download/38827109/McPherson_Civilian_Witnessing.pdf

McPherson, E. (2015b). *ICTs and Human Rights Practice*. University of Cambridge Centre of Governance and Human Rights. Retrieved from https://www.repository.cam.ac.uk/handle/1810/251346

Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web* (pp. 677–686). ACM Press. http://doi.org/10.1145/2566486.2568012

Nunez, M. (2016, September 5). Former Facebook Workers: We Routinely Suppressed Conservative News. Retrieved July 4, 2016, from http://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006

Pariser, E. (2012). *The Filter Bubble: What The Internet Is Hiding From You*. New York: Penguin.

Silverman, C., & Tsubaki, R. (2014a). Creating a Verification Process and Checklist(s). In C. Silverman & R. Tsubaki, *Verification Handbook. A definitive guide to verifying digital content for emergency coverage.* (1st ed., p. 122). The Netherlands: European Journalism Centre. Retrieved from http://verificationhandbook.com/book/chapter9.php

Silverman, C., & Tsubaki, R. (2014b). *Verification Handbook. A definitive guide to verifying digital content for emergency coverage.* (1st ed.). The Netherlands: European Journalism Centre. Retrieved from http://verificationhandbook.com/book

Sindelar, D. (2014, August 12). The Kremlin's Troll Army. *The Atlantic*. Retrieved from http://www.theatlantic.com/international/archive/2014/08/the-kremlins-troll-army/375932/

Vis, F. (2014, April 24). To tackle the spread of misinformation online we must first understand it. *The Guardian*. Retrieved from http://www.theguardian.com/commentisfree/2014/apr/24/tackle-spread-misinformation-online

Appendix A – Project Profiles

A.1 Banjo

Banjo is a powerful social media monitoring tool that "instantly organizes the world's social and digital signals by location, giving an unprecedented level of understanding of what's happening anywhere in the world, in real time"[7]. As a proprietary platform that allows for aggregation social media into one platform, it holds great

[7] http://ban.jo

verification potential to "[allow] images and events to be cross-checked against each other", as noted in the European Journalism Centre's Verification Handbook (Silverman & Tsubaki, 2014, p. 210).

Banjo is best known to most consumers as a social media monitoring app, developed in 2011 and used by millions worldwide. However, the developers are now releasing Banjo Enterprise. This latest instantiation of the service is something more of an 'event-detection engine'. As described by Inc. Magazine,

"*Through a Playskool-simple Web interface, Banjo turns a system built around "following" people into one organized by location. It shows only geolocated public posts made from mobile devices; those posts are drawn from…a dozen major social networks (and counting), from Twitter to Instagram to Russia's VKontakte to China's Weibo…[Banjo] combines location, photo classification, analytics, and the ability to "rewind" each social media network in time.*" (Bourne, 2015)

The proprietary software developed by Banjo that monitors this 'world feed' is able to detect, given its memory of each 'space' on a virtual grid across the globe, when unusual activity is occurring in a particular location. Thus, some of the first companies to use Banjo were media outlets such as NBC and ESPN; the software allows for incredibly fast detection and real time monitoring of breaking stories such as the Boston Marathon bombings or a school shooting at Florida State University.

Banjo's software has potential implications for not only assisting first responders or monitoring ongoing news stories, but for personal privacy and surveillance. The software collects data from various social media platforms where the user has designated their content as public; however, many users on these platforms struggle with managing the privacy settings, and so some content might be analyzed without the consent of the creator, as is the case with much secondary data use. But on this front, Banjo has been careful to "[engineer] privacy protections into the product, including a patented method for routinely cycling through its databases to scrub any posts that have been pulled down or turned private by their authors" (Bourne, 2015).

A.2 BBC User-generated Content (UGC)

The BBC's UGC (user generated content) hub officially came into existence in the wake of the devastation of the 2004 Indian Ocean tsunami. The well-known news organization was inundated with emails: eyewitnesses sharing their accounts and families of victims trying to get the word out about lost loved ones. Since then, the UGC hub has grown from a pilot of three journalists to a full-fledged team[8] that finds, solicits, verifies, and publishes material directly from users. The verification practices refined within this organization hold potential for increased awareness and accountability surrounding human rights violations.

Background

The BBC's UGC Hub has become a "central clearing-house for content", as team members collaborate with colleagues in BBC Trending, BBC Monitoring, and general website, tv, and radio production. Initially, the team focused on reviewing unsolicited material, but now a good portion of their resources go toward finding content by scanning various social media platforms, as well as engaging with users via other platforms such as texting, email, and phone calls.

Description of Process

[8] http://www.bbc.com/news/world-30421631

Users upload material to a platform used by BBC journalists, who are able to view the content in a 'picture console' where they are then able to add verification notes and organize the content accordingly based on whether it has been verified or not. One of the best tools for this process is a response form at the bottom of online stories to encourage users to contact the BBC.

When it comes to discovering breaking news, the team uses social media to find more intelligence by scanning various platforms and connecting with eyewitnesses. Widely available tools that the team uses for this process include Tweetdeck[9] and Geofeedia[10].

After finding or receiving content, the team then conducts an authentication process, greatly informed by traditional journalistic practices, which includes:

- Closing the feedback loop by identifying and contacting the original source to ask basic questions
- Put the content through a series of technical checks (image manipulation or checking, cross-referencing with satellite imagery)

More Information

The BBC has published a list of FAQs related to its UGC practices[11]. Additionally, the December 2014 piece, published in BBC News[12], provides an overview of specific examples, illustrating how the news organization has used this team and its practices to refine its reporting. The organization is continuing to integrate the UGC Hub and its social media verification practices throughout its productions.

A.3 CameraV

CameraV[13] is "the easiest way to capture and share verifiable photos and video proof on a smartphone or tablet, all the while keeping it entirely secure and private". It assists users to generate metadata around their photos, allowing their images to be verified more easily for use in courts and through other accountability mechanisms.

CameraV is the official app from the InformaCam project[14], a collaboration between Witness[15], the International Bar Association[16], and Guardian Project[17].

In addition to the CameraV app, the project is also focusing on building "desktop software that can decrypt, decode, and visualize media generated by the app, and creating "a system [for] safely accepting and storing

[9] https://tweetdeck.twitter.com/
[10] https://geofeedia.com/
[11] http://www.bbc.co.uk/terms/faq.shtml#8
[12] https://www.journalism.co.uk/news/how-bbc-outside-source-is-forging-a-new-digital-style-of-live-video-news/s2/a565342/
[13] https://guardianproject.info/apps/camerav
[14] https://blog.witness.org/2013/01/how-informacam-improves-verification-of-mobile-media-files/
[15] https://witness.org/
[16] http://www.ibanet.org/
[17] https://guardianproject.info/

InformaCam media as submitted by app users"[18]. I.e., this system includes the ability to securely share material over Tor network, and the ability to share more than one kind of data: media, metadata in a plain txt file, or secure IDs and links to the media's location.

When it comes to verifying data received via CameraV, the project has also devoted resources to developing a testbed that can determine if the "media file or sensor data [was] tampered with since the time of capture".

The aim of this project is to assist human rights activists and journalists in more efficiently verifying the torrent of social media often created around crises and human rights violations.

A.4 Citizen Desk

Citizen Desk[19] is a platform used for aggregation, verification, and publishing of content created by citizens. It allows organizations to integrate community involvement, particularly citizen journalism, into their work. The code for the platform is open source, and developed/supported by SourceFabric[20], a Czech nonprofit organization.

The Citizen Desk platform has three basic functions:

- The platform collates information from various feeds and sources (e.g., social media as well as mobile phone texts).
- Organizations can send content through an editorial workflow, allowing for increased verification of submissions.
- Finally, the output can be sent to any of a variety of publishing platforms.

Other notable aspects of CitizenDesk include its focus on mobile technology (as the most useful way to communicate in areas with limited Internet connection) and the 'chat now' support functions and user forums, providing organizations with low technical resources assistance in incorporating the platform.

More information can be found at the Citizen Desk website, including a case study on Verdade[21], a widely read news source in Mozambique that used CitizenDesk to help monitor elections with the help of citizen journalists. SourceFabric can be followed on Twitter @Sourcefabric.

A.5 Amnesty International's Citizen Evidence Lab

The Citizen Evidence Lab[22] is an online toolbox for verification. The intention is to support human rights researchers and give them tools to best authenticate and verify text, sound and images. The website also includes a Youtube video analysing tool. The Lab is part of the Amnesty International's Sensor Project, and is founded by Christoph Koettl.

[18] https://dev.guardianproject.info/projects/informacam/wiki/System
[19] https://www.sourcefabric.org/en/citizendesk/
[20] https://www.sourcefabric.org/
[21] https://www.sourcefabric.org/en/home/casestudies/2380/
[22] http://citizenevidence.org/

A.6 Coal Scam

CoalScam.org[23] is an initiative spearheaded by three groups (mines, minerals & PEOPLE, Greenpeace India, and Amnesty International India) to "to map the true impact of coal mining in India on people's lives and habitats"[24]. For those in close proximity to coal mining, the process can have "devastating effects on land, water and air, affecting health and livelihoods and key human rights"; when officials and owners engage in corrupt practices, these effects are exacerbated. CoalScam.org is an issue-targeted effort to hold governments and companies to accountability in this arena.

Background

The main purpose of CoalScam is to collect, verify, and disseminate information from the ground, in order to create a better understanding of the real cost of coal mining. The project includes collection of media stories, ground reports and first-hand testimonies. Then, "reports on the site are uploaded after they are verified and moderated for veracity." The project members then attempt to contextualize each report, "seeking responses from companies and government officials as far as possible."

The CoalScam website is a submission platform, repository, and eventual site of publication for these reports. Its maintains a simple webform for report submission, easily navigated by those without extensive technological expertise and allowing for collaboration and contextualization of reports via linked media.

Coalscam also provides **resources** for educating interested and submitting parties. It also maintains an interactive visualization of its data, a map of reported events.

A.7 DataShift

DataShift[25] is an initiative by international NGO Civicus[26] aiming to "build the capacity and confidence of citizens and civil society to generate and use data"[27]. Along with partners the engine room and Wingu, Civicus has launched a two-year effort to ameliorate the amount, quality, and sources for data used in monitoring progress towards international development goals - particularly from local sources. This effort includes development and support of new technological mechanisms for citizen reporting.

For example, the initiative plans to release a DataShift Dashboard in 2016:

"[A] platform for sharing information about the coverage and comparability of citizen monitoring and attempt to promote accountability in development through the support and coordination of harmonisation efforts. Actual data will be made available through user-friendly methodologies and visual representations to provide alternative metrics according individual countries and goals".[28]

[23] http://coalscam.org
[24] http://coalscam.org/page/about
[25] http://civicus.org/thedatashift/
[26] http://civicus.org/index.php/en/
[27] https://www.devex.com/news/from-development-information-to-a-data-revolution-84625
[28] http://civicus.org/thedatashift/background/

This platform, along with other efforts of DataShift aimed at improving civilian engagement and data use, will provide new general resources for protecting human rights. More information on the DataShift can be found **at the official website** or by emailing datashift@civicus.org.

A.8 EchoSec

EchoSec[29] is a Canadian-based company that focuses on location-based search and social media monitoring. Its consumer based is varied; the company's products "[provide] public safety, security, journalism, and intelligence professionals actionable knowledge". For example, EchoSec recently was used to find evidence of Russian military personnel in Ukraine, as noted on the **citizen journalism website Bellingcat**[30].

More information can be found at the EchoSec website, and EchoSec can be followed on Twitter @echosec_search

A.9 Emergent

Emergent[31] is a project aiming to track real-time rumor development, run by Columbia's Tow Center for Digital Journalism[32]. It consists of a platform that allows users to track the propagation of misinformation via the media; the platform lists out various 'stories' and current status as to verification. Clicking on each story shows a breakdown of reporting sources, edits, and shares.

Emergent encourages submission of misinformation or rumors via email for tracking and verification. It is part of a larger project at the Tow Center that "draws on qualitative and quantitative data to test and analyze strategies and best practices for debunking misinformation". The project aims to eventually "provide actionable guidance for newsrooms on how to best debunk misinformation"[33]. Such guidance could be incorporated by human rights activists as well.

A.10 Eyewitness Media Hub

Eyewitness Media Hub (EMH) is "a non-profit organisation established to support the creation, discovery, verification and publication of eyewitness media."[34] It aims at developing and setting standards–ethical, legal, and logistical–from the standpoints of both creators and disseminators. In addition, EMH is also involved in the development of a verification platform for images (Verified Pixel[35]) and directing a coalition of interested media-related organizations around the use of eyewitness media.

Research and Dissemination

[29] https://www.echosec.net/
[30] https://www.bellingcat.com/news/uk-and-europe/2015/02/19/how-echosec-found-evidence-of-a-russian-fighting-in-ukraine/comment-page-1/
[31] http://www.emergent.info/about
[32] http://towcenter.org/
[33] http://www.craigsilverman.ca/2014/09/02/researching-rumors-and-debunking-for-the-tow-center-at-columbia-university/
[34] http://eyewitnessmediahub.com/
[35] http://www.verified-pixel.com/

EMH creates various resources (digital packages, broadcasts, workshops) on topics such as verification of content and protection of sources. It then disseminates these resources via live events, workshops, digital packages, and mentoring schemes. EMH research has focused on topics as varied as how the major global news outlets use eyewitness media[36] (and repercussions for the producers of that media) and the impact of 'vicarious trauma' in the media[37]. This research is both qualitative and quantitative, as they try to understand values and practices from the sides of users and news outlets. Much of their work in this sphere focuses on convening workshops and meetings for major news outlets to explain their findings and prompt discussion.

Coalition Development

Eyewitness Media Hub has received funding in the past from the Tow Center for Digital Journalism (based in Columbia), and in June 2015 received $2M for the development of First Draft Coalition - whose "aim is to open up the conversation around the use of eyewitness media in news reporting with a strong focus on ethics, verification, copyright and protection". This coalition is comprised of several well known outfits in social media and journalism: Bellingcat, Emergent, Meedan, Reported.ly,Storyful and Verification Junkie, and supported by Google News Lab.

Platform Development

Currently, EMH is also focusing on building a platform for Verified Pixel, a Knight Foundation-funded project that attempts to collate third party verification tools into one space to ameliorate the process of eyewitness media verification for journalists. They are developing this in conjunction with Source Fabric, a Czech-based software developer that creates open tools for journalism.

EMH's work is particularly relevant for the social media implications of human rights reporting. By developing platforms to verify user generated content, conducting research on user experience and values when it comes to use of eyewitness media, and creating spaces for discussion on these topics, EMH is contributing to better understanding and better tools in the reporting and verification space.

A.11 eyeWitness App

The following description was authored by Wendy Betts, Director of the eyeWitness App Project.

Information and communication technology has vastly enhanced our ability to report on and learn about the human rights situation in even the remotest parts of the globe. The internet is rife with examples of footage showing horrific abuses from conflict zones and other troubled regions. However, the internet is also increasingly full of apologetic posts from media outlets and others who further disseminated, what turned out to be, false footage. As a result, the media and ordinary information consumers are recognizing that seeing is not necessarily believing. This skepticism, while justified, allows the perpetrators of atrocities to claim that incriminating images are false and to rely on denial as a plausible, and even persuasive, line of defense.

[36] http://eyewitnessmediahub.com/research/user-generated-content/executive-summary
[37] https://medium.com/@emhub/making-secondary-trauma-a-primary-issue-24a802abe461#.a545c6yna

Nonetheless, human rights defenders, journalists, and ordinary citizens around the world continue to capture and upload videos showing the violent and oppressive actions of abusive regimes and groups. As problematic as the mistaken belief in falsified footage may be, disbelief of the authentic footage, collected by these courageous individuals at great risk, is tragic. Moreover, it is a lost opportunity to use this footage to help hold the perpetrators of international atrocity crimes accountable.

Background

eyeWitness to Atrocities[38], a project of the International Bar Association[39], aims to address the problem of verifying citizen captured video by providing a mobile app designed to record video and take photos in a manner that will facilitate authentication of the footage. The eyeWitness app is a tool that can elevate the authentic footage above the din of propaganda and misinformation by ensuring it can be believed.

Function

The eyeWitness app records and embeds metadata at the time the footage is captured that verifies where and when the footage was taken. The metadata also allows eyeWitness to confirm that the footage has not been edited or digitally altered and to trace the chain of custody. The user submits the videos or photos directly from the app to the eyeWitness organization. Footage sent to eyeWitness will be held in an off-line repository, hosted by LexisNexis. LexisNexis has created a secure cloud environment for the storage and management of data uploaded by eyeWitness users. The eyeWitness legal team will analyze relevant footage and seek out the appropriate legal authorities to investigate the situation further. In the interim, the secure repository will function as a virtual evidence locker, safeguarding the original, encrypted footage until it is needed for an investigation or trial.

Future Development

Putting in the hands of citizens on the ground the technology to capture verifiable video provides an opportunity for these individuals to help bring to justice the perpetrators of atrocities. However, while this information is necessary for accountability, it is not sufficient to bring it about. It is incumbent upon the international community to ensure there is an appropriate forum, whether national, regional, or international, to air this information and respond to these abuses.

A.12 First Draft

First Draft[40] aims to create a space for development of best practices in handling eyewitness media, which has substantial implications for verification of evidence in the human rights context. Supported by Google News

[38] http://www.ibanet.org/Article/Detail.aspx?ArticleUid=11e76b66-d949-4738-9347-e67fbfbb9441
[39] http://www.ibanet.org/Default.aspx
[40] https://medium.com/1st-draft

Lab and run by members of Eyewitness Media Hub[41], First Draft is currently working on a "new destination website that will feature essential training materials, plus a database of case studies, resources and tools".

The First Draft Coalition is currently comprised of several 'thought leaders and pioneers of social media journalism', including: Bellingcat, EyewitnessMediaHub, Emergent, Meedan, Reported.ly, Storyful and Verification Junkie. All of these members contribute via "regular articles, interviews and reviews covering all aspects of handling eyewitness media, including the most effective techniques for discovery and verification alongside ethical and legal guidance for publishing and broadcasting." For example, material already published includes a list of 'essential geolocation tools for verification'[42] by Eliot Higgins of Bellingcat. Such curation and collation of resources from First Draft will be of use to activists and journalists alike in verification practice.

A.13 The Gaza Platform

The Gaza Platform[43] is a comprehensive interactive map of Israeli attacks during the 2014 Gaza conflict. The project, realised by Amnesty International and Forensic Architecture, aims to serve as a tool to reveal "emerging patterns that could be used to pressure those with influence over the warring parties to take action". It achieves this by aggregating text reports, photos, videos, audio recordings and satellite imagery documenting the war, collected by the Al Mezan Center for Human Rights and the Palestinian Centre for Human Rights (PCHR), as well as Amnesty International.

The platform holds a database of over 2,650 individual event which can be browsed and filtered through a variety of variables including: over time, by type, by mode of firing by casualties, by media and by data source.

This is possible given the characteristics each incident is required to be described against (a full list of these can be found on their website under the heading "methodology").

The platform is based on pattrn[44] by Forensic Architecture (who are delivering their pilot project with the Gaza Platform) for aggregation capabilities, data visualization and interactive mapping.

A.14 Geofeedia

Geofeedia is a proprietary platform built for "location-based social media monitoring". The software aggregates user-generated content based on geotagging and public posting. This content can include that generated by phones as well as laptops, and is collated from the biggest incumbent social media platforms: YouTube, Instagram, Twitter, Flickr, Picasa, etc. Given its potential for monitoring and analytics based on

[41] http://ictandhr.tumblr.com/post/124907555483/eyewitness-media-hub
[42] https://medium.com/1st-draft/searching-the-earth-essential-geolocation-tools-for-verification-89d960bb8fba#.34vz1h9ma
[43] http://gazaplatform.amnesty.org/
[44] http://pattrn.co/#about

location, it is a tool used by journalists at the BBC Social Media Hub[45] and explored as a tool for first responders[46] in crises.

Though Geofeedia is a private company whose products are aimed at other companies wishing to monitor consumers' social media usage, it has been identified as a good tool in crisis management, public safety, and investigative journalism. As noted by multiple sources in 2012[47], during a public beta of the product, Geofeedia could potentially provide one of the most promising uses of social media in disasters. Additionally, as noted by Verification Junkie[48], Geofeedia "can assist in the verification process, by cross-referencing posts within a particular area to see if details match".

With its release of a new app in recent months, Geofeedia might prove to be even more useful to first responders and investigative journalists, given greater allowance to monitor "geo-tagged real-time social media data on the go". However, the aggregation and recording of these posts must be done with awareness of the privacy rights of the users, even when they post publicly.

A.15 Mexicoleaks

Mexicoleaks[49] is an "independent whistleblowing platform for revealing information in the public interest in Mexico". Its mission is to "amplify" the impact of the information revealed through their platform, on a social and legal level.

It operates by giving a secure submission link and protecting the identities of its whistleblowers. The information received is then verified, analysed and submitted to the group of partners of Mexicoleaks who will publish it through their channels, in order to build more capacity for NGOs, civil organisations and media outlets to take action in the interest of the public. In their own words: "Mexicoleaks is a tool that allows people to send information of public interest to media outlets and civil society organizations through secure technologies that protect the identity of the source".

Mexicoleaks currently partners with:

- Animal Politico
- emeequis
- Másde131
- Periodistas de a pie
- PODER
- PROCESO

[45] http://www.bbc.co.uk/academy/journalism/skills/social-media/article/art20140728140907290
[46] http://www.standbytaskforce.org/tag/geofeedia/
[47] http://www.standbytaskforce.org/tag/geofeedia/
[48] http://verificationjunkie.com/
[49] https://mexicoleaks.mx/english.html

- Red en Defensa de los Derechos Digitales (R3D)
- Aristegui Noticias

A.16 Pattrn

Pattrn[50] is a new project that aims to provide "data-driven, participatory fact mapping for research and analysis" to entities such as Amnesty International. Pattrn allows organizations to aggregate "diffused and partial bits of information" - for example, a text report here and an image there into an analytical and visual platform that gives them a birds-eye view of an event or situation.

Pattrn is run by Forensic Architecture, a research consultancy/project with ESRC funding based at Goldsmiths, University of London. They "undertake spatial and media analysis for the investigation of human rights violations", partnering on Pattrn with **Tekja** - a data visualization and analytics firm in London.

As further described on the Pattrn website, after the data has been collated, it can "then be explored through a visualisation platform, which simultaneously provides access to the details of each individual event and, through interactive graphs and charts, to the big picture of an overall situation…[T]he tool enables its users to build collaboratively a database of events with space and time coordinates, and to add tags, media, and content to these events."

The first Pattrn project is the Gaza Platform by Amnesty International, which is a comprehensive interactive map of Israeli attacks during the 2014 Gaza conflict, based on data collected by Amnesty and local NGOs over the period of months. For journalists and activists, Pattrn can be used to map crises, protests, and conflicts over a space of time, providing an aggregation tool for verification and accountability.

A.17 People's Intelligence

People's Intelligence (PI) is a crowd-sourcing media verification project, based in the Netherlands, and currently in the development stage. It aims to help "victims and witnesses better document and verify their stories and provides them as well as relevant organizations with actionable information, thereby facilitating early warning and targeted assistance…[i]t supports syntactic and semantic analysis and allows networking between affected individuals, communities, relevant organizations and experts through the use of ubiquitous technologies".

One of the aims of the project is widespread, low cost and easy-use tools for submission and verification, on both the sides of the citizen and NGOs. The envisioned tool will also be appropriate for use in a variety of contexts. For example, "PI's technology is highly portable and can be applied in multiple domains and settings, including but not limited to citizen journalism, election monitoring, environment degradation, corruption and good governance, labour rights and businesses' accountability."

Current partners include Amnesty International, Free Press, Liberia Peacebuilding Office, TU Delft, and the Peace Informatics Lab at Leiden University in the Netherlands. Organizations supporting the project include USAID, The Humanitarian Innovation Fund, and ELRHA.

[50] http://pattrn.co/#about

A.18 PHEME

PHEME[51] is a European Commission-funded project aimed at evaluating the truthfulness of social media data.

There is a real problem in that it's not currently possible to carry out the complex task of analysing mass-data in real time, and on most occasions, each entry would have to be verified through a manual process. This poses a problem to organizations to whom it would be helpful to have access to verified social media content to build e.g. medical information systems and for digital journalism.

Given that 2 of the 3 Vs (Gartner) have been dealt with more extensively (volume and velocity), PHEME is concerned with mainly veracity as one of the largest computational challenges.

PHEME uses a combination of big data analytics, advanced linguistics and visual methods to produce powerful algorithms to process social media information. The platform employs the following categorization of data:

- **Speculation**
- **Controversy**
- **Misinformation**
- **Disinformation**

PHEME is a particularly interesting project as it could potentially save NGOs a lot of time when it comes to processing information, particularly with reference to human rights violations. The projects builds on the GATE text mining platform[52] (Sheffield), cross-media linking and contradiction detection, as well as Ushahidi's swift-river[53] media filtering and verification platform.

PHEME plans on releasing the algorithms as open-source upon completion.

A.19 REVEAL

REVEAL[54] works to advance the necessary technologies to facilitate a higher level analysis of social media content. In their own words, "The project will enable users to reveal hidden 'modalities' such as reputation, influence or credibility of information".

REVEAL acknowledges that media organisations can no longer act as "gatekeepers", deciding what information is communicated to whom. Instead, individuals have the power to access information directly from the primary source, and to share it through social media. However, this content is usually bare, and REVEAL aims to "reveal" more, by determining how trustworthy the information is, the impact potential of the contributor of the information and to what extent it all affects reputation or influence.

[51] https://www.pheme.eu/
[52] https://gate.ac.uk/
[53] https://wiki.ushahidi.com/display/WIKI/SwiftRiver
[54] http://revealproject.eu/about-reveal/

It works through the following stages:

- Analyze
    - "Automatic inference of community-type relations between members by continuous analysis of their social interactions" based on Social Media graph analysis, privacy preserving data analytics, crowdsourcing through gamification and psychology-based sentiment analysis.
- Organize
    - Focused analysis on large volume of data based on computational stylometry along with forensic image and video analysis
- Deduce
    - Understanding the context of social media to determine relevance, through tracking the content back to its origin
- Portray
    - "Effective presentation layer that will help users understand and further utilise such higher level concepts" based on crowdsourcing.
- Protect
    - Protecting personal data.

A.20 Storyful

Storyful is a private venture that provides verification of content distributed on social media. Combining journalistic expertise with technological tools, Storyful has a dedicated team that 'sources, dates, and geolocates'[55] various types of content that are being distributed online.

Often this content is from eyewitnesses of events where human rights are possibly being infringed upon. When such content pops up on popular platforms such as YouTube and Facebook, Storyful verifies these postings for newsrooms and producers.

In this way, Storyful is a means of connecting citizens with the traditional media, and has additionally made attempts to do so in a collaborative and transparent way. This was the impetus behind the Open Newsroom project[56], an experimental space with contributions from well-respected civil society organizations and human rights activists, curated and assisted by the Storyful team. Contributors have included Eliot Higgins and James Miller, NGO experts like Peter Bouckaert (Human Rights Watch) and Christopher Koettl (Amnesty International).

[55] http://blog.storyful.com/
[56] http://blog.storyful.com/2013/06/27/experimenting-with-open-journalism/#.VT-l9WSrS2w

As a joint project with Witness[57], Storyful powers the youtube Human Rights Channel.

Inclusion/Exclusion:

Owned by one of the largest media companies in the world (NewsCorp), Storyful's main customers are other news organizations. As an intermediary, it is able to assist traditional media in verifying sources. But as an intermediary, it is also reliant on its partner civil society organizations bringing relevant content to its attention, and its own resources in scanning various social media platforms. It is not the main disseminator of what it verifies; it is not an organization that has the final say on what makes the news and what does not.

A.21 Taarifa: The Real Life Bug Tracker

Taarifa[58] was established as a mechanism for people to stay engaged with their local governments, on matters such as infrastructure. Taarifa is an open-source web application, initially based out of Tanzania, where it was utilised to manage waste water and to allow for citizens to report sightings of related problems. Ultimately, the goal of Taarifa is to avoid the damages and hazardous consequences of children's exposure to waste materials in particular.

Submissions are sent to appropriate governments or NGOs, whom are notified of the issue, effectively establishing a communication channel between the public and decision makers. Although Taarifa allows input for detailed meta-data, it does not elaborate on whether or to what extend information is verified.

Having started at the WaterHackathon in 2011 and won the the Global Sanitation Hackathon, Taarifa partnered with Geeks Without Bounds (GWOB[59]), an accelerator helping humanitarian startups develop, to bring their platform to Tanzania. They were aided by GWOB in setting up a hackathon with the purpose of getting more developers involved to run the project.

The current workflow of the project allows for **information collection**, **visualisation,** and **interactive mapping**. Data is collected via **sms, webforms, email** or **twitter,** making the application easily accessibly for all levels of tech proficiency.

A.22 TweetCred

TweetCred is a "a real-time, web-based system to assess credibility of content on Twitter". It currently exists as a Chrome extension[60] and browser version[61]. Tweetcred assigns a credibility score to individual tweets, based algorithmically on 45 different factors, including "content, properties of user who posted the tweet, external URLs or pictures shared in the tweet".

[57] https://witness.org/
[58] https://github.com/taarifa/TaarifaAPI
[59] http://gwob.org/
[60] https://chrome.google.com/webstore/detail/tweetcred/fbokljinlogeihdnkikeeneiankdgikg
[61] http://twitdigest.iiitd.edu.in/TweetCred/tweetcred-browser-demo.php

TweetCred was developed via "insights that researchers have gained from studying massive databases of tweets surrounding major news events", as profiled in the New Yorker[62]. The algorithmically determined score is being improved by user feedback; since machine learning is at the core of the app's functionality, it will only improve as use grows.

The TweetCred score, as "an indicator of the trustworthiness of information…[and] credibility of resources (pictures, videos, URLs)", can serve as a tool to help journalists and activists verify reports. Additionally, if widely used, it might act as a deterrent when individual users are considering retweeting potentially false information.

A.23 Ushahidi

Ethnic violence and unrest following the 2007 Kenyan elections were the results of an unexpected election fraud, which also catalysed the role of the citizen journalist. Originally designed to allow journalists to report on incidences of violence across the country, Ushahidi[63] saw 45,000 users utilize the platform extensively as a means of communicating eye witness reports with appropriate authorities, via mapping incidents geographically with accompanying picture, time and location.

Ushahidi takes its point of departure in handling large crowd-generated data, which is otherwise missed or hidden in the vastness of information.

It works by collecting information through user input on various platforms including: email, text messaging, Twitter and RSS feeds; visualizing it, and mapping it. Its back-end facilitates the process by categorising the information and geolocating it, and finally publishing it to an interactive map.

Being an open-source web application, Ushahidi has most popularly served as the core for context-specific applications by other non-profits and NGOs, including the Louisiana Bucket Brigade, who developed their iWitness platform on top of Ushahidi to allow its citizens to report damages and other incidents caused by flooding. Interestingly, in the case of Louisiana, the Brigade wanted to allow the citizens employed by BP, whom are otherwise contractually bound to non-disclosure, to report on cases where oil that accompanies the flooding had caused significant damage in their homes or elsewhere, without fear of losing their jobs. This adds an element of privacy/anonymity (it is unclear to what extent it protected the identity of the user), which could be useful when transposed to circumstances under for instance authoritarian regimes, where citizens might want to report human rights abuses.

Most recently, Ushahidi deployed the Uchaguzi[64] initiative to monitor the first Kenyan elections since the new constitution, in 2013. Uchaguzi was used as a platform for rapid reporting and alerting system, in addition and cooperation with the more traditional monitoring done by the Elections Observer Group (ELOG). Vast crowd-generated information undergoes a verification process to determine how credible the information is, and the verified information is finally delivered to organisations who can intervene. As a final step, it monitors the response received, in order to measure its effectiveness.

[62] http://www.newyorker.com/tech/elements/can-an-algorithm-solve-twitters-credibility-problem
[63] https://www.ushahidi.com/
[64] https://wiki.ushahidi.com/display/WIKI/Uchaguzi%20-%20Kenyan%20Elections%202013

A.24 Verified Pixel Project

Developed by Source Fabric and Eyewitness Media Hub, Verified Pixel is a project that aims to "build a service that will unify several popular existing image verification tests into an automated workflow."[65] By creating this platform, the developers hope to "make image verification accessible to a wider audience".

The Verified Pixel Project is currently prototyping with funding until 2016 via the Knight Foundation. Specifically, they are proposing to "integrate several popular existing verification tests in one workflow tool using original code and layered API calls." Such an approach, aimed at a tool for individual journalists, will hopefully cut down on the amount of time spent on individual photo verification and facilitate faster, more efficient use of images in the various contexts of reporting.

Current partners to the project include the Knight Foundation, Tin Eye[66], and Izitru[67].

A.25 Veri.ly

Veri.ly is "an experimental web application designed to rapidly crowdsource the verification of information during humanitarian disasters"[68]. Inspired by the 2009 DARPA Network Challenge, which asked teams to locate 10 red weather balloon distributed across the United States, Veri.ly is built by members of the successful MIT team. Their winning strategy in the DARPA challenge was to harness the power of public participation via social media.

As described by Victor Naroditskiy, one of the application's creators, the subsequent "rationale for Verily is that the collective effort of people searching for the truth will be fruitful…[t]he lack of verifiability of content posted on social media is the main reason preventing humanitarian and news organisations from making a wider use of it."[69]

According to the Veril.y website[70], the platform has already been piloted at major news organizations, and it already offers means for humanitarian organizations to request a verification procedure. It also includes several educational resources for the average citizen who wishes to assist in the verification process, such as a list of suggested tools.

A.26 Wikimapia

Wikimapia is described as a "multilingual open-content collaborative map"[71]. Alternatively, it could be described as a crowd-sourced version of Google maps, complete with human-written descriptions. The project is not associated with the better-known Wikimedia (the hosting organization of Wikipedia), but aims

[65] http://www.verified-pixel.com/what-is-vpp/
[66] http://www.tineye.com/
[67] https://www.izitru.com/
[68] https://veri.ly/about
[69] http://phys.org/news/2014-07-wikipedia-fact-checking-natural-disasters.html
[70] https://veri.ly/
[71] http://wikimapia.org/about/

to provide the average user with the means to document and explore local space. Wikimapia is a tool that can be used by journalists and activists to verify the location of a purported event. For example, the Verification Handbook documents how Wikimapia was used to pinpoint the location of a demonstration in Egypt[72].

A.27 Witness App

Witness[73] is a "panic button" app for individuals in high-risk circumstances, and as a largely new feature on the market, it also tracks incidences in real-time. It was built to answer the question: "if I find myself in an emergency and all I have is my phone, what is the best my phone can do to protect me?"

Users can start streaming and recording a precarious situation as it unfolds, and have the app send it to pre-defined contacts who will be able to receive a broadcast of location, audio and video. The app requires only one touch to activate, and is suitable to any smartphone user.

Having been compared to a private version of Twitter's Periscope, the app was originally developed for people in situations such as walking home alone at night, and it resonates well with recent incidents of police brutality. An example used was the event of the officer's brutality against teens at a Texas pool party[74]; on this occasion the incident was recorded on a video and then uploaded to social media. Witness would be the medium which would have allowed the incident to be reported immediately and as it unfolded.

Witness automatically calls the pre-defined contact and texts them the location of the incident, and provides a link to a live-stream herein. In addition to broadcasting the incident, the app uses the phone's internal storage in situations where it is unable to communicate via cellular or data networks.

Witness won the TechCrunch Disrupt NY 2015 Hackathon, and has recently unveiled its public release.

A.28 Yomapic

Yomapic[75] is a free tool that visualizes geographic searches for geotagged images, pulling content from Instagram and Russian social media site VKontakte[76]. Similar to other location-based search tools such as EchoSec, Yomapic has been cited as a useful tool by citizen journalists in investigation and verification.

For example, as recently noted by Eliot Higgins of the citizen journalist initiative Bellingcat,

"*An interesting feature of Yomapic is the possibility to select an individual account by clicking on the username, displaying all images on the account, geotagged or not. The below image shows on Instagram user as they travel from the Seychelles to Kenya, then onto Turkey and finally Raqqa, Syria.*

While the obvious use for Yomapic…is looking for posts related to specific events in an area, it's also useful in another way. When geolocating images it might not be possible to find street view imagery or Panoramio photographs, but thanks to Yomapic and EchoSec it's possible to see if anyone in the local area may have posted an Instagram photo or YouTube video showing the

[72] http://verificationhandbook.com/book/chapter5.php
[73] https://getwitness.com/
[74] https://www.theguardian.com/us-news/2015/jun/08/texas-pool-party-police-dajerria-becton-eric-casebolt-rude
[75] http://www.yomapic.com/
[76] https://new.vk.com/

location you're interested in. This adds to the range of option you have when searching for reference images as part of any geolocation, and by combining these tools it makes efforts to verify content far more effective."[77]

[77] https://www.bellingcat.com/resources/how-tos/2015/07/25/searching-the-earth-essential-geolocation-tools-for-verification/