|By Rebekah Larsen

The Digital Information Verification Field

A snapshot of current actors, practices, and trends.

27 October 2015

whistle@hermes.cam.ac.uk | www.thewhistle.org | @WhistleReporter

Notes on Research

Research contributions were made to this report from the following members of The Whistle team:

- Matt Mahmoudi – many thanks for his help in brainstorming the categories, research into individual initiatives, and intrepid, creative development of the visualizations.
- Dr Ella McPherson – many thanks for her insight into the field, endless enthusiasm, and expert editing.

Much of the data in these graphs was assigned based on limited information given project resource constraints. As a result, the below findings and those in the online visualizations are a result of our own interpretations, and any misrepresentation or mistake is ours. We welcome suggestions, corrections, updates, and additions at whistle@hermes.cam.ac.uk.

Acknowledgements

As a team, we would like to thank those individuals and organizations that generously gave of their valuable time and knowledge during our research. We interviewed several representatives of non-profit, academic, and commercial initiatives in the digital information verification space, gaining not only specific information regarding their operations but also valuable insight into the field's development. Many also provided us with further contacts and resources. In return, we hope that this report and the accompanying data will serve to illuminate the current verification space—and perhaps be of use in further innovation and collaboration.

Introduction

This report aims to map the field of initiatives involved in the verification of digital information. The impetus for this research was the ongoing development of The Whistle, a web application that aims to connect civilian witnesses who are digitally reporting human rights violations with human rights organisations that can help them pursue accountability. Based on the premise that verification can be a bottleneck for these organisations' use of civilian witness information as evidence, The Whistle aims to speed verification practices with an eye to increasing the variety and volume of civilian witness voices in human rights reporting.

This report is the result of The Whistle team's review of the current digital information verification field. 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.

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.[1] 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.[2] 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.[3]

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.[4] 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.

[1] McPherson, E. (2015). *Digital Human Rights Reporting by Civilian Witnesses: Surmounting the Verification Barrier. In R. A. Lind (Ed.), Produsing Theory in a Digital World 2.0: The Intersection of Audiences and Production in Contemporary Theory (Vol. 2, pp. 193–209).* New York, NY: Peter Lang Publishing.

[2] Czuperski, M., Herbst, J., Higgins, E., Polyakova, A. and Wilson, D. (2015) Hiding in Plain Sight: Putin's War in Ukraine. Think Tank. Washington, DC: Atlantic Council, p. 40. http://www.atlanticcouncil.org/publications/reports/hiding-in-plain-sight-putin-s-war-in-ukraine-and-boris-nemtsov-s-putin-war

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

[4] McPherson, E. (2015) 'ICTs and Human Rights Practice', *Centre of Governance and Human Rights, University of Cambridge*. Available at: https://www.repository.cam.ac.uk/handle/1810/251346

The Digital information verification Field

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.[5], "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"[6]

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.

[5] 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.

[6] Silverman, C. and Tsubaki, R. (2014) 'Creating a Verification Process and Checklist(s)', in *Verification Handbook.* 1st edn. The Netherlands: European Journalism Centre, p. 122.

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 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. Accompanying this report is a visualization of these findings on the Whistle website that gives a detailed overview of the categorizations below and provides information on each individual initiative studied. This visualization takes the form of a concentric circle, with each layer representing a category and branching subcategories. Each category is clickable, providing explanations and links.

METHODOLOGY

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. (Our current map predominantly consists of English-language projects, but we aim to expand on this going forward and welcome suggestions of projects in other languages for inclusion via email at whistle@hermes.cam.ac.uk.)

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.

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.

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.

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).

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.

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.

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).

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.

Each of these categories has been subdivided into more explicit aspects of practice and characterization. More information on the categories, including subcategories and the initiatives that fall into each, can be found in the interactive visualizations on The Whistle's website.

After categorizing each initiative, we compiled the resulting data into charts, weighing initiatives based on their maturity. We also compared each initiative in marimekko charts based on which characteristics, practices, or focuses they presented. These can be access by click on the categories in the online visualization. The 'marimekko 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 Input – Data mekko char, one can see that social media is the most frequent type of input, with 24 initiatives utilizing this data type (given the largest column width); the 'other' category in contrast has only three initiatives and the column width is correspondingly small. These categorizations were not exclusive—some initiatives fall into multiple categories in each chart.

Field and Trends

Below are select highlights and takeaways from each of the categories. The online visualization also includes links to more information on each of the individual initiatives. Given that this is a selection, The Whistle team plans to post periodic additional insights from the information collected on the project website.

While the information below is representative of a comprehensive review, we have tried to pay particular attention to two 'metrics' by which to evaluate practices in particular: pluralism (the inclusion and empowerment of as many voices as possible) and speed of processing.

Characteristics

In addition to reviewing the verification practices of each initiative, we also looked at the characteristics of each: the level of development, sources of funding, and type of organization engaging in these practices.

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.

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.

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.

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 (the EU's research and innovation program).

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 EU'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.

Practices

Every one of the initiatives reviewed engages in some type of practice meant to facilitate digital information verification (see the descriptions of the categories above). These categories are not exclusive; some initiatives fall into multiple.

Input

This category includes those initiatives that accept information for evaluation; it excludes those initiatives whose main purpose is to provide a space for discussion and coalition building.

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.

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.

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).

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.

Process

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.

Human v. 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.

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.

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.

OUTPUT

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.

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.

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.

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).

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.

EDUCATION

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.[7] 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.

Focus

- 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.

[7] McPherson, E. (2015). *Digital Human Rights Reporting by Civilian Witnesses: Surmounting the Verification Barrier. In R. A. Lind (Ed.), Produsing Theory in a Digital World 2.0: The Intersection of Audiences and Production in Contemporary Theory (Vol. 2, pp. 193–209).* New York, NY: Peter Lang Publishing.

- 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 project);
    - Intelligence professionals (including law enforcement), investigators embedded in think tanks, and agencies that are increasingly using such tools in open source intelligence.[8]

Nature

- 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).

[8] Czuperski, M., Herbst, J., Higgins, E., Polyakova, A. and Wilson, D. (2015) *Hiding in Plain Sight: Putin's War in Ukraine*. Think Tank. Washington, DC: Atlantic Council, p. 40. Available at: http://www.atlanticcouncil.org/publications/reports/hiding-in-plain-sight-putin-s-war-in-ukraine-and-boris-nemtsov-s-putin-war

- 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.

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.

THE WHISTLE: THE NEXT STEPS

> *"Through speeding and simplifying the verification process, The Whistle aims to support the variety and volume of civilian witnesses' participation in human rights reporting."*

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.
- Illuminate the field (as with this report) so that initiatives can combine forces and resources.

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 young age 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, as a project with a focus on pluralism and human rights, 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".[9]

[9] Hindman, M. (2009) *The Myth of Digital Democracy*. Princeton University Press. (17)

One of The Whistle team's aims in increasing pluralism in this space is to continue trying to illuminate it. Stay tuned! We will be publishing a series of blog posts over the coming months, focusing on topics such as security and privacy, closing the feedback loop, human-centred verification, and highlighting specific practices. We welcome any insight, additions, clarifications, and general feedback at whistle@hermes.cam.ac.uk.