**Myanmar Witness: Summary of Methodology**

*September 2021*

1. **Introduction**

Myanmar Witness is a collaboration between international OSINT experts and Burmese journalists and civil society representatives, managed by the Centre for Information Resilience. The project’s goal is to increase accountability for human rights abuses and violence in Myanmar. The project uses open-source investigation techniques to identify, document and verify instances of alleged human rights abuses and violence. Its methodology is sensitive to gender and conflict dynamics and builds the capacity of local actors in parallel with conducting investigations. The project disseminates data and investigations in a form that can be used by:

* International legal accountability mechanisms, specifically the Independent Investigative Mechanism for Myanmar (IIMM), to support their investigations into violations of international law;
* Governments and international bodies to inform strategy and action to respond to human rights abuses and violence in Myanmar, and;
* Civil society and the media to raise awareness and increase pressure on governments to hold actors in Myanmar to account.

Myanmar Witness commenced operations in April 2021 and publicly launched in June 2021. Further details are available on the Myanmar Witness microsite[,](https://www.info-res.org/myanmarwitness) [Twitter](https://twitter.com/MyanmarWitness) and [Facebook](https://www.facebook.com/MyanmarWitness2021) pages.

1. **Methodology**

Myanmar Witness follows an eight-step process to collect, preserve, catalogue, verify, analyse, review, investigate and report on data relating to acts of violence and interference with human rights in Myanmar. This process aligns with best practice as set out in the Berkeley Protocol on Digital Open Source Investigations. A visual representation of our data processing framework is set out in Annex 1.

**Step 1: Data Collection**

 Data content is collected from three key sources:

* **User-generated content on open-source social media,** with relevant data identified through keyword, date-based and hashtag searches and monitoring of online sources which generate high levels of relevant content
* **Anonymous submissions made by individuals** through the project’s online submissions form
* **Third-party content** shared by journalists, civil society organisations, public bodies and international organisations

We constantly review our data sources to maintain objectivity (for example, balancing open-source material collected from both pro-democracy movement and pro-Tatmadaw social media channels) and to address data gaps e.g., in relation to specific geographical areas.

**Step 2: Preservation**

Each piece of content is assigned a unique identifier, logged in the Myanmar Witness database, downloaded and saved in a secure drive (the archive) in its original form. In addition to file downloads, we save screenshots of social media posts to ensure they are archived and available even if subsequently removed from platforms. This original data is then SHA3-512 hashed and publicly timestamped (via Twitter) in a transparent process which provides assurance it has not been interfered with. The archive is regularly backed up onto a secure external hard drive.

**Step 3: Cataloguing**

Where it exists, metadata is extracted from all preserved media and logged in the Myanmar Witness database, alongside data type, source, reported date and location and violence level (which triggers additional traumatic imagery handling protocols).

**Step 4: Verification**

Wherever possible, we verify **where** and **when** a piece of content was taken through:

1. *Geolocation*: Using visual forensics ([here](https://www.youtube.com/playlist?list=PLrFPX1Vfqk3ehZKSFeb9pVIHqxqrNW8Sy) for MW Lead Investigator Ben Strick's guide to these techniques) to establish the exact coordinates (latitude and longitude) of an image. These are recorded in the database, along with a brief description of how that geolocation was established.
2. *Chronolocation:* Using similar techniques, satellite imagery, alternative angles of footage and shadow analysis to establish the date and, where possible, time of a piece of footage.

No content is marked as verified unless it has been geolocated to specific coordinates. However, it is not always possible to chronolocate to equivalent levels of precision (i.e., a specified timeframe). Content will be marked as verified where:

1. Reporting from multiple sources concurs on the data published
2. All accessible evidence available suggests that the footage was taken on the date indicated - i.e., it is consistent with satellite and other imagery and reporting taken at the time;
3. There is no evidence to suggest it appears earlier than the recorded incident date, for example a reverse image search would show no earlier results.
4. Weather conditions and other visual clues align with the recorded incident date.

The data entry will be marked making it clear that the date and time has not been confirmed.

We assess whether the geolocation and chronolocation results align with the metadata where available and the reported location and time of incident. We note any limitations or reservations with geolocation and chronolocation results in the database.

**Step 5: Analysis**

This stage focuses on establishing **what** each piece of content shows. Each piece of content is assigned a privacy rating (which triggers additional data protection protocols); incident category (ies) and a variety of analysis tags pertaining to its content. The presence of different actors is recorded (along with a confidence level), as well as use of arms and ammunition and a gender count (see Section 6 of this paper for detail). A textual description of the content is also recorded.

**Step 6: Quality Control of Analysis**

Every piece of analysed content is reviewed by a second team member with expert reviewer status, who confirms they are satisfied the analysis has been done correctly and is accurate, and helps resolve any queries or issues.

**Step 7: Archive and Data Management**

Reviewed data is then transferred onto a master spreadsheet, which forms the basis of project reporting, ensuring only verified and reviewed data is used in reports and communications. This is backed up regularly to the external hard-drive for preservation and security purposes.

**Step 8: Investigation and Reporting**

This stage focuses on:

* Identifying and analysing trends in the data e.g., relating to specific types of incidents and patterns of behaviour
* Identifying and analysing specific incidents or series of events in the data

We use a WHO, WHAT, WHERE, WHEN framework of investigation, which aims to establish a timeline and geographical map of incidents; who was involved and in what capacity and; what took place. To do this, we compare data we hold in our database with high-quality media outlets and other credible third party (NGO, UN) reporting and eyewitness testimony. We also use specialist tools and databases - such as NASA FIRMS Fire databases - to corroborate findings.

1. **Data Quality and Limitations**

The strength of Myanmar Witness data comes from both the verification of individual pieces of content to international best practice standards and the subsequent triangulation of those pieces of content with each other, and with third party reports. By geo-locating and chrono-locating multiple pieces of content relating to the same incident, provided by different sources and taken at different angles and times, and comparing these against eyewitness testimony, media reports, historical footage and satellite imagery, we are able to build up a robust chronological and spatial analysis of incidents on the ground. To further verify our findings, we cross-check key pieces of information such as images of visible insignia or weapon types against both our own datasets (we are developing the first weaponry index for the current conflict in Myanmar) and third-party sources.

However, there are limitations to the representativeness and quality of the data we are able to collect, which inevitably limits the scope of our verification techniques. The combination of internet restrictions and unequal access to the internet and social media based on location, class and gender means that we have less content from rural and more remote areas of the country, and that content is harder to verify. As described in Section 3 of this paper, localised and time-limited communication shutdowns also create data gaps in our analysis of specific incidents.

Some forms of content are more challenging to verify than others. It is difficult to geolocate or chronolocate audio and text content, poor quality content, footage taken at night or in poor visibility and footage taken with no identifying landmarks, - although it is possible to make some judgements on veracity based on triangulation with other sources. These challenges are amplified by climatic and weather conditions in Myanmar, where pervasive cloud cover in some areas can limit the usefulness of satellite imagery (although this is now being partially addressed through access to higher quality satellite imagery). This means that we are less likely to be able to verify certain types of incidents, or incidents taken in certain locations.

As described in Section 3, it is often not possible to be as precise with chronolocation as with geolocation.

While we triangulate content and reporting to gain as full a picture as possible of both incidents and the contexts in which they occur, it is not always possible to establish the context and chain of events surrounding a particular incident. This is particularly the case where we receive isolated pieces of content, which do not relate to reported events. Where these feature in project reporting, we clearly state what is known and not known in relation to these pieces of content.

1. **Gender Analysis**

We conduct a gender analysis on all content logged in our database. Our goal is to identify trends in the data in relation to gender (for example if women or men are more likely to be affected by certain types of incidents) and to identify and address potential biases in our data. For example, we recognise that women are less likely to have access to an internet-enabled device; less likely to post or feature in content on social media and; more likely to be affected by less visible forms of violence.

The most meaningful form of gender disaggregation for Myanmar Witness is a gender breakdown of those who are subject to possible interference with their human rights. This is highly challenging because: i) many pieces of content do not contain clearly identifiable subjects of human right interferences; ii) where content does feature data subjects of human rights interferences, their gender is not always clear and; iii) we frequently have multiple pieces of content relating to the same incident, creating a risk of double-counting.

For each piece of content, we assess on a case-by-case basis using set criteria whether it is possible to identify individuals who are subject to a possible interference with their human rights and then count and log these as male, female, other or unknown. Where there are multiple identical or highly similar pieces of content relating to an incident, we only record gender for the first piece of content. Where there are different and separate pieces of content relating to the same incident, we record genders for each piece of content. This is because it is often not clear at the outset which pieces of content relate to precisely which incident (thus trying to define by incident would create data flow issues) and because of the resource requirements involved in trying to triangulate gender counts across multiple pieces of content for multiple incidents. Where possible, we will identify an overall gender estimate for prominent incidents which are the subject of detailed investigations. We also record qualitative notes on gender disaggregation and gender issues.

Given the challenges described, it is important to emphasise that the gender disaggregation does not represent an accurate gender breakdown of the number of subjects of human rights interferences documented by the project. Instead, it should be seen as indicative analysis that is able to signpost trends and characteristics in the incidents taking place on the ground. Gender trends in the first phase of project data will be covered in a forthcoming trend analysis of Myanmar Witness data.

In addition to gender disaggregation, we also use a data coding system to log possible cases of sexual and gender-based violence (SGBV). Possible cases of SGBV are then copied to a dedicated datasheet for analysis. Incidents of possible SGBV are discussed in detail in the relevant thematic and incident reports. However, from a methodological perspective it is important to note that - to date - content relating to possible incidents of SGBV has largely been impossible to verify. This is because they either contain reports of SGBV with no supporting imagery or footage to analyse, or imagery that is not possible to geo or chrono-locate. It may be possible to address this in later phases of the project through cooperation agreements with specialist agencies working on SGBV issues.

**Annex 1: Data Processing Framework**

