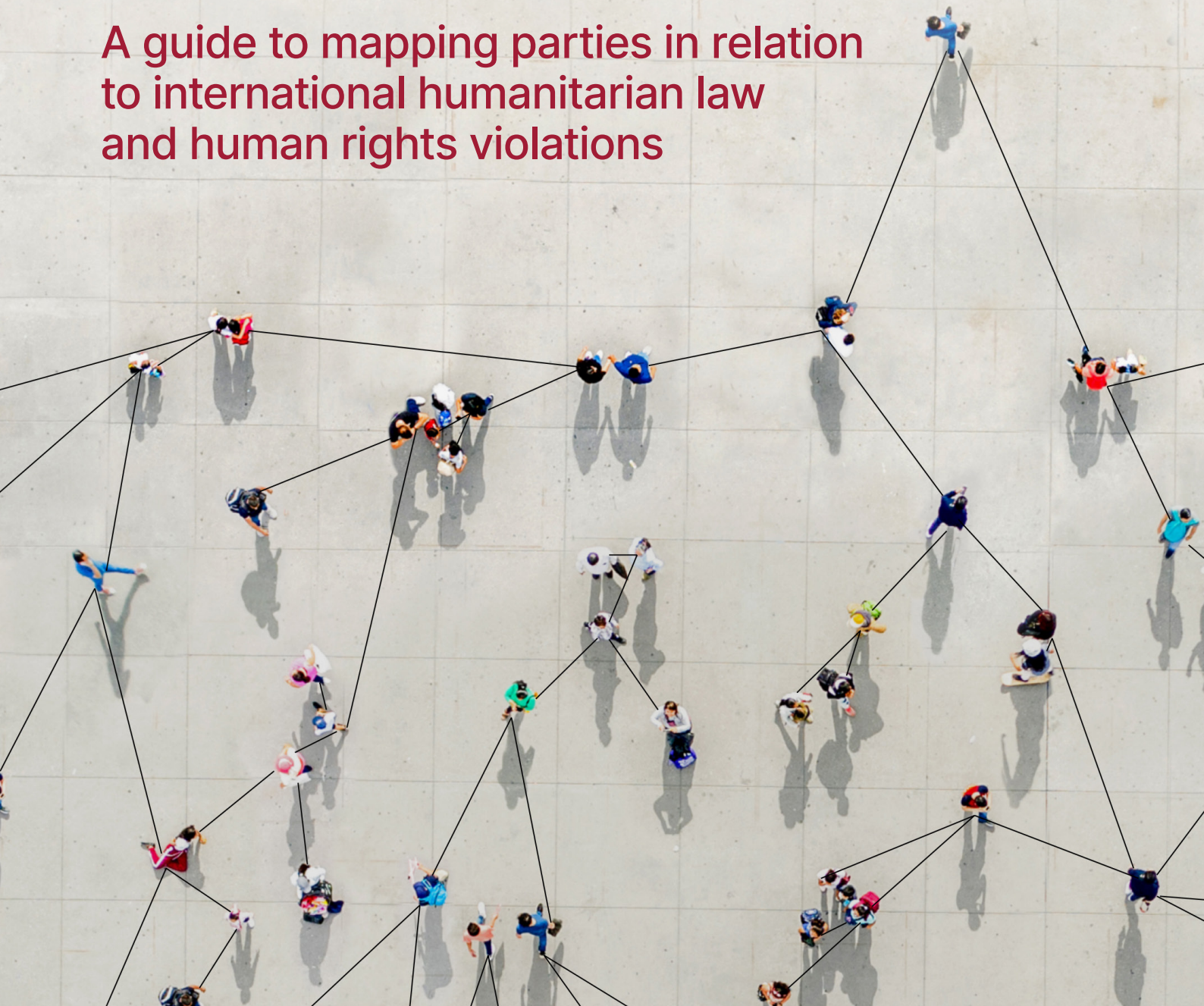
Public Interest Advocacy Centre

HUMAN RIGHTS CENTER
UC Berkeley School of Law

# Investigating Perpetrators

## A guide to mapping parties in relation to international humanitarian law and human rights violations



Public Interest Advocacy Centre

The Public Interest Advocacy Centre (PIAC) is a leading social justice law and policy centre. We are an independent, non-profit organisation that works with people and communities who are marginalised and facing disadvantage.

The Truth and Accountability Program at PIAC advances truth-telling and accountability for large-scale, systemic violations of human rights and international humanitarian law.

**Address:** Level 5/175 Liverpool St, Sydney NSW 2000, Australia

**Telephone:** +61 2 8898 6500

**Email:** media@piac.asn.au

**Website:** https://piac.asn.au/

**Social media:** @PIACnews

HUMAN RIGHTS CENTER
UC Berkeley School of Law

The Human Rights Center at the University of California, Berkeley, School of Law conducts research on war crimes and other serious violations of international humanitarian law and human rights. Using evidence- based research methods and innovative technologies, we support efforts to hold perpetrators accountable and to protect vulnerable populations. We also train students and advocates to research, investigate, and document human rights violations and turn this information into effective action.

**Address:** 2224 Piedmont Avenue, Berkeley, CA 94720

**Telephone:** +1 510 642 0965

**Email:** hrc@berkeley.edu

**Website:** humanrights.berkeley.edu

**Social media:** @hrcberkeley

## Authors:

Daniela Gavshon
Mary Flanagan
Nadeshda Jayakody
Erol Gorur

# Contents

# Contents (continued)

# Commonly Used Terms

| Term | Definition |
|---|---|
| **Armed group** | An organised armed group distinct from the armed forces of a State or paramilitary. |
| **Open-source materials** | Publicly available information that any member of the public can observe, purchase or request without requiring special legal status or unauthorised access. |
| **Open-source research/ investigation** | Use of open-source information for information and evidence gathering functions. |
| **Paramilitary** | An organisation whose structure may be similar to that of a military but operates separately to a State's security force or sometimes alongside or under the control of a State's security force. |
| **Parties** | Includes units and individuals within state security forces, other armed groups or paramilitary groups associated with a conflict or mass human rights violations, which may be perpetrators. |
| **Security force** | The armed forces, police and other law enforcement agencies of a State, including intelligence services. |

# Abbreviations and Acronyms

| | |
|---|---|
| **GUI** | Graphical user interface |
| **IHL** | International humanitarian law |
| **IHRL** | International human rights law |
| **INGO** | International non-governmental organisation |
| **NGO** | Local non-governmental organisation |
| **ORBAT** | Order of battle |
| **OSINT** | Open-source intelligence |
| **SF** | State security forces |

# 1. Introduction

Establishing a clear understanding of the parties to a conflict or those involved in specific events/incidents is a crucial preliminary step in any investigation into potential perpetrators of international humanitarian law (IHL) and international human rights law (IHRL) violations.

There is a vast amount of disparate open-source information available in the public domain regarding parties to conflicts or specific events. The challenge is knowing how to find the information and how to meaningfully piece it together. When this information is systematically collected, organised and analysed, it is possible to create a rich and powerful information map of the structure and location of parties through time. This mapping can be used as a dynamic analytical tool to understand parties and to investigate their potential involvement as perpetrators of IHL and IHRL violations.

Mapping parties is a complicated and resource intensive undertaking. It requires careful thought on a range of matters including scope, project management, security, technology, and methodology. There is very little guidance available for those seeking to conduct this work. The purpose of this guide is to support people and organisations seeking to map parties to a conflict or parties involved in specific events. It does so by providing practical guidance and a proven research methodology for the process of assembling the disparate information regarding parties. This allows for a coherent understanding of the structure and location of parties, as well as their potential involvement in IHL and IHRL violations.

The discipline of open-source research in IHL and IHRL investigations has been progressing at a record pace over the last decade. We are seeing increased professionalisation and sophistication. This progress is to be welcomed but there is still some way to go. This guide is intended to contribute to the growing professionalism of open-source intelligence (OSINT) work and to serve as a useful tool for OSINT practitioners working on IHL or IHRL investigations.

The authors would like to acknowledge the productive partnership between the Public Interest Advocacy Centre (PIAC) in Sydney, Australia and the Human Rights Center (HRC) at UC Berkeley, California. In drafting this guide, PIAC has drawn on its prior experience conducting mapping projects. HRC has drawn on its expertise as a leader in the professionalisation of the OSINT field. The HRC Co-Executive Director, Alexa Koenig, has served as an advisory committee member on some of PIAC's mapping projects, and the HRC team has trained PIAC staff on OSINT methods and the Berkeley Protocol on Digital Open Source Investigations. PIAC and HRC would like to particularly thank John Ralston for his guidance and expert input into this work and Michelle Ke for her assistance in preparing an earlier draft of this report.

## 1.1 What does this guide cover?

This Guide covers the process for collecting, organising and analysing open-source information to map parties and their alleged involvement in incidents that may constitute IHL and/or IHRL violations. Parties include the Security Forces (SFs) and other armed groups. This Guide can be used to map parties only, or it can be used to go further and map parties as potential perpetrators of incidents. The work can be referred to as parties mapping or, where relevant, perpetrator mapping.

This Guide focuses on the use of open-source research to conduct parties mapping. The Guide does not examine how to map IHL or IHRL violations. There are other guides that provide this information.[1] This

1. See for example: 'How to prepare for a database', *HURIDOCS* (Web Page) <https://huridocs.org/resource-library/monitoring-and-documenting-human-rights-violations/how-to-prepare-for-a-database/>; Steven Spittaels and Filip Hilgert, International Peace Information Service, *Handbook: Mapping Conflict Motives in War Areas* (August 2008) <https://www.ipisresearch.be/maps/Handbook_Aug2008.pdf>; See also Public Interest Advocacy Centre, *Conflict Mapping and Archive Project Methodology* (2020) <https://ipisresearch.be/mapping/webmapping/lka/context_documents/CMAP_methodology.pdf>.

Guide also does not examine how to conduct witness interviews to map parties and potential perpetrators.[2]

This Guide is divided into six sections:

- Section 2: issues to consider when deciding the scope of the parties mapping work;
- Section 3: practical considerations, such as staffing and security;
- Section 4: issues to consider in the technical design and development of the parties mapping work;
- Section 5: the research and analysis phases; and
- Section 6: issues to consider when using or sharing the analysis.

## 1.2 Why use this guide?

The purpose of parties mapping is two-fold:

- to build an information base about parties of interest, including information on command positions and structures and organisational hierarchy; and
- to map individuals' and units' alleged involvement in incidents that may amount to IHL and/or IHRL violations.

Mapping parties, and in particular building detailed profiles of key units and individuals, is useful for understanding how and where parties operate(d); their command structures; the career histories of individuals of interest; and, where relevant, the current roles being held by individuals of interest.

Open-source mapping can also provide valuable insight into which periods or geographical areas were more widely or accurately reported, and those periods or areas where more targeted investigations are necessary to address information gaps.

Parties and perpetrator mapping can support focused investigative work for accountability processes, including but not limited to criminal prosecutions, civil cases, sanctions, and various forms of vetting.

## 1.3 Who is this guide for?

The intended audience for this guide includes international non-government organisations (INGOs) and non-government organisations (NGOs), multilateral organisations, and others working to map parties that may be implicated in IHRL and/or IHL violations.

2. See other guides for how to interview witnesses, including remotely, for example, Public Interest Advocacy Centre, *Restricted Access Interviews: A Guide to Interviewing Witnesses in Remote Human Rights Investigations* (April 2021) <https://piac.asn.au/wp-content/uploads/2021/08/Restriced-Access-Intervew-Guide.pdf>; United Nations, Office of the High Commissioner for Human Rights, 'Interviewing' in *Manual of Human Rights Monitoring* (HR/P/PT/7/Rev. 1, 2019) Ch 11, <https://www.ohchr.org/Documents/Publications/Chapter11-MHRM.pdf>,

# 2. Scope of Mapping

## 2.1 Purpose of mapping

The purpose of mapping work and how the parties mapping will be used must be considered at the outset. This includes how it will add value to the existing body of work in a given context. The purpose will determine the scope of the mapping exercise including, in particular:

- the parties to map;
- the incidents to focus on in situations where perpetrator analysis is mapped;
- the time period to cover;
- the geographical areas to consider; and
- the scope of analysis and whether and to what extent conclusions and evaluations will be made.

**Tip:** Do not be too ambitious – be strategic about your scope and then expand it if you have the resources and time. Mapping parties can take a lot longer than initially anticipated.

## 2.2 Mapping as an investigative lead

Mapping provides a nuanced and informed starting point for further investigations. The nature of the mapping work as an investigative lead should be recognised and made clear to intended end users of the mapping, particularly if the mapping includes perpetrator analysis. The following factors should be disclosed:

- **the nature of the source material on which the mapping information is based**, for example, if the information is based solely on open-source materials, this may limit the analysis;
- **the range and breadth of source material used**, for example, the extent to which sources considered reliable or less reliable for particular information are used;
- **the scope of the mapping exercise**, for example, the extent of the parties, incidents, time period, geographical areas that are mapped.

As all mapping work will have unique parameters, the methodology followed should be clearly documented and, where possible, made available to intended end users.

In all cases, because mapping work is a lead for other work, it is helpful for end users if the following is provided:

- the ability for end users to access the original underlying source for each piece of information presented;
- reliability assessments of the sources used based on agreed criteria; and
- confidence assessments of the information presented.

## 2.3 Parties

Parties to map may include SFs, paramilitary groups and other armed groups, such as non-state actors, and units and individuals within them. The range of parties to be mapped will be guided by the purpose of the mapping work. For instance, if the mapping work is solely to support vetting of SFs, such as United Nations peacekeeper vetting, then only the SFs and their potential connection to incidents would need to be mapped. Alternatively, if the mapping work is intended to assist with wider accountability efforts such as criminal prosecutions, or visa vetting processes, a broader number of parties might be mapped, that is, not solely members of the SFs.

Mapping parties can include mapping:

- command structure and organisational hierarchy, for example, superior and subordinate units, connections between units;
- individual commander and other important positions, as well as connections between individuals;
- the location of key parties during the time period covered;
- possible connections between units/individuals and incidents; and
- other relevant information about parties, including weaponry, modus operandi, unit origins etc.

## 2.4 Incidents

Where perpetrator analysis is included in the mapping, a list of “incidents to consider” must be created. It is

important to set parameters around the types of incidents that will be relevant to the perpetrator mapping aspect of the work. Incidents should be prioritised according to agreed selection criteria. Such criteria may include factors like:

- the extent of open-source reporting on the incident;
- the gravity of the incident;
- the range of potential perpetrators;
- the geographical location of the incident; or
- whether the incident forms part of a pattern of violations of particular interest.

If incidents have already been mapped in a separate exercise, this will help define the scope of the parties mapping. For example, it may be apparent from incident mapping:

- what geographic or temporal scope to cover;
- whether certain party units should be mapped first because these units were known to be located near where the incidents occurred; or
- allegations that implicate certain parties in incidents or patterns of incidents.

If incidents have not been mapped prior to parties mapping, initial research work[3] will still need to be undertaken to identify the types of incidents or events that will be relevant. This is the case even if perpetrator analysis is not being done as incident mapping will assist the team to define the scope of the parties mapping.

**Tip:** Priority incidents should be identified from the start. This will assist in limiting the scope and will help ensure any mapping work remains relevant to the most pertinent incidents.

## 2.5 Time period

Mapping requires a significant amount of time and resources, particularly if the aim is to map parties relatively comprehensively. Given the size of the task, depending on the context, parameters should be set around the time period that will be mapped. When choosing the relevant time period, the following factors should be considered:

- **the availability of open-source material**: for example, for a conflict or events that have spanned many decades, if the mapping is solely based on publicly available information, there will be less open-source material available from earlier time periods. It may, therefore, be preferred to exclude or limit those time periods if they are unlikely to yield much information;
- **key phases of a conflict or events**: for example, there may be periods of a protracted conflict or events that are more violent than others or which resulted in particular violations of interest. In such cases, it may be preferred to focus on the time periods with a greater number of violations or which resulted in particular types of violations of interest; and
- **the age of potential perpetrators**: for example, for a protracted conflict the mapping work may only be relevant for a particular, more recent, phase of the conflict as some accountability efforts cannot be pursued if the key perpetrators are deceased.

**Tip:** Be strategic about the time period being covered – both in terms of the likelihood of finding available information and the likelihood of the information being usable.

## 2.6 Geographical scope

Factors to consider when defining the geographical scope include the following:

- **active conflict zones**: for example, some geographical regions may not have formed part of the active conflict zone and may therefore be excluded from the mapping exercise;
- **sites of particular incidents or incident clusters**: for example, some geographical areas may be the site of particular incidents of interest or incident clusters which may require specific focus, such as detention sites; and
- **known locations of units and people of interest**: for example, units and people of interest may be known to have moved through, or to have been stationed in certain geographical areas. These locations may form the basis for a particular geographic focus.

3. There are several guides detailing how to map incidents. See for example: 'How to prepare for a database', *HURIDOCS* (Web Page) <https://huridocs.org/resource-library/monitoring-and-documenting-human-rights-violations/how-to-prepare-for-a-database/>; Steven Spittaels and Filip Hilgert, International Peace Information Service, *Handbook: Mapping Conflict Motives in War Areas* (August 2008) <https://www.ipisresearch.be/maps/Handbook_Aug2008.pdf>. See also Public Interest Advocacy Centre, *Conflict Mapping and Archive Project Methodology* (2020) <https://ipisresearch.be/mapping/webmapping/lka/context_documents/CMAP_methodology.pdf>.

# 3. Practical Considerations

## 3.1 Staffing

When assembling a team for mapping work, a team should comprise people with varied experience and expertise. This can include:

- a team leader;
- legal officers/investigators/analysts/researchers (senior and junior);[4]
- technology officer(s), ideally embedded within the team; and
- consultants who can provide specific expertise as needed (e.g., military expertise).

The following skills are necessary for parties mapping:

- **project management**: ability to plan, initiate, execute, monitor, control and close the body of work;
- **legal knowledge**: an understanding of IHRL, IHL and international criminal law;
- **contextual knowledge**: a sound knowledge of the conflict (if applicable), country context and language skills;
- **technical skills for information systems and database design**: ability to develop and refine information systems to gather and preserve information and process it into an appropriate data structure;
- **technical skills for open-source intelligence (OSINT) tools**: ability to implement sophisticated technological tools to support OSINT;
- **OSINT skills**: ability to conduct rigorous open-source research using the latest investigative techniques having regard to the context being mapped;
- **military expertise**: specialised military consultants as needed from time to time on matters such as structure, operations, weaponry etc.; and
- **strong research and analysis capabilities**: a large research team with strong research and analysis skills across the range of potential types of sources that will be accessed.

**Tip:** Junior level researchers are crucial for mapping work. To this end, pro bono partnerships with law firms, and partnerships with universities where students can undertake paid internships or research work for course credit can be helpful for resourcing at a junior level.

**Tip:** An embedded technology officer is essential to ensure the technical elements support the work of researchers and end users. Funding proposals should include this type of role.

## 3.2 Security

It is necessary to conduct a detailed security analysis at the outset and establish security procedures and supporting documentation to identify the threats and mitigate the risk of those threats occurring during the course of the work. This is especially pertinent given that parties mapping work is often carried out in contexts where certain actors, such as the governments involved, are hostile to the work. Key documentation includes:

- **threat assessment matrix and risk mitigation strategy**: a document which identifies the types of security threats and sets out strategies to mitigate the risk of those threats occurring; and
- **security accountability framework**: a document which outlines the roles, responsibilities and accountabilities of the team.

A security management team can be established to identify threats, mitigate risks and help maintain appropriate security. The responsibilities of the security management team can include the following:

- conducting regular security threat assessments and updating the documentation regarding threat assessments and risk mitigation strategies as security threats evolve;

4. Referred to as "researchers" throughout this document.

- engaging specialised experts as needed, especially relating to cybersecurity, as the knowledge required to accurately assess and mitigate cybersecurity threats is often not available in-house;
- creating and implementing procedures to ensure that where required, the team complies with the measures set out in security documentation;
- providing training as required to all members of the team on their roles, responsibilities and accountabilities under security documentation; and
- establishing procedures for responding to emergency incidents and managing responses if such an incident occurs.

**Tip:** Taking a serious approach to security involves engaging a variety of experts including for in-country knowledge; and hardware and software knowledge.

**Tip:** Find a balance between having rigorous security protocols to effectively manage risks and avoiding overly onerous security obligations that make the work too difficult to carry out.

**Tip:** Clear communication about security protocols is required to maintain security across a large team who may be working remotely. Documentation must be simple and easy to follow. The importance of security must be emphasised during induction processes.

## 3.3 Inductions and training

If the mapping is part of an ongoing body of work, continuous training and information sharing will ensure the methodology is applied consistently across the team. This is especially required if a team is large, researchers are working remotely and/or working on a part time basis and if researchers are likely to change over time. Strategies to maintain efficient information sharing across the team include:

- thorough induction processes;
- regular team meetings for all team members;
- regular communications regarding the latest methodology decisions, including by building this into the information system; and
- dedicated working sessions where team members come together either in person or virtually to conduct research.

**Tip:** Subject to security considerations, it will often be more time efficient to have an initial induction recorded (and updated as necessary) if there is relatively high turnover of junior researchers. This will also ensure training is consistent.

## 3.4 Managing vicarious trauma

Researchers undertaking parties mapping work and in particular, perpetrator analysis, are likely to be exposed to distressing content. Precautions to minimise the harmful effects of the work, such as vicarious trauma, should be taken.

Practical measures can include:

- watching video without audio where the audio is not relevant for the analysis;
- stopping video auto-playing;
- not showing thumbnails of videos;
- requiring users to flag or grade content (for themselves and others);
- having a balance of tasks; and
- regular de-briefing.

Team leads will need to ensure researchers are provided with information on best practice for minimising the harmful effects of this work, as well as structuring work and teams in a way that facilitates this.

Researchers should be trained in identifying signs and symptoms of vicarious trauma in themselves and others and how to manage the risks. There should be organisational support provided to individual researchers if necessary. Funding proposals should include resources for managing vicarious trauma.

# 4. Technical Design

This section introduces issues to consider in the design and development of the technical systems that underpin the parties mapping work. Annexure A provides more detailed guidance, outlining suggestions for an implementation of the data model in a relational database and a graphical user interface (GUI) for researchers.

## 4.1 Information system

It is useful to view the mapping work as being undertaken within an information system. This information system supports the phases of research and analysis detailed in Section 5, and should be developed in line with the scope and purposes of the mapping work. Processes and elements of the information system to consider include:

- **Finding relevant sources**: how will researchers find sources and information, whether online or in hard copy, and how will this process be managed so that it is systematic and efficient?
- **Storing and preserving content**: how should gathered content be stored (if at all), what requirements are necessary for its preservation for different purposes, and what are the risks and implications of this?
- **Database and data model**: how will the data be organised to meet the requirements and purposes of the work?
- **Designing interfaces for research and analysis**: how will researchers use interfaces, including workflows to enter information, process data and conduct analysis?
- **Outputs and end uses**: how will the database and workflows facilitate end uses, whether by direct access to the database, by export, or by producing other content based on the data (e.g., written products)?

## 4.2 Designing a data model

The data model should be developed by the technology officer(s) alongside the mapping methodology. It should proceed with a clear understanding of the requirements, purposes and end uses of the mapping work. It may be useful to have external input from specialists or other organisations with relevant experience, especially at the design stage. Some factors include:

- **Flexibility**: while a clear and rigorous structure is useful from the outset, as with the methodology, the model should be flexible and re-developed as necessary throughout the life of the mapping work. It is essential to plan for ongoing review of the data model, especially as the team gains more knowledge of the conflict or incident(s), the parties, and the kinds of uses the mapping work will facilitate.
- **Complexity**: teams should consider the extent to which the aims of the mapping work are facilitated by a data model that tries to accurately represent a complex real-world environment (a conflict or patterns of violations involving many actors over an extended period of time). While a data model could be infinitely complex, the extent to which this detracts from its usability, intuitiveness and accessibility by researchers and end users should be considered. This, again, will depend on the purpose of the mapping work.
- **Sources**: as discussed in Section 5.3, the research and analysis will generally begin with documents and other content. The data model should consider how this content will be stored and accessed, as well as other implications such as security and continuity, for example, what will happen to the content after the end of the mapping work? Importantly, the data model should allow for all data to be able to be supported (or linked to) one or more sources.
- **Querying/searching data**: the data model should account for how researchers and end users will query the data, including through keyword searches (e.g., for names of individuals, units, locations), searches for particular incidents, spatial queries, temporal queries, and searches for documents or other content.

**Tip:** Technology officer(s), when embedded within the team, will be more able to proactively identify ways to improve the data model to better facilitate the goals of the mapping work.

## 4.3 Choosing a database

Many different kinds of databases could satisfy the requirements of the mapping exercise, ranging from proprietary software to open-source tools, and different types including relational or graph databases. In deciding on the appropriate choice of database, teams should consider the following:

- Purpose and scope of the mapping work, including what kinds of information are within the scope and likely to be raised by the key sources (see Section 5.1 on conducting a digital landscape assessment), and what outputs or end products are required.
- End use and continuity questions, such as who will require access to the data, in what form, and for how long, and whether integration with other external databases or information systems will be required.
- Ease of integration of the database into the information system that underpins the mapping exercise.
- Adaptability and the extent to which the data model or design can be refined in accordance with the needs of the mapping work.
- Costs, including development, ongoing updates and changes, maintenance, licenses, and continuity beyond the life of the mapping work.
- Accessibility by researchers, including minimum hardware requirements, remote access and internet access issues.
- Security requirements and risk assessments.
- GUI design requirements for data entry and analysis by researchers, and, if appropriate, end users.

## 4.4 Designing interfaces for researchers

Due to the range and complexity of the data that the mapping work will collect, designing how researchers will interact with the database is important and should be given early consideration. The work of researchers should be facilitated by interfaces such as dashboards and workflows that align with the phases of research and analysis discussed in Section 5. These interfaces should be considered important elements of the information system.

User interface and user experience design are broad fields of study and commentary with varied insights that do not require repeating here. However, in designing interfaces for researchers undertaking mapping work, there are some pertinent considerations:

- Ensuring that interfaces only show relevant information for the task being conducted. This involves striking a balance between presenting enough information for researchers to make decisions, without overwhelming them with information. The GUI can become difficult to work in given the amount of information that is presented, which is largely text-based, and will often involve repetitive tasks.
- Minimising the impact of distressing content to manage vicarious trauma (i.e. putting considerations in Section 3.4 into the user design). Some suggestions include: requiring researchers to flag or grade content (both for themselves and other researchers); automatically hiding flagged content; not showing thumbnails for videos, playing videos without sound as the default, and not auto-playing videos; requiring breaks after certain periods of work in certain workflows, among others.
- Designing features for communication within the research team, including messaging, reviewing, and updating researchers. Review by senior members of the research team is important for training new researchers, but also review by team members at the same level is valuable for ensuring consistency of data and analysis. These features should be easily accessible within workflows and display on dashboards.
- Designing features for researchers to make notes to themselves, which can encourage reflection, allow for researchers to take breaks and continue where they left off, and improve decision-making.
- Designing features to address security risks, for example, automatically logging out after a period of inactivity, and limiting access to or visibility of information that poses a security risk.

**Tip:** A close working relationship between researchers and technology officer(s) is essential to ensure the GUI best facilitates the research and analysis, and that researchers are aware of useful features in the GUI. The technology officer(s) should have routine meetings with researchers to understand how to refine or change the interfaces where necessary, and to update and train researchers on new features as they are developed. It may also be useful for the technology officer(s) to routinely work in the interfaces themselves to identify potential improvements.

### 4.4.1 End users

A final, critical consideration is how the data model and researcher interfaces will facilitate the end uses of the mapping work. This follows directly from the purposes described in the methodology. Key questions are who will require access to the data, in what form, and for how long, and whether integration into other databases or information systems is required.

If building interfaces for the end user, some considerations include:

- Creating interfaces that allow multiple access points to the data, for example, units and individuals (showing detailed profiles), incidents (showing incident details and alleged perpetrators), and documents (showing data points that came from the document).
- For party profiles, demonstrating the breadth and depth of information, without overwhelming the end user. One example is reducing the amount of text fields immediately visible, and instead having these visible only when clicked or hovered over.
- Highlighting the most important information, which will typically be perpetrator analysis.
- Demonstrating how many sources there are for a data point and confidence in the data point, visually and intuitively.
- Embedding source documents, as end users may want to view the sources themselves to follow the analysis provided.
- Providing easily accessible guidance for end users, such as a guide to key functions, walkthroughs, and more detailed documentation on the methodology, data model and data dictionary.

**Tip:** Feedback from end users on their requirements and preferred features is essential. This should happen throughout the mapping work, for example through beta testing end user layouts, and will also help to inform the ongoing development of the research methodology.

# 5. Research and Analysis Phases

This section provides an overview of best practice for how to collect, research, analyse and document the relevant information used in mapping work. The research and analysis phases include:

- Phase one: **digital landscape assessment** – conducting an overview of potentially relevant sources to determine which ones to use initially. The list of sources will be amended during the course of the mapping.
- Phase two: **scoping incidents and parties** – researching incidents (whether already mapped or not) to determine the focus and geographical scope of the parties mapping, and units and individuals of interest in relation to these incidents.
- Phase three: **systematic review of sources** – thoroughly reviewing sources looking for key information on parties, such as a parties' place in the hierarchy, commanders of units, and the location of units and individuals. In particular, any units or individuals identified in phase two should be focused on.
- Phase four: **review, revise, research** – consolidating information gathered in phase three to develop detailed profiles of units and individuals of interest.
- Phase five: **perpetrator analysis** – investigating key incidents and using the research and analysis from phases two to four to identify connections of interest between the incidents and units/individuals.
- Phase six: **final review** – filling gaps in the research to develop detailed profiles of units and individuals of interest.

It is necessary to maintain a confined and realistic timeframe for research. A one to two week pilot should be conducted for each research and analysis phase. This will help to determine a realistic timeframe for each phase as well as support any early adjustments needed in the methodology.

The methodology development should be an evolving process. Regular team meetings are necessary to review the mapping methodology and the research processes. This will ensure team resources are maximised and research processes remain efficient. Ideally, a technology officer should attend these meetings to ensure the methodology continues to be supported by the data structure and data input processes as it develops.

All decisions should be documented and shared with researchers and induction videos modified if necessary.

**Tip:** Continually question the approach to research and analysis throughout each phase to ensure it is efficient and yielding the intended results. Revise the approach as necessary.

## 5.1 Phase one: digital landscape assessment

### 5.1.1 Conducting the digital landscape assessment

For mapping work that relies on online open-source material, a digital landscape assessment should first be conducted. A digital landscape assessment requires identifying key online sources that contain relevant information on parties.

The types of sources to consider include:

- Official websites of parties – a good starting point is the official websites of, for example, the army, navy, air force, police, ministry of defence etc. These sources can be particularly helpful for unit hierarchy and commander information. Where possible, researchers should search for organograms and order of battle (ORBAT) diagrams. If the websites provide contemporaneous reporting during a conflict or a period of violations, they can also be useful sources of party location information, particularly if, for example, troop movement maps are published.
- Local and international media sources – this includes information from the UN, INGOs, NGOs, blogs and information that has been leaked and is now publicly available etc.
- Biographies and memoirs – writings by or about key SF or armed group personnel active during a conflict may be useful depending on the conflict context. This can be helpful for information on particular campaigns and battles and the units and commanders involved.

- Social media – research should be done to identify the social media platforms and digital technologies (e.g., mobile phones) commonly used in the geographic region being investigated, including by the parties themselves. These platforms can then be mined for relevant information.
- Internet archives – websites such as Wayback Machine should be systematically used to access archived material from certain sources that reported on parties and their structures/movements in real time and that may no longer be available on the live version of the relevant website. The technology officer(s) should consider efficient ways for researchers to review these sites.

When identifying potentially relevant sources, time should be spent reviewing the source to determine:

- what type of information the source provides;
- whether it is worth the time to systematically review it for information;
- where systematic review efforts should be focused for the source (e.g., one webpage on a website rather than the entire website); and
- whether an internet archiving tool should also be used for the source.

If the mapping work uses other types of sources (for example, confidential memos or physical documents), consideration should be given to issues regarding provenance, chain of custody, consent, security and confidentiality. Where necessary, protocols for the handling of various types of sources should be developed and followed.

**Tip:** There are numerous OSINT tools available to enable deeper, more efficient and systematic research. For example, tools can keep track of what has been researched, especially using which search terms on which websites. Other examples include tools that can run searches in online databases not accessible via conventional search engines, and tools that can automate the collection of different kinds of data.

**Tip:** Downloading and generating locally stored copies of online sources may be useful to protect against sources becoming unavailable in future. Downloaded content may allow advanced text searches that may not be available through online search engines. Cybersecurity, operations security and the implications of capturing and storing this content should be considered before doing so.

### 5.1.2 Source ratings

Sources will have differing levels of credibility and reliability for different types of information. It is important to take a position on the relative reliability of each source for the information it provides. Each source should be considered against an objective assessment criteria and given a value rating, for example, “high”, “medium” or “low”.

The value of sources should be considered for the different *types* of information. This is because the credibility and reliability of a source’s information can vary depending on what it covers. For example, an official SF website may be credible and reliable for information on SF structure and location information but not for information on who was responsible for a particular type of violation, given their partiality.

Annexure B contains sample source rating criteria.

The application of source value ratings in the database is explained further below.

**Tip:** Ratings should be decided through an internal consultative process as different researchers may have more experience working with some sources over others. Ratings for each source should be documented.

**Tip:** At this first phase, the digital landscape assessment might be rudimentary to enable the team to get started. As the work develops, the team will add to the sources being used via a consultative process.

## 5.2 Phase two: scoping incidents and parties

As noted in section 2.4, preparing a list of key incidents and key parties is a necessary preliminary step to determine which parties to map, which incidents to focus on if perpetrator analysis will be done, the time period to cover and the geographical areas to consider.

### 5.2.1 Incidents

Whether incident mapping has been done already or not, the core team should conduct a scoping exercise to develop a list of key incidents. If incidents have already been mapped, this can be done using the existing incident mapping database.

If incidents have not already been mapped, the team should use the digital landscape assessment to identify key sources that reported on relevant incidents, including well-known incidents and patterns of incidents. The sources should then be systematically reviewed for information on these incidents.[5] This initial incident scoping work will assist to define the scope of the mapping work and should be done even if perpetrator analysis of incidents will not be included in the mapping work.

### 5.2.2 Parties

At the scoping phase, an initial list of key parties should be developed. ORBATs and troop movement maps found during the digital landscape assessment will help to identify key operational parties. The review of key incidents described in section 5.2.1. above will also assist to identify key parties. Some time should be spent researching publicly available or known attribution allegations for key incidents and developing a list of key parties.

It is important to have an initial broad understanding of the structure of the parties being mapped to assist with the research carried out during the systematic review and targeted research phases. The team should ensure that hierarchy relationships and command structures are understood so they are accurately described during the systematic review phase. Preliminary research should be undertaken to understand the background of parties, their function, capabilities, and the typical hierarchy formations they followed. Relevant ORBATs and other hierarchy diagrams found during the digital landscape assessment will be crucial for this exercise. The understanding of party structures will become more sophisticated as the research progresses.

**Tip:** Time spent at the beginning of the mapping work to prepare preliminary research memos on various parties that can be shared with the team will provide a valuable knowledge base. This may include information on command structure, how specific units of a party worked (standard operating procedures, modus operandi), typical weaponry etc.

## 5.3 Phase three: systematic review of sources

### 5.3.1 Information to capture during systematic review

Using the digital landscape assessment, the team should identify which sources to use for the systematic review of sources phase. The type of information to capture from the sources will depend on the nature and extent of the mapping work and in particular, whether perpetrator analysis will be carried out.

Given the vast amount of information on parties that may be available in open sources, there should be strict parameters on what information researchers should extract in the systematic review process. Furthermore, priority types of information should be identified to assist researchers to move efficiently through sources.

Depending on the scope, the type of information to enter into the database can include the following:

- Unit information:
  - unit hierarchy information, that is, superior and subordinate unit information over time, as well as less conventional organisational information between units (e.g., parallel command structures or cross-links between sections of the SFs that would not be obviously apparent).
- Individual information:
  - unit commander and unit membership information over time; and
  - individual hierarchy information, as well as less conventional organisational information between individuals.
- Information regarding the location of units and individuals over time.

5. When gathering information on incidents, several issues should be considered. See Public Interest Advocacy Centre, *Conflict Mapping and Archive Project Methodology* (2020) <https://ipisresearch.be/mapping/webmapping/lka/context_documents/CMAP_methodology.pdf>.

- Specific allegations that a unit or individual was responsible or potentially connected to an incident.
- Other potentially useful information about a unit or individual such as the date a unit was raised, photos of an individual, dates of birth of individuals, whether a key individual is now deceased, medals awarded to individuals, battles an individual or unit was involved in etc. There should be specific guidance provided on what type of information researchers should enter in this field.

**Tip:** Mapping individual to individual hierarchy information comprehensively may not be necessary as the user interface should display this information automatically from the mapping of units and unit commanders.

**Tip:** Given limited resources, it will not be feasible to map detailed membership information for every unit. Researchers should focus on only entering unit membership information about individuals (non-commanders) who were members of known problematic units, or membership information of individuals themselves who were known to be of concern.

### 5.3.2 Location information

Capturing party location information is an important step in the research to assist with the investigation of co-located incidents at a later stage. Location information should be captured, stored and presented in a manner that is useful for any perpetrator analysis that will be done. Entering and categorising location information raises a range of issues, which the research methodology will need to address, in consultation with the technology officer(s). The following considerations may be useful:

- **Precision of location information**: Sources will refer to locations with different levels of precision: sometimes more generally (e.g., "somewhere in X town"), and sometimes more precisely (e.g., "in front of the X hospital"). The research methodology should provide guidance on how to differentiate and record these levels of precision.

  Where location information is particularly imprecise or vague, the research methodology should address whether or how to record this information. For example, "somewhere in X province", may not be useful to record in some instances, but may be useful in others, depending on the context and the type of information.
- **Geocoding (entering coordinates)**: Locations should be geocoded, that is, location data should include coordinates. The methodology should provide a process for this. Depending on the context, integration with an existing source of geospatial data like OpenStreetMaps may be useful, or researchers may be required to manually find the coordinates of locations. The research methodology should provide guidance on which sources are more authoritative or should be consulted first.

  The research methodology should also provide a process for treating locations whose coordinates cannot be located.
- **Polygons and points**: whether it is necessary or useful to record locations as polygons (a geospatial shape that is made up of multiple coordinates), as opposed to singular points. Depending on the level of detail available for unit areas of command, it may be useful to record locations as polygons. However, for most location information, singular points (albeit with varying degrees of precision) are often sufficient.

### 5.3.3 How to systematically review sources

The digital landscape assessment will assist the team to determine which sources need to be systematically reviewed and which can be used for targeted information gathering.

The key sources should be systematically reviewed. For example, official unit websites may be particularly helpful for finding information on unit commander periods of tenure and should be systematically reviewed. When systematically reviewing websites, time should be spent scoping the website to determine where the most useful information is located on the site. The systematic review of the most useful parts of the website should be prioritised.

Other sources can be reviewed in a more targeted manner for specific types of information on units and individuals of interest. Certain sources can be deprioritised.

The decision to systematically review a source should be continuously reassessed to determine whether:

- the source is providing useful information;
- the scope of the source review needs to be narrowed;
- further resources are needed to systematically review the source; or
- systematic review of the source should cease as it is no longer worth the resources.

Document search tools are valuable during this phase to conduct targeted searches of archived or downloaded material.

### 5.3.4 Entering information as data points

Researchers should use the information gathered in the systematic review phase to enter data points into the database. These data points are the core content of the mapping process, showing relationships between and information about units, individuals and incidents. Importantly, this information is variable over time, meaning that each data point should have start and end dates.

Each data point should be supported by and linked to at least one source. This source or sources should also be contained in the database, for example as PDFs or in other easily accessible formats. The data point should also contain a reference for where in the source the information came from, for example, a page number. This allows researchers and end users to easily see the information's source, how information has been interpreted and entered in the database.

Each data point should be given a confidence rating, that is, how confident the researcher is that the data point is correct (see 5.3.6 below).

### 5.3.5 Making decisions about data points

When entering information, researchers must make decisions about whether to add information to the database as a new data point, to add information as a *further source* for an existing data point, or not to add the information at all. These decisions can be informed by the following principles:

- If consistent information is already entered as a data point with a high confidence rating, the information would likely not need to be entered. This avoids unnecessary duplication of work for obvious or uncontested pieces of information.
- If consistent information is already entered as a data point, but the confidence rating is low, the information should be entered as a further source for the existing data point, and the confidence rating updated accordingly.
- Where information does not exist in the database, a new data point is likely required.
- Where information is inconsistent with data in the database, a new data point is likely required, which notes the inconsistency.

These decisions are interpretive, as it may not be readily apparent whether information is substantially similar or, instead, is distinct enough to justify creating a new data point. Some examples include:

- **Location data points**: for example, one data point indicating a unit was "in Town X" and another that says, "north of Town X". These should likely be two data points as the nuances may be important (e.g., for shelling incidents where the direction of fire is critical).
- **Data points over time**: for example, if one source reports that a unit is in a location on one date, and another source places the same unit in the same location one year later, this information could be entered as one data point with a one-year long date range, or two data points each with a single date.

The methodology should provide guidance for researchers in making these decisions. They will be informed by contextual knowledge of the SFs, for example, regarding whether units are mobile, information about superior and subordinate units, how the source is rated and so on.[6]

**Tip:** Guides should be developed to assist researchers as they enter data points in the database. This includes guiding researchers on how to identify when an individual or location they come across during their research is already in the database but with different spellings, to avoid duplicate records in the database.

### 5.3.6 Confidence rating

As part of the data entry process, confidence assessments should be applied to each data point. Confidence assessment criteria will need to be developed in order to decide which confidence assessments to apply to a data point. The ratings applied to sources will be important for this process. See Annexure C for sample confidence assessment criteria.

In making confidence assessments, researchers should assess the confidence of both the information entered in the data point (e.g., linking a unit to a location) *and* the date range for the data point, including whether the date range entered was based on an inference between multiple sources. Refer to the discussion and footnote above at section 5.3.5.

6. The Security Force Monitor Methodology provides insightful discussion of this issue, see Security Force Monitor, *Research Handbook* (2022) 'Data integrity measures' <https://help.securityforcemonitor.org/en/latest/data_integrity.html#timebound-data>

The confidence rating for each data point should be dynamic and updated when researchers enter new information in the database. For example, a data point with a low confidence rating may be given a higher rating after a new source is added that supports it. Alternatively, a data point with a high confidence rating may be given a lower rating if an inconsistent data point is entered. The confidence ratings will also be informed by the credibility and reliability of the sources (see section 5.1.2 below).

### 5.3.7 Language and data entry conventions

The sources used in the mapping exercise may vary in language depending on the context. All information analysis should be in a consistent language. Consistent language and conventions can assist in searching information, coding and processing data, whether that information is stored in a database or other information system.

A style conventions guide can help ensure consistency. This should include:

- how to refer to alleged attribution, for example, whether words like "alleged" or "suspected" will be consistently used;
- how foreign language texts will be used, and if they will be inserted in their original format with translations into English included;
- how sources will be cited and differentiated within any analysis provided, especially where there are differing sources and multiple pieces of information from each source;
- how root sources are dealt with/ expressed, for example, is the source being used basing their information on another source like a press release or a government spokesperson, or on independent research/investigations;
- how to address common language variations when conducting online research and when entering foreign language words, including developing spelling suggestions for researchers (e.g., for location and individual names) to ensure consistency between information recorded and to ensure research is thorough;
- which date format to use, and numbering formats (spelling or numerals);
- what tense to use in text descriptions;
- how to refer to key parties, for example, non-state actors, an opposing party etc;
- how to address language that might appear partial but is useful, for example if the word "rebel" is used it could either be entered in quotation marks or could be switched to the preferred terminology in the style guide;
- how to record unit and individual location information when locations are reported vaguely in sources. For example, if a source reports that a unit was 10 km north of X TOWN, it is important to place the unit "10 km north of X town" rather than in "X TOWN" to maintain accuracy; and
- how to name individuals when their full name is not known.

**Tip:** A written guide for the systematic review of sources phase should be provided to all researchers with updates discussed and shared. This must include a language and data entry conventions guide that is regularly updated. Free text fields are easier to search when researchers use a prescribed format, language and spelling.

**Tip:** All alternative spellings/names of individuals should be captured in the database allowing end users to locate the individual in the database irrespective of the name they have.

## 5.4 Phase four: review, revise, research

### 5.4.1 Secondary review

The systematic source review work in phase three should mainly be carried out by researchers after an initial training. The systematic review work should then be reviewed by senior members of the team and individual feedback provided. Issues to be considered during review include:

- Consistency with project methodology and style guide.
- Accuracy of information; logic of inferences; and how this is accounted for in confidence ratings.
- Coherence of data points and whether any data points should be consolidated or separated. Keep in mind any nuances that need to be accounted for, for example "a unit was in town X" versus "a unit was north of town X".

**Tip:** Due to the volume of information entered, not every data point will be reviewed. Secondary review needs to be strategic and targeted. This can be done by doing randomised spot checks, by reviewing the work of the newer, less experienced researchers more closely at the start.

**Tip:** A review process should be integrated into the information system that tracks when data points are reviewed, who reviewed them and when, and allows for reviewers to share feedback with researchers.

### 5.4.2 Targeted research

Following the systematic source-review phase, there will inevitably be gaps and inconsistencies in the information profiles for key parties. During or shortly after the secondary review phase, a profile overview should be compiled to identify information gaps and inconsistencies in the profiles of key parties. Targeted research should then be carried out to address these issues. Filling information gaps for party hierarchy, party location information and unit commander information should be prioritised.

The effectiveness of targeted research requires thorough training and frequent information sharing across the team on proven search strategies and techniques. A regularly updated guide should be shared with the team providing instruction on best practice for OSINT search techniques including:

- the usefulness of various sources for different types of information;
- the usefulness of different search engines;
- the relevance of alternative spellings;
- useful key words and search term combinations;
- site searching techniques;
- useful search operators; and
- conducting searches of downloaded archives of online sources.

**Tip:** Given limited resources, set parameters around the scope of the targeted research phase. Key parties linked to key incidents should be prioritised and a clear time frame for this phase of the work agreed.

## 5.5 Phase five: perpetrator analysis

Perpetrator analysis involves investigating incidents to determine which units or individuals may have been responsible for the incident.

Where only parties are being mapped and not parties in relation to incidents, this phase of the research methodology can be excluded.

**Tip:** Online evidence collection applications are valuable tools to deploy across the team during this phase.

### 5.5.1 Priority incidents

The incidents identified in phase two should be reviewed to prepare an investigations priority list having regard to the following criteria:

- likelihood that the incident was committed by one of the parties being mapped;
- relevance of the incident to the issues under consideration in the mapping work;
- gravity of the incident; and
- extent of existing open-source information available on the incident.

### 5.5.2 Conducting analysis

Once the initial priority list of incidents is prepared, the team should begin investigating the incidents. Note, the priority list may be regularly revised throughout this phase.

The following approaches are useful for investigating who may have been responsible for an incident:

- leveraging the existing information assembled in the database, including in particular the co-location information for parties in relation to the location of incidents and any preliminary leads about units/individuals connected to incidents recorded during the systematic source review phase;

- thoroughly reviewing all sources available on the reporting of the incident;
- if relevant, preparing troop movement maps for certain periods in a conflict to assist with determining how troops moved in relation to territory captured;
- considering other known information for a party (unit or individual) based on investigations into other incidents, comparing these and using them as research leads to build a better understanding of that party; and
- additional in-depth targeted research into incidents, including analysing video footage, photographs, satellite imagery.[7]

Depending on the purpose of the mapping work, researchers may be only looking at commanders, or they may be looking at all members of implicated units, or a hybrid. The purpose of the mapping should be clear to ensure researchers do not waste time on detail that may be irrelevant.

**Tip:** The team should approach the analysis from different entry points, including structuring the research around a single incident, a unit or individual of interest, a cluster of incidents or a geo-temporal pattern of violations. Researchers working on similar or related entry points should communicate regularly to share tips and ideas, ensure consistency in analysis and check biases.

### 5.5.3 Entering analysis into the database

Perpetrator analyses should be captured and represented on the profile of the relevant parties and the incident in the database. As much information as possible should be provided for the analysis conclusion. Explanations for each conclusion should be clearly outlined and all supporting research documentation should be captured, stored and easily accessible to end users.

It is very important to be clear to end users about the limitations of any perpetrator analyses. The factors affecting the limitations of the mapping work are discussed above at section 2.2.

Investigation outcomes may result in some conclusions being stronger or weaker than others. The team should develop agreed terminology for distinguishing between the potential varying strength of analysis conclusions so that these distinctions are clear to end users. If appropriate, the team may decide to rate their confidence in the perpetrator analysis.

### 5.5.4 Secondary review

All perpetrator analyses and the accompanying explanations must be reviewed by senior team member(s). Reviewers should pay close attention to:

- accuracy of the analysis;
- logic of inferences;
- basis for assertions and conclusions; and
- accuracy and appropriateness of language used in the analysis.

## 5.6 Phase six: final review

At the end of the substantive research, time should be dedicated to conducting a final review to determine whether any further research is needed for particular parts of the data. For example, the profiles of parties with connections to incidents may need to be further reviewed to identify and fill information gaps or a review may be needed to identify and resolve any errors or inconsistencies on parties' profiles.

If the mapping work is in relation to an ongoing conflict or unrest, then there may be no fixed end date in sight. If this is the case, then periodic reviews should be scheduled to assess if there are gaps or new avenues that should be explored.

Depending on the purpose of the mapping work and the range of potential end users, it may also be important at this stage to do thorough trialling of the final database interfaces to ensure the information is presented as usefully and intuitively as possible.

7. For OSINT tips, see for example, 'Resources', *Bellingcat* (Web Page) <https://www.bellingcat.com/category/resources/>; Sam Dubberley, Alexa Koenig and Daragh Murray (eds), *Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability* (Oxford University Press, 2020).

# 6. Using and Sharing the Analysis

## 6.1 Developing guides to assist end users

If sharing the mapping work with others, the team should consider preparing support documentation for end users, including:

- **research methodology**: a document clearly setting out the research methodology followed to create the mapping work. Ideally this information was continually documented throughout the research; and
- **database guide**: a document setting out how to navigate the mapping data, and how to read and understand the information presented.

## 6.2 Questions of continuity

Continuity is a significant challenge with all technology dependent work. When funding is being sought or administered for this kind of work, thought must be put into longevity. Questions include:

- How long should the information and analysis be available for?
- Where will the information continue to be stored?
- Is the mapping work part of a fixed project but are the violations continuing? If so, will the information be updated beyond the term of the project and who will do this?
- What ongoing maintenance, security or subscriptions will the database require, and who will fund this?
- Who will need future access to the database and how might it be used?
- Are there any ongoing security concerns to consider regarding staff or anyone else involved in the work?

# 7. Conclusion

Investigators of IHL and IHRL violations are faced with enormous challenges and limited resources. Open-source information has become an ever-increasing, critical resource for investigators to capture and leverage, particularly at the earlier stages of an inquiry when investigation strategies are tested and formulated. The vast scale and disparate nature of open-source information presents a daunting challenge for analysts. There is often little guidance available, particularly for systematic and methodical reviews of large troves of open-source information. Yet the value derived from capturing, analysing and structuring this information cannot be overstated.

The most valuable mapping exercises are those that are carefully planned, rigorously executed and that present information in a way that is accessible for users to retrieve and leverage.

The guidance and methodology set out in this document will support practitioners to create a thorough and well-structured map of parties to conflicts or particular events. The mapping will serve as an integral tool for the investigation of parties' involvement in IHL and IHRL violations during the conflict or events. By setting out necessary considerations in parties mapping, this guide contributes to the continued professionalisation of the OSINT community, a critical component in the fight to end impunity for IHL and IHRL violations.

# Annexure A: Suggested Data Model and Interface Design

This annexure advances on the considerations discussed in section 4. It outlines a suggested implementation of parties and perpetrator mapping in a relational data model and a GUI for researchers working in the database.

# 1. Relational Data Model

## 1.1 Entities

To map parties and potential perpetrators of IHL and/or IHRL violations, the data model's primary purpose should be to hold information on **Units** and **Individuals**, and to connect this information to **Incidents**. If an incident mapping exercise has already been conducted, the parties mapping database should be integrated with the incidents database. This data model does not discuss the creation of an incidents database (or events database), which are conducted elsewhere.[8]

In a relational database, these correspond to three entities or tables:

- **Units.**
- **Individuals.**
- **Incidents** (may be derived from a separate database).

Further entities include:

- **Source materials:** holds documents and other content that is the basis for information contained in the database.
- **Data points:** holds the information about units and individuals.
- **Data point sources:** links data points to their respective source materials (one or more for each data point).
- **Locations:** holds information about locations.
- **Attributions:** links units or individuals to incidents.

An entity relationship diagram is included at the end of this Annexure.

## 1.2 Data points

The data relating to parties (i.e. data about units and individuals) face two key issues:

1. Variability over time: for example, commanders of units change, the location of units change, and unit hierarchies change.
2. Inconsistencies between sources: some sources may differ in their accounts from others. Recognising consistencies and inconsistencies is essential to the phases of research and analysis (see, for example, section 5.3.4).

To address these issues, the data model structures the data relating to parties as **Data points**, separate from the party's 'inherent information'.

Inherent information would instead be stored in the **Units** or **Individuals** entity respectively. This information could be a unit or individual's name or their affiliation (e.g., that they are affiliated with the army), which are pieces of information that are unlikely to change over time or be the subject of inconsistencies between sources.

### 1.2.1 Types of data points

**Organisational structure and hierarchies**: data concerning the organisational structure and command hierarchies of parties can be entered in the form of links. These are:

- Units – Units.
- Units – Individuals.
- Individuals – Individuals. Note that many of these will correspond to the hierarchy established for units. However, there may be other relevant links.

**Locations**: geospatial information is essential to a parties mapping exercise, and the data model must enable the location of parties to be recorded over time.

This data model structures **Locations** as a separate entity. This choice helps to ensure that locations are entered accurately and consistently, to avoid duplication (and the associated risk of errors), and to better assist researchers in analysis. It would also allow the data to interface with online sources of geospatial data, for example, OpenStreetMap or Google Maps. Databases that are spatially enabled, that is, those which are specifically designed to hold spatial information, may also be useful for these reasons.

8. See Judith Dueck, Manuel Guzman and Bert Verstappen, *HURIDOCS Events Standard Formats: A Tool for Documenting Human Rights Violations* (HURIDOCS, 2nd edn, 2001) <https://huridocs.org/wp-content/uploads/2020/11/HURIDOCS_ESF_English1.pdf>; Patrick Ball, *Who Did What to Whom? Planning and Implementing a Large-Scale Human Rights Data Project* (American Association for the Advancement of Science, 1996) <https://hrdag.org/whodidwhattowhom/contents.html>.

Parties' location information can also be considered a form of link as follows:

- Units – Locations.
- Individuals – Locations. Note again that many of these links will correspond to the locations of units. However, there may be other relevant location links.

**General information about parties**: the data model should allow for further information to be entered about a unit or individual that does not fall into one of these categories, for example, photos, dates of birth and medals awarded. As with all other data points, this information should be linked to a source.

### 1.2.2 Fields for data points

**Data points**, which include links relating to organisational structure, locations, or general information, should hold a range of information:

- Type of relationship:
  - For a 'unit – unit' or 'individual – individual' link, is it a hierarchical relationship where one is superior and the other subordinate, or is it some other form of organisational relationship?
  - For a 'unit – individual' link: is the individual a commander of the unit, do they occupy another position within the unit, or is the individual just a member of the unit?
- Dates: the period of time for which the data point is accurate. This should be accompanied by a text field commenting on why the researcher made the decision. See discussion at section 5.3.4.
- Description: a text field in which the researcher enters narrative, analysis or commentary about the data point that other fields do not capture.
- Confidence rating and assessment: discussed at section 5.3.6, an assessment of confidence in the data point. This should be accompanied by a text field commenting on why the researcher made the decision.

## 1.3 Data point sources

Data points should each be supported by one or more sources contained in the source materials entity. Implementing this in a relational database will generally require creating an additional entity, here called **Data point sources**.

A 'data point source' is effectively a link from a data point to a source. This link should contain further unique data:

- Reference: where in the document or content the relevant information is found (e.g., 'page 5')?
- Root source: in the source material, who or what is attributed as the original ('root') source of the information?

This data can be used by researchers in making decisions about the data point, for example in the confidence rating and assessment.

## 1.4 Attributions

**Attributions**, or links between parties and incidents, present a suggestion that a unit or individual may have been involved in an incident. How exactly these are analysed, justified and presented will depend on the purposes and scope of the mapping exercise.

Attributions are also entered as links:

- Units – Incidents.
- Individuals – Incidents.

Importantly, attributions should be able to be supported by (and linked to) multiple other data points, such as hierarchy or location data points. This linking enables the attribution link to be supported by information from across the database. This requirement follows from the phases of research and analysis set out in section 5.5, in which making attribution links comes after data points are entered.

The attribution link will at least contain a text field in which the researcher enters narrative, analysis and evaluates information, drawing on all linked information in order to support the claim.

In addition, it may be useful to have a field that relates to the strength or weakness of the attribution. The content of this field should be determined according to the scope and purpose of the mapping exercise.

Finally, the research and analysis process may require linking additional documents or content directly to the attribution link, for example, intelligence reports or documents resulting from targeted research, which are not otherwise linked to data points (see section 5.4.2). The data model should allow for additional sources to be linked to the attribution link.

# 2. Researcher Interface Design

In designing how researchers will interact with the database, most often through a GUI, there are a range of important considerations.

The design of workflows should accord with the phases of research and analysis as well as the underpinning information system. In the parties mapping exercise, this will likely require the following interfaces:

1. **Collecting and storing source materials.** This is a workflow in which researchers can easily (or automatically) store relevant documents and other content, with appropriate metadata.
2. **Entering relevant information found in documents as data points.** This workflow should facilitate decision-making by the researcher about:
    a. What kind of data point could the information be entered as: for example, is it a commander relationship to a unit, or a different position within a unit? (section 5.3.1)
    b. What similar data points already exist in the database: for example, should the information be entered as a new data point, as a source for an existing data point, or not entered at all? (section 5.3.4 – section 5.3.5)
    c. What confidence rating should the data point be given? (section 5.3.6)

    Making content directly viewable in the GUI will greatly aid researchers, for example through a PDF viewer or video player. This is especially so for data points with multiple sources, where researchers should be able to easily switch between those sources and change or enter data alongside it.

    Visualisations, for example of geospatial data, are also helpful for researchers.
3. **Profiles of parties**. After data points about parties have been entered, a review stage will follow, one part of which requires reviewing information entered about parties (section 5.4). To facilitate this, an interface that effectively displays a 'profile' of a party, containing a range of information should include:
    a. For units:
        i. the unit's position in a hierarchy or organisational structure, including its superior units, subordinate units, and other units it had relations with, over time;
        ii. its commanders and other unit members, and their positions, over time;
        iii. its locations over time, ideally displayed on a map; and,
        iv. any further general information about the unit, over time.
    b. For individuals:
        i. the individual's position in a hierarchy or organisational structure, including their superiors, subordinates, and other individuals they had relations with, over time (this may be achieved by displaying the contemporaneous hierarchy of units the individual was linked to);
        ii. units the individual commanded or was a member of, including their position in the unit, over time;
        iii. their locations over time, ideally displayed on a map (this may be achieved by displaying any contemporaneous location data points of units the individual was linked to); and,
        iv. any further general information about the individual, over time.
4. **Perpetrator analysis – linking parties and incidents**. This interface is perhaps the most complex, as it requires researchers to be able to run queries for data points across the database, view multiple documents and incorporate targeted research, in order to conduct in-depth analysis.

    As discussed in section 5.5, this phase requires multiple entry points, for example, the analysis required to make attribution links could begin from the context of incidents or from the context of parties. One example of a suggested workflow is:

    a. Incident context: researchers select an incident or group of incidents.
        i. Researchers set criteria to run queries on data points in the database, for example, location data points within a certain radius of the incident, time period around the incident, or keywords relating to the incident or an alleged perpetrator.

ii. From the data points returned by the query, researchers select a unit or individual to inquire further about. Researchers can then see all the data points related to that unit or individual.

iii. Researchers can then conduct further research outside of the database, finding additional content to support or contextualise the attribution link.

iv. Researchers select one or more of the data points displayed in the database, and upload any additional content they have found.

v. Researchers make an attribution link and enter the analysis that supports the attribution. The attribution link should contain the selected data points (and their underlying source documents), as well as any additional content uploaded.

Attributions, their associated data points and analysis will often be relevant for more than one party for the same incident or group of incidents, for example, a unit and its commander, or other affiliated units (noting that the attribution may be of different strength). The workflow should consider how to apply similar analysis and data points across multiple parties.

5. **Features for reviewing, messaging, note-taking and updating:** additional features should be incorporated across all of the workflows to increase the quality of data and to improve the experience of researchers.

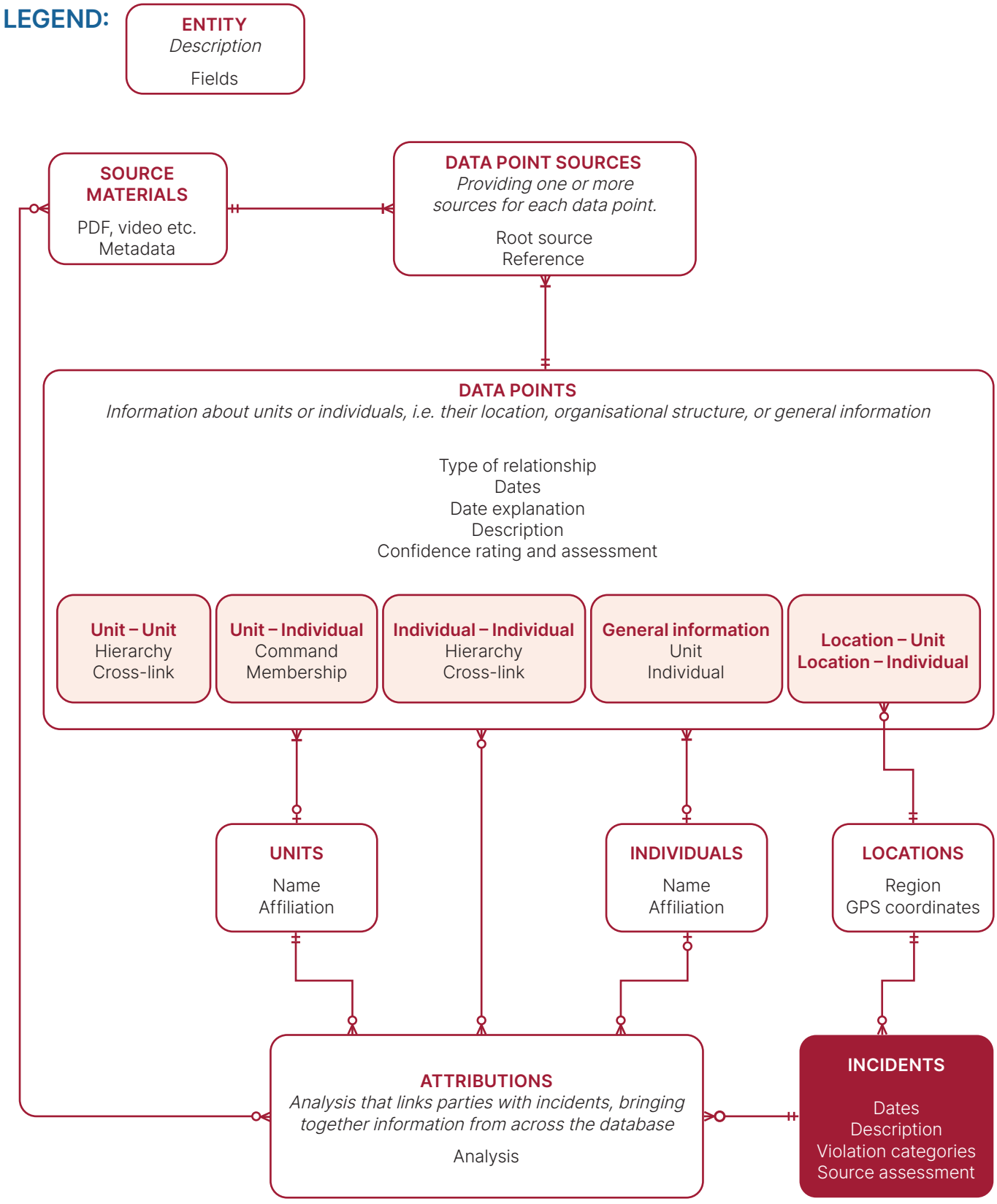
   a. Review: to allow reviewing of data where necessary, with a system to track what has been reviewed, by whom, when, and with any comments.

   b. Messages: to encourage communication between researchers, and also with reviewers, such as asking for opinions, clarifications, or drawing on different expertise.

   c. Notes: notes researchers make to themselves, which can encourage reflection, allow for researchers to take breaks and continue where they left off, and improve decision-making.

   d. Updates: a feature that allows reviewers or senior researchers to easily communicate information to the broader team, for example, a rolling list that appears on users' dashboards.

# 3. Entity Relationship Diagram



# Annexure B: Sample Source Ratings Criteria

## Purpose of source ratings

1. Information presented in the database is drawn from multiple sources.
2. Each source has been given a "rating". The rating has been chosen based on the decision criteria set out below.
3. The purpose of the source ratings is to:
   a. help researchers give an overall confidence rating to data points at each stage of data entry and analysis; and
   b. help end users of the database to assess the reliability of individual data points.

## Limitations

4. The ratings are not intended to make any general reliability or credibility conclusions about the sources used in the database or the quality of reporting by particular sources.
5. The ratings have been created for the mapping work only and sources have been considered according to the specific categories of information set out below.

## Confidentiality

6. This document is confidential and should not be shared beyond those involved in the mapping work.

## Source categories

7. Sources are rated based on the following categories of information:
   a. hierarchy: information about the structure of parties
   b. location: information about the location of parties;
   c. individual: information about an individual's membership of a unit (including whether they were the commander) and information about the relationship between individuals;
   d. general information: miscellaneous information about parties, for example the date a unit was established, promotions of individuals, medals an individual has received, education an individual has undertaken;
   e. incident information: information about the details of an incident, such as, type of violation, how many people were killed etc; and
   f. incident attribution information: analysis regarding who may have been responsible for an incident that might be a IHRL and/or IHL violation.

## Source ratings

8. There are four source ratings:
   a. high value;
   b. medium value;
   c. low value; and
   d. unknown.

## Decision criteria

9. A source is rated as **high value** for a category of information if:
   a. the source has first-hand knowledge of the information provided because it relates to the source's "core business"; and
   b. there is no reason to suspect that the source is biased in reporting this type of information.

   Where the information is provided by a secondary source:
   c. the source consistently relies on primary sources considered reliable for this type of information; or
   d. the source has a track record of using credible witness information for this type of information; and
   e. the source is typically specific for this type of information.
10. A source is rated as **medium value** for a category of information if:
    a. the source is not known for having developed expertise or knowledge on this type of information; or
    b. the information is not always consistent with information provided by sources considered credible for this type of information; or
    c. the primary sources for this type of information are not always credible sources for this type of information or are not always known.

11. A source is rated as **low value** for a category of information if:
    a. the information reported has been routinely contradicted by independent and thorough fact-finding by credible and competent entities; or
    b. language used in the source generally indicates a bias for this type of information; or
    c. the source only reports allegations by chosen parties without any fact-checking; or
    d. the source features one or more authors which are unknown, for example, an online blog.
12. A source rating is **unknown** for a category of information where it is not possible to determine the accuracy of a source for the information provided.
13. A source rating is **N/A** for a category of information where the source does not report on that type of information.

# Annexure C: Sample Confidence Assessment Criteria

A confidence assessment criteria will need to be developed in order to decide which confidence assessments to apply to a data point. This criteria should be read alongside the Sample Source Ratings Criteria (see Annexure B).

The following principles may be followed when giving confidence assessments to data points:

**a.** A "very confident" confidence assessment is applied if:

- **i.** the data point is based on information from one or more sources that are deemed to be high value for the type of information provided; or
- **ii.** the data point is based on information from different sources that is consistent and specific; or
- **iii.** the information is particularly detailed; or
- **iv.** there is no other logical counter option for the inference made; and
- **v.** the date range for the data point is reasonably clear and specific in the source.

**b.** A "moderately confident" confidence assessment is applied if:

- **i.** the data point is based on information from one or more sources that are deemed to be medium value for the type of information provided; or
- **ii.** the data point is based on an inference that is likely; or
- **iii.** the information in the sources is generally consistent but there are some minor inconsistencies; or
- **iv.** the date range for the consolidation is not clear in the source or the date range is based on an inference or a temporal contiguity assessment between two dates.

**c.** A "not very confident" confidence assessment is applied if:

- **i.** the data point is based on one or more sources that are deemed to be low value for the type of information provided; or
- **ii.** the data point is inconsistent with another data point; or
- **iii.** the data point is based on an inference in a source with no known basis; or
- **iv.** the data point is based on an inference that is a mere possibility only; or
- **v.** the date range for the data point is broad and uncertain.

**d.** An "unknown" confidence assessment is applied if an assessment could not be made.