Syria Justice and Accountability Centre

Data Collection and Documentation for Truth-Seeking and Accountability

Prepared by
Megan Price and Patrick Ball

January 23rd, 2014

Memorandum Series:

Documentation in Transitional Justice

Data Collection and Documentation for Truth-seeking and Accountability

Introduction

The United Nations describes transitional justice as "the full range of processes and mechanisms associated with a society's attempt to come to terms with a legacy of large-scale past abuses, in order to ensure accountability, serve justice and achieve reconciliation." [17] Comprehensive transitional justice may include criminal prosecutions, reparations, institutional reform, "truth recovery" via truth commissions, commissions of inquiry, or other fact-finding missions, and memorialization. This memo will primarily focus on "truth recovery" via data collection and documentation.

Societies undertake truth-seeking when there is a truth which has been obscured, denied, or ignored, so it is often difficult to determine what is actually the truth. As asked by Chapman and Ball [3] in the specific context of their examination of lessons learned from truth commissions, we must begin by reflecting on the question, "what kind of truth is being sought?" As Parlevliet [11] puts it, truth "is so commonly used that it seems to be a transparent notion, clear to all who are involved or interested in redressing past abuses, but 'truth,' like 'justice' and 'reconciliation,' is an elusive concept that defies rigid definitions."[1]

Therefore, such truth-seeking projects should be viewed as ongoing, evolving processes. Every piece of data collection and documentation is a building block in that process. For example, in the late 1990s numerous groups, including Physicians for Human Rights, Human Rights Watch, and the American Bar Association/Central European and Eurasian Law Initiative collected information about victims of the conflict in Kosovo. This information was analyzed by researchers at the American Association for the Advancement of Science, and presented in a series of reports and testimony to the International Criminal Tribunal for the Former Yugoslavia (ICTY) [1]. Over a decade later, the Humanitarian Law Center in Belgrade is continuing this work, completing the Kosovo Memory Book 1998-2000, which commemorates victims of this conflict. Similarly, records collected by the International Center for

*The content of this memo is largely the institutional memory of the Human Rights Data Analysis Group. Anita Godhes, Tamy Guberek, Jule Krüger, and Romesh Silva all contributed additional content via a team discussion and subsequent review and feedback on earlier drafts of this memo.

[1] For examples of further discussions of different kinds of truth within the context of transitional justice, see, among others, works by South African jurist Albie Sachs [14] and The Berghof Foundation [2].

Human Rights Research in Guatemala in the mid-1990s were crucial inputs into the 2013 trial of General Jose Efrain Rios Montt for charges of acts of genocide.[2]

Accountability

Accountability is strongest when it can be based on a truth-seeking mechanism in the transitional justice process. In our experience, accountability mechanisms have not, historically, been very useful mechanisms for truth-seeking. Court cases do not produce the truth. They produce a very specific set of information, governed by (not-very-transparent) rules of evidence. For example, portions of significant statistical evidence, including two datasets, were ruled as "hearsay" in the ICTY case Prosecutor v. Milan Milutinovic, Nikola Sainovic, Dagoljub Ojdanic, Nebojsa Pavkovic, Vladimir Lazarevic, and Sreten Lukic, despite the acceptance of Patrick Ball as an expert witness and the admission of his and his team's work as evidence. As a further example from this same case, Hoover Green [8] highlights[3] "…the differences between academic argumentation (truth-seeking) and the goals of adversarial legal argumentation (invalidating the opponent's argument)."In contrast, truth-seeking mechanisms are quite valuable, and often under-utilized, to inform accountability mechanisms. The informal effect of truth-seeking processes is to bring a lot more information to light; society as a whole and people who participate become vastly better informed through truth-seeking.

In contrast, truth-seeking mechanisms are quite valuable, and often under-utilized, to inform accountability mechanisms. The informal effect of truth-seeking processes is to bring a lot more information to light; society as a whole and people who participate become vastly better informed through truth-seeking.

We have not seen many accountability mechanisms that utilized information from truth commissions or other truth-seeking mechanisms; Guatemala is the exception to this observation. In addition to the Rios Montt case mentioned in the previous section, another example from Guatemala includes the Myrna Mack case. Data from the Historical Clarification Commission (CEH[4]) was introduced in the Guatemalan Supreme Court Case investigating the death of Myrna Mack.[5] These records provided important historical context for her death and established that her murder was likely politically motivated.[6] Although these are the too few examples of which we are aware of truth-seeking mechanisms informing accountability mechanisms, we hope that future accountability mechanisms will be more likely to follow these examples. The incorporation of information from truth-seeking processes can not only lend additional credibility to accountability processes, but also provide the broader community with more of a sense of ownership of a specific, potentially narrow, accountability process.

[2] See HRDAG Guatemala project page.
[3] Referencing an interview with ICTY defense expert Eric Fruits.
[4] The acronym refers to the commission's name in Spanish Comision para el Esclarecimiento Historico
[5] Myrna Mack was murdered in 1990; in 2004 the Guatemalan government acknowledged its responsibility for her death. See the Myrna Mack Foundation for more details.
[6] Principal ruling of Guatemala's Third Court.

Data Collection

It is too early to know precisely which "truth" will be sought, much less which transitional justice mechanism(s) may be employed, once the ongoing conflict in Syria comes to an end. As a result, we suggest that data collection efforts should focus on the question "what do you need to remember?" There are many threats to remembering, ranging from those affecting the victims and witnesses themselves (i.e., the initial possessors of the information) to those affecting the documentation once it has been collected. For example, as time goes by victims and witnesses may become so dispersed as to be difficult or impossible to locate for data collection purposes, or their perceived incentives for telling their stories may be reduced, or their ability to recall salient details may degrade. Ultimately, the victims and witnesses may die, as a result of conflict violence, accident, illness, or simply old age.

Once information has been collected from these individuals, it still needs to be protected from loss, via computer crashes or theft, or outdated technology that can no longer be accessed. Data collected now, in the midst of the conflict, should be viewed as starting points for future investigations, not as data which will be used directly in future transitional justice processes. Transitional justice will be shaped by the mechanisms which emerge either through imposition by the victory of one side or through negotiation once the conflict ends. For now, data collection should focus on preserving key details for future investigation. As will be developed throughout the rest of this memo, this preservation involves not only documenting events and victims, but keeping that information safe and accessible for the future.

In our experience, data loss can be the result of a variety of causes. Data preservation should focus not only on securely backing up information (preferably outside the country) but also on saving information in a format and organizational structure that is likely to remain applicable and useful months, years, even decades in the future. Encrypted software such as Martus or SpiderOak or commercial products such as Google documents or Dropbox or other programs enabling a central collection headquarters and cloud-based back-up should be employed. Crucially for collecting data on the conflict in Syria, neither the central collection organization nor the cloud-based back-up should be located in or visibly dependent upon either the United States or Russia. It is critical that data collection and storage is viewed as neutrally and objectively as possible.

Collection of meta-data is also important. What we mean by meta-data is the contextual information around the data collection process: what group collected the information, how they found it, what technology they used to preserve it, how corrected information was introduced when errors were detected, how they identified duplicated information and what they did with the duplicates. This documentation explains the data generating process, which will be crucial when generalizing from what is observed and recorded to the broader universe of violence that occurred during the conflict.

It is also important to think about data generating processes in the midst of active violence versus post-conflict. The data available now may differ significantly in terms of type, content, and collection method from what may be available following the conclusion of the conflict (at which point the

possibility of formal transitional justice mechanisms, such as a truth commission, may provide new and valuable sources of data). We need to not only leave open ways to incorporate these other types of data that are not currently available, and may only come into existence months or years from now, but also to refocus resources on enabling future data collection. Specifically, this means devising a data storage and preservation scheme that is sufficiently flexible to preserve an increasing amount of data and a wide range of types of data (unstructured text, audio, photo, and video, in addition to more conventionally structured databases). It also means developing a team of data collectors who think innovatively in terms what might constitute "data."

One of the key points here is that different types of data can be triangulated later. This can (potentially) provide an analytical advantage if the strengths and weaknesses of the various data types are complementary. The more diversity in data types and sources the better the chances for such an advantage.

3.1 Different Kinds and Sources of Information

Sources of data include victims and witnesses, social and traditional media, and bureaucratic and administrative records. The benefits and drawbacks of each of these sources, plus specific examples from each of these broad categories, will be elaborated in this section.

Memories of victims and witnesses are in most cases the primary source for transitional justice. It is worth noting that these memories may be voluntarily offered—the narratives and testimonies that a victim or witness provides—or elicited through pro-active pursuit. In the case of the former there may be fundamental differences between victims and witnesses who are sufficiently motivated to volunteer their stories versus those who respond to an interview or survey. In the case of the latter researchers must think beforehand about what needs to be asked of victims and witnesses to elicit the most pertinent details. For example, the frequency with which certain violations were reported to the Truth Commission in Sierra Leone (volunteered narratives) differed substantially from those collected by a retrospective survey (elicited narratives) (see Gohdes, 2010 for details, [5]). Both are valuable, and both are necessary, to determine a more complete picture of what occurred during the conflict.

Social media sources, including Facebook, YouTube[7], Twitter, and personal and institutional webpages, provide a source somewhere in between these two types of accounts. Typically the information provided to these sources is volunteered by victims or witnesses, but it may also be indirectly elicited by the emerging cultural norm of sharing such information online. The type and amount of information shared will also be shaped by the medium. Twitter only allows a limited number of characters; YouTube lends itself to video; Facebook blends text, photographs, and video, but all in limited quantity; personal blogs may lend themselves to longer narratives. It should also be

[7] Of note is the Human Rights Channel, part of a collaboration between WITNESS and Storyful; there are also several Syria-specific channels.

noted that all of these sources require additional verification and data derived from such sources should be backed up independently from the original social media source.

Although emerging technology is providing a new and valuable source of data generally, and in Syria specifically, it is crucial that researchers bear in mind the inherent limitations and complexities of this data source. For example, social media reports, like many other sources, are subject to selection bias.[8] When generalizing from patterns in social media data researchers must explore the category of individuals who are likely to have access to and choose to use social media—younger, technologically savvy, motivated individuals—versus those whose stories are likely missing from social media—older, more remote, lacking access or ability or motivation to use technology.

This same thought experiment should be performed regardless of the data source. Researchers should always ask themselves, "whose stories are not captured by this source? how are those stories, victims, and witnesses likely to differ from the ones documented by this source?"

We must also consider issues around recall and disclosure bias and the construction of memories. Recall and disclosure bias are two different ways that a victim or witness's narrative (whether voluntary or elicited) might differ from objective truth. Recall bias is specifically related to an individual's ability to accurately remember and describe events. The challenges this poses to retrospective surveys has been extensively documented. As an example, Checchi and Roberts [4] find (perhaps not surprisingly) that longer recall periods (i. e. , the more time has passed since the event) result in more uncertainty around dates and event details and bias recall toward the most violent events.

Disclosure bias is related to a victim or witness's incentive (or disincentive) to include certain events or details in his or her narrative. For example, victims and witnesses may shape their narrative based on what they believe is most likely to benefit them. Utas [18] studied Sierra Leonean refugees, and concluded that the very high rate of reported rape was a result of the refugees' perception that victims of sexual violence were more likely to receive assistance from aid agencies. Alternatively, an individual may respond in a certain way to meet perceived social or cultural norms [10] or to fit their memories into the broader narrative to which they have been exposed (see Pennebaker and Banasik [12] for a thorough treatment of individual and collective construction of memory).

Another valuable source of information is what we consider "found"[9] data. This category of data includes an almost infinite set of possible sources, and emerging technology is expanding this list even more. Examples include bureaucratic and administrative records, such as the Historic Archive of the National Police in Guatemala (AHPN[10]) [6, 15, 13] and documents generated by the Documentation

[8] Also often referred to as "reporting bias," defined by Kruger et. al. as "…the likelihood that a given event is reported varies with characteristics of the event itself, or with characteristics of the agency collecting reports." See also this HRDAG blog post on raw data and this HRDAG blog post on convenience samples for further discussion of sources and types of bias and the potentially large impact such biases can have on substantive conclusions.

[9] See Klingner and Silva [9] for a more extensive discussion of found data.

[10] The acronym refers to the organization's name in Spanish Archivo Historico de la Policia Nacional.

and Security Directorate (DDS) in Chad [16]. Border crossing records, such as those kept by the Albanian border guards during the 1999 conflict between Yugoslavia and NATO, can provide vital information on refugee flow, as demonstrated in analyses by the American Association for the Advancement of Science [1].

Such official, bureaucratic records, often kept by the perpetrators themselves, are invaluable sources of information. Not only can they provide a different perspective on events, they are often more difficult for skeptics or apologists to refute or ignore than statements from victims and witnesses. As an example, analyses of documents from the AHPN have played an important role in the prosecution of two police officers for the disappearance of Edgar Fernando Garcia and in the establishment of the former Police Chief's command responsibility for that disappearance.[11]

We should be open to and prepared to preserve a wide variety of these potential types of data. For example, when Reed Brody and Olivier Beurcault found the DDS records the initial struggle was simply in photocopying them. Patrick Ball's ability to scan the Albanian border crossing records in the field was vital to preserving that valuable data source. Civil society groups working in newly liberated Syrian towns to document recently occurring violence should be prepared with portable scanners and the means to back up those files on a server outside the country. Here too it is important that meta-data is included in the documentation process—where did the records come from? How were they discovered? Who had access to them? How were they preserved?

Lastly, technical evidence, such as that generated by forensic anthropology, autopsies, exhumations, medical examinations of victims, and satellite imagery, is also a vital component of "truth recovery." The incorporation of these types of sources should also be considered and prepared for in any data gathering project.

3.2 Specific Information Characteristics

Several questions should motivate the specific information that is collected. What information that is available now will be most needed to carry out further investigations post-conflict? What information that is available now is most likely to be lost post-conflict? Have we accurately and completely captured a victim's or witness's story? What are likely to be the pertinent questions raised during transitional justice mechanisms?

It may not be possible to collect all of the details of interest. We know that with respect to Syria questions are already being asked about the age and ethnicity of victims, and this is very difficult information to collect. But we can prioritize additional effort and resources aimed at these details, knowing that post-conflict they are likely to be among the questions raised. For example, field experts could review photos and videos of funeral processions, in an attempt to identify unique markers of ethnicity for victims.

[11] See Guzmán 2011 [7], Reuters, and the National Security Archive.

Another specific challenge in the current state of data collection in Syria is the lack of perpetrator information. The question of who is responsible for crimes is obviously already of great importance, and will only continue to be so once transitional mechanisms are put into place. Unfortunately, the vast majority of current records are missing this key piece of information. Is this because those collecting the data are afraid to ask? Because victims and witnesses are afraid to say? Or because they truly do not know?

Given the current state of the conflict in Syria, specifically the presence of many different groups of combatants, some of whom may not have readily recognizable traits, such as a uniform, it may be impossible to collect this information. And in fact identification of perpetrators may be biased toward those who do have recognizable traits. Additional effort and emphasis should be placed on collecting perpetrator information when possible, and clearly stating how the victim or witness identified the perpetrator.

It is crucial that subject-matter experts partner with data collectors and analyzers to frame quantitative questions. The absence of this partnering risks both failing to record critical details needed for subsequent analyses and simply sloshing through different troughs of data in search of a question. For example, one could ask "what is the pattern of all conflict-related deaths over time in Syria?" With the right data and analysis this question is answerable. But it does not immediately lend itself to actionable interpretation. In contrast, asking "what happened in Homs in the first quarter of 2012?" is a question informed by context—it is well established that a government offensive into Homs was launched in February 2012.[12] Data collection motivated by this question will be much more precise, directed, and has a higher chance of generating data that will be useful to future analyses and further investigations.

Generally, as much information on individual victims as possible should be collected—including name, sex, date of birth, date of event/violation, ethnicity, language, combatant/non-combatant status, location of event/violation, religion, home location/region, occupation, police, military, or other security affiliation and rank—by as many different (ideally independent) information providers as possible. This will enable later analysis to search for patterns of violence which link back to definitions of crimes against humanity, for example, that seek to demonstrate widespread or systematic nature of crimes. The more disaggregated and detailed each record is, the more useful the data are for detecting and adjusting for selection bias.

Conclusions

It is impossible to predict which specific transitional justice mechanisms will be preferred or even possible following the conclusion of the conflict in Syria. But in the meantime we can prepare for a wide variety of future research efforts by preserving data available now and thinking critically about questions that are likely to influence those investigations.

[12] As cited by the BBC and other news outlets.

With respect to data preservation, the most important goal is simply not to lose the data, through technical failure, disorganization, or theft. As mentioned earlier, software such as Martus or SpiderOak or commercial products such as Google documents or Dropbox or other programs enabling a central collection headquarters and cloud-based back-up should be employed and actual or implied affiliation with the United States or Russia should be carefully avoided.

We suspect there will be many more subsequent investigations in Syria, as compared to other conflicts because, candidly, situations in which the United States opposes the falling government tend to result in trials (e.g., former Yugoslavia, Rwanda, Iraq) while regimes supported by the United States tend toward amnesty-based approaches (e.g., El Salvador, Guatemala, South Africa).

With respect to possible future investigations - what are likely to be the questions and comparisons of interest? We know that questions about the age and ethnicity of victims have already arisen, but this information is very difficult to collect. Can and should we expend additional resources trying to fill in these gaps in existing documentation? How can we use information about the structure and organization of various opposition groups to better determine whether a victim or perpetrator is affiliated with any of those groups? Inevitably these details will be necessary, not only to address questions of accountability, but to gather a better understanding of what happened during the conflict. For example, how do patterns of violence change when control over a region changes hands?

Lastly, we need to be open to and prepared to preserve a wide variety of types of data. Official archives and bureaucratic records are among the most prevalent and useful types of "found" data. As areas of Syria change hands from opposition groups to regime control and back again, what happens to government office buildings? What happens to police headquarters? Are records secured and preserved or lost and destroyed? We need to train those on the ground to inform them how valuable these data sources are and to develop, design, and implement strategies for preserving them.

References

[1] P. Ball, W. Betts, F. Scheuren, J. Dudukovic, and J. Asher. Killings and Refugee Flow in Kosovo, March – June, 1999 (A report to ICTY), 2002.

[2] Berghof Foundation. Berghof Glossary on Conflict Transformation: 20 Notions for Theory and Practice, 2012.

[3] A. R. Chapman and P. Ball. The Truth of Truth Commissions: Comparative Lessons from Haiti, South Africa, and Guatemala. Human Rights Quarterly, 23:1–43, 2001.

[4] F. Checchi and L. Roberts. Interpreting and Using Mortality Data in Humanitarian Emergen- cies: A Primer for Non-Epidemiologists, 2005.

[5] A. Gohdes. Different Convenience Samples, Different Stories: The Case of Sierra Leone, 2010.
[6] D.R. Guzman, T. Guberek, Shapiro G.M., and P. Zador. Studying Millions of Rescued Documents: Sampling Plan at the Guatemalan National Police Archive (GNPA). In Proceedings of the Joint Statistical Meeting, Survey Research Methods Section. American Statistical Association, 2009.

[7] D. Guzman. Speaking Stats to Justice: Expert Testimony in a Guatemalan Human Rights Trial Based on Statistical Sampling. CHANCE, 24, 2011.

[8] A. Hoover Green. Collective Violence and International Criminal Justice: An Interdisciplinary Approach, chapter Learning the Hard Way at the ICTY: Statistical Evidence of Human Rights Violations in an Adversarial Information Environment. Intersentia, Antwerp, 2010.

[9] J. Klingner and R. Silva. Counting Civilian Casualties, chapter Combining Found Data and Surveys to Measure Conflict Mortality. Oxford University Press, Oxford, 2013.

[10] Z. N. Mneimneh. Interview Privacy and Social Conformity Effects on Socially Desirable Reporting Behavior: Importance of Cultural, Individual, Question, Design and Implementation Factors, 2012.

[11] M. Parlevliet. Considering the Truth, Dealing with a Legacy of Gross Human Rights Violations. Netherlands Quarterly of Human Rights, 142, 1998.

[12] J.W. Pennebaker and B. L. Banasik. Collective Memory of Political Events: Social Psychological Perspectives, chapter On the Creation and Maintenance of Collective Memories: History as Social Psychology. 1997.

[13] M. Price, T. Guberek, D.R. Guzmán, , and P.and Shapiro G.M. Zador. A Statistical Analysis of the Guatemalan National Police Archive: Searching for Documentation of Human Rights Abuses. In Proceedings of the Joint Statistical Meeting, Survey Research Methods Section. American Statistical Association, 2009.

[14] A. Sachs. Experiments with Truth: Transitional Justice and the Processes of Truth and Rec- onciliation, chapter Different Kinds of Truth: The South African Truth and Reconciliation Commission. Hatje Cantz, Ostfildern-Ruit, 2002.

[15] G.M. Shapiro, D.R. Guzman, P. Zador, T. Guberek, M. Price, and K. Lum. Weighting for the Guatemalan National Police Archive Sample: Unusual Challenges and Problems. In Proceedings of the Joint Statistical Meeting, Survey Research Methods Section. American Statistical Association, 2009.

[16] R. Silva, J. Klingner, and S. Weikart. State Coordinated Violence in Chad under Hissene Habre: A Statistical Analysis of Reported Prison Mortality in Chad's DDS Prisons and Command Responsibility of Hissene Habre, 1982-1990, 2010.

[17] United Nations Secretary General. Guidance note of the secretary-general: United nations approach to transitional justice, 2010.

[18] M. Utas. Victimcy, Girlfriending, Soldiering: Tactic Agency in a Young Woman's Social Navigation of the Liberian War Zone. Anthropological Quarterly, 78, 2005.

About HRDAG

The Human Rights Data Analysis Group is a non-profit, non-partisan organization that applies scientific methods to the analysis of human rights violations around the world. This work began in 1991 when Patrick Ball began developing databases for human rights groups in El Salvador. HRDAG grew at the American Association for the Advancement of Science from 1994–2003, and at the Benetech Initiative from 2003–2013. In February 2013, HRDAG became an independent organization based in San Francisco, California; contact details and more information is available on HRDAG's website (https://hrdag.org) and Facebook page.

The materials contained herein represent the opinions of the authors and editors and should not be construed to be the view of HRDAG, any of HRDAG's constituent projects, the HRDAG Board of Advisers, the donors to HRDAG or to this project. This memo was requested and funded by IREX on behalf of the Syrian Justice and Accountability Centre.

About The Syria Justice and Accountability Centre

The Syria Justice and Accountability Centre (SJAC) is a non-profit, multilaterally-supported organization that envisions a Syria where people live in a state defined by justice, respect for human rights and rule of law.

The SJAC is collecting, preserving and analysing information on human rights violations and other relevant data to contribute to and inform a transitional justice process for Syria. Employing an unbiased and non-partisan approach, the SJAC also builds and maintains close relationships and partnerships with Syrian individuals, organizations, communities and international actors working towards justice and accountability for all Syrians.