Human Rights Researchers’ Data Analysis and Management Practices

Lu Xiao, Isioma Elueze
Faculty of Information & Media Studies
The University of Western Ontario
London, Canada N6A 5B7
{l Xiao24, ielueze}@uwo.ca

Jillian R. Kavanaugh
Center on Media and Child Health
Boston Children’s Hospital
Boston, USA 02115
jill.kavanaugh@childrens.harvard.edu

ABSTRACT
The impetus to assist human rights researchers in data analysis is stronger than ever; however, little is known in the literature on human rights researchers’ practices in collecting, managing, and analyzing their research data. In an attempt to address this gap, we interviewed human rights researchers and conducted an online questionnaire to understand the characteristics of the data they analyze, as well as their data analysis and management practices, such as their experiences with data analysis software programs. We also explored their expectations with respect to a qualitative data analysis (QDA) software program.

Keywords
Human rights research, data analysis practices, qualitative data analysis, QDA tool development

INTRODUCTION
Human rights researchers use a wide range of data analysis techniques, from statistical analysis and modeling to examining correlations among different factors of human rights violation and to studying the impact of human rights violation events ethnographically. Researchers also interact with relevant data in various formats (e.g., textual, audio, and video) and conduct both qualitative and quantitative content analyses. These data are increasingly accessible as Internet-based resources. Prominent examples include the Center for Human Rights Documentation and Research at Columbia University, the Human Rights Documentation Initiative (HRDI) at the University of Texas at Austin, and the University of Connecticut Human Rights film collection.

While these trends in the generation and access of human rights violation data offer more data analysis opportunities to human rights researchers, they also present challenges in the qualitative data analysis because of the ever-growing size of the data resources. Computer scientists and digital humanities scholars have started to explore computational approaches to address these challenges (e.g., Miller, Li, Shrestha, and Umapathy, 2013). However, there is limited understanding about human rights researchers’ data analysis practices.

Addressing this gap, we surveyed human rights researchers through interviews and online questionnaires to understand the characteristics of the data they analyze, their data analysis and management practices such as their experiences with data analysis software programs. We also explored their expectations with respect to a qualitative data analysis software program.

RELATED WORK

Data Analysis Methodologies in Human Rights Research

In human rights research, human rights indicators are often the basis for both governmental and non-governmental agencies to measure human rights violations. In 1990, when determining the global need for human rights indicators, the United Nations acknowledged the importance of the addition of qualitative data to pre-existing quantitative data analysis practices in economic, social, and cultural human rights. According to the Progress Report, the qualitative elements of economic, social and cultural rights are as crucial as the quantitative statistics (Türk, 1990). In Green (2001), the author reviews the history of human rights indicators (including Türk’s report). While Green states that the emphasis is on the numerical data, definitions provided from United Nations, various NGOs, and several government agencies have historically called for the essential inclusion of qualitative data analysis when developing and evaluating human rights indicators.

While quantitative data analysis of human rights research, specifically in the area of indicator development, may be
featured more prominently in the literature, qualitative analysis is crucial, and exclusion of this type of analysis can lead to issues. According to Rosga and Satterthwaite (2009 p. 4), the reliance on quantitative data in both the identification and measurement of human rights can lead to “tests of measurability [that] often prevail over accurate and contextually sensitive assessments of substance or actions.” Another problem that can affect both quantitative and qualitative data analysis is the problem of “information effects,” defined by Clark and Sinkkink (2013, p. 540) as “patterns in the data stemming from the process of information collection and interpretation, rather than the process that actually gives rise to human rights violations or their mitigation.”

Human rights researchers also reportedly used qualitative data analysis methods (e.g., Safari, 2003; Drew et al., 2011; Scorgie et al., 2013). For example, Drew et al. (2011) used the established WHO QualityRights Assessment Tool to determine human rights violations in mental health facilities. This tool, which provides a series of interview questions and observation guidelines for those assessing the practices within a mental health facility, encourages documentation and discussion of qualitative findings, and includes space for such analysis within the tool templates.

Bakewell (2007) identified that the biggest remaining obstacle to the use of qualitative data by development organizations such as NGOs is the lack of appropriate cost-effective methods for its analysis. Similarly, Madden and Ross (2009), who critiqued qualitative and quantitative methodologies for assessing crimes of violence, expressed that even the publicly available images on Google Earth and Virtual Earth are helpful in providing quantitative data that support qualitative data analysis of human rights violations and mass atrocity narratives. They also posited that within a broad geographic area, the quantitative numbers would support the testimonies of personal narratives to empower displaced persons in courts of international justice.

The Importance of Understanding Researchers’ Data Analysis Practices

Prior studies about the data analysis practices in academic communities acknowledged several important issues in researchers’ data analysis processes, such as the need of using mixed analytical approaches (Schillerup, 2008) and transcribing, coding and interpreting the collected data (Merritt & Labbo, 2004; Whiffin, Bailey, Ellis-Hill, & Jarrett, 2014), the preparatory work for archiving and sharing data (Cheshire, 2009), and the use of analytical aids like ATLAS/ti (Schillerup, 2008) and NVivo (Welsh, 2002). The number of researchers involved/collaborated in the data analysis processes also affects the data analysis practices (Sweeney, Greenwood, Williams, Wykes, & Rose, 2013).

Nonetheless, The research context has been recognized as a critical factor in the data analysis practices. Schillerup (2008) observed that for some researchers, procedures for data analysis refers primarily to the tasks of coding, indexing, sorting, retrieving, or otherwise manipulating data, while for some others data analysis involves only the work of interpretation. Although there have been a lot of studies about the researchers’ data analysis and management practices, we have not found any study that examined this issue in the human rights research’s domain except for a few articles about the investigation of the research methodologies such as Madden and Ross’s (2009) work which we mentioned in the previous section.

RESEARCH DESIGN

Research Questions

To help us better understand the data analysis and management practices in human rights research, we sought to answer the question: what are the human rights researchers’ data analysis practices? To shed more light into the design of qualitative data analysis software programs that aim at improving the human rights researchers’ data analysis practices, we also asked: what are the human rights researchers’ expectations and concerns on qualitative data analysis programs? To answer these research questions, we adopted both interview and questionnaire methods.

Interview

Our interview had four sections. The first section was about the interviewees’ research background and the data that they have worked with in the research. For example, we asked interviewees to describe the data they often analyze, such as the format (photos, text, video, etc.), the size, the type (primary or secondary data), and the source (publicly available or private). In the second section, we asked the interviewees to think of a concrete example of data analysis in their projects and describe the analysis process. We also asked them to provide details like what they were looking for in the data, whether they used software programs for analysis, and if so, the name of the programs, and the software program features that they found most useful in the analysis. Interested in how researchers coordinate on shared data during the analysis process, we asked them how many people are generally involved in data analysis projects, and if there are indeed multiple analysts, we asked how they coordinate with each other in the process of analyzing data and integrating the results. In the last section, we sought their expectations on a software program that would support qualitative analysis of human rights data, either small or large scale. We asked for concerns or issues that should be addressed, three most important design features desired in such a program, and features they considered ineffective and should be excluded (based on their experiences).

Through purposive and snowball sampling techniques, we interviewed 13 North American academic and non-academic researchers of human rights violation research. For example, we emailed the authors of the reviewed papers who have either Canadian or American contact information, and emailed the researchers whom we know have been
doing human rights violation research. Some interviewees provided additional contacts. All interviews were conducted in a semi-structured format. The interviews were 60 – 90 minutes long except one interview which lasted 25 minutes because the interviewee’s background and experiences were different from what we were looking for. One interview was via phone, one via Skype, and the rest were conducted face-to-face at locations of the interviewees’ choices (e.g., their offices). The first author and a research assistant were both present in the interviews taking the roles of asking questions and taking notes. All interviews were audio recorded. Interviewees were compensated CAD $25 or USD $25 for their time. One researcher kindly refused to accept the compensation after being interviewed, which we understood as a voluntary support of our study.

We performed a thematic analysis with the interview transcripts using open coding process. The interview design, i.e., the four interview sections, provided an initial structure in our first reading, e.g., the data format, data sources, etc. We then identified more specific reoccurring themes within and across individual transcripts, and searched for repetitions within and across narratives and field notes (Ryan & Bernard, 2003). Once the salient themes had been identified, we selected dialogue from interviews that supported the major themes we had identified. For example, to support the idea that terminology differs across disciplines, we included quotes that highlighted the contrary definition of the term coding.

Online Questionnaire

Our questionnaire had 19 questions that reflected the four sections of the interview: background, data, analysis, and QDA software programs.

We used several resources to reach out to potential respondents. First we sent the email invitation (with the web link to the online questionnaire) to all authors who have articles published in International Journal of Human Rights (publications years 2010 – March 2013), Human Rights Law Review (2005 - March 2013), Human Rights Review (2000 – June 2013), Journal of Human Rights (2010 – April 2013), and Journal of Human Rights Practice (2009 - March 2013). Due to a limited number of responses we received from this approach, we contacted non-profit organizations in human rights research or human rights activities. One major resource we used to identify these organizations is the Human Rights Web Archive, an initiative by Columbia University Libraries and Information Services (CUL), and its Center for Human Rights Documentation & Research (CHRDR). We selected the organizations that communicate in English or French (as our questionnaire has English and French versions) and sent our email invitation with the web link.

As indicated in the information letter (that is the first page of the online questionnaire), we donated CADS$2 to United Way for each completed questionnaire. In total, we sent out 772 email invitations (this did not include two reminder emails for each non-responded invitation) and received 145 responses. We discarded 3 that did not consent to the study (as no data was provided), and 29 that gave their consent to the study but did not respond to any questionnaire item. Of the 118 responses, 73 were complete.

FINDINGS

Human Rights Researchers’ Academic Background and Experiences

Our interview and questionnaire results illustrate that researchers working on human rights related issues have diverse academic backgrounds. The thirteen interviewees, including two non-academic researchers, covered eight disciplines such as political science, statistics, and sociology. Among our questionnaire respondents, 54% are academic researchers, 18% non-academic researchers, and 27% selected “other” and provided their profession (e.g., lawyer, independent consultant, and program director). The respondents represented close to 30 fields, with the most common fields being law, political science, international studies, sociology, and anthropology.

The researchers we interviewed have a wide range of experience, ranging from a few years to over 35 years. When asking for this information in the questionnaire, we provided several options from “0-5 years” to “20+ years”. Median is 6-10 years. 18 respondents had 11-15 years of experience, and 14 respondents had over 20 years of experience. We consider the questionnaire respondents to be modestly experienced in general in performing human rights research.

Data Characteristics in Human Rights Research

All but one interviewee has dealt with primary data, with some being publicly accessible, some required special access. These data varied in formats and include text, number, audio/video, and artifact (e.g., images, museum objects). They came from various sources, such as police records, court decisions, U.S. Department of State Human Rights Reports, and interviews conducted by the researchers themselves. The size of the data in the qualitative analysis is within hundreds of pages (e.g., 40 interviews, documents of 20 countries, etc.).

Our questionnaire result is somewhat similar, whereby out of 77 respondents, 48% reported to have used both primary and secondary data, 26% used primary data only, and 25% only used secondary data. About 56.8% of the respondents reported to have used public data, 30.5% have used private data, and 12.7% have used semi-public data. Respondents used text, images, videos, audios, and artifacts. Some provided additional details about the data, e.g., discussions, focus group or interviews, site visits, and even participatory research where victims doing their own research.

Hafner-Burton and Ron (2009) posited that most scholars of human rights draw on two key sources of cross-national data: the Political Terror Scale (PTS) and the Cingranelli and Richards Index (CIRI), noting that they are the most complete cross-national sources for violations of personal
integrity rights data for many countries. However, our questionnaire results show that there was only one response that noted the use of PTS datasets. In fact one respondent explained that he/she would not use PTS because their processes are not reliable and have validity issues. Additionally, the results suggest that the researchers used various data resources. 51 of 78 respondents (65%) used data from Amnesty International, 39 of 78 respondents (50%) used private interviews, and 27 of 78 respondents (34%) used data from the U.S. Department of State Country Reports on Human Rights Practices. Respondents who used publicly available interviews were 26 of 78 (33%), while 19 of 78 respondents (24%) used data from Truth and Reconciliation Commission Reports. Also, 16 of 78 respondents (20%) used CIRI Human Rights Dataset (20%), and 6 respondents out of the 78 who responded to this question (7.6%) used The WomanStats Project. In addition to the cases of data sources that we provided in the questionnaire, respondents were also given the option of listing ‘other’ data sources that they use for their human rights research projects. From this, 53 of 78 respondents (68%) noted that they used other data sources, and listed examples such as - academic literature, fact descriptions in case law, journal articles; HRW reports; legal decisions, government statistics, media reports, NGO reports, newspapers and newspaper articles, budgets, world bank reports, social media, email communications, in person interviews, research missions, country reports outside US, UN reports, shadow reports, on-line data bases assembled by the World Bank, IMF, and the UN, data bases on grave violations against children’s rights, court records and government statistics (where available), survey data, on-line administrative data, research and news publications, e.t.c. A respondent also noted that they draw on research that comes in through the equality and human rights helpline and Citizens advice Bureau, which helps them to monitor emerging trends, while another remarked that they used data from Anti-bullying Alliance, Article 19, British and Irish Legal Information Institute, British Institute of Human Rights, Care Quality Commission, Children’s Commissioner, Crown Prosecution Service, Human Rights Watch, Joseph Rowntree Foundation, Justice, Law Commission, Law Society, Legal Action Group, LexisNexis, Local Government Ombudsman, Low Pay Commission, National Housing Federation, Nursing Standard, Open Democracy, UK Human Rights Blog. We are Spartacus, Westlaw etc. These ‘other’ sources were categorized into academic literature or publications, media or news sources, international institution data (e.g. World Bank, UN), legal documentations and proceedings, databases, and government publications.

Original data collected by the researcher was another common response that our questionnaire respondents stated they used for their data analysis human rights projects, which were not listed in our instrument.

With respect to the size of the data, we asked “How many items do you analyze in a project? Please select all that apply”. Out of the 79 researchers who responded to this question, 34 (43%) reported that they analyzed less than 50 items in a single project, 25 (31.6%) analyzed between 50 to 100 items in a project, 12 people (15.2%) analyzed 101 – 500 items, 8 people (10.1%) analyzed between 501 to 1000, and 10 (12.6%) analyzed more than 1000 items in a project. Note here that instead of the term document, we used item because we expected that not all researchers analyzed documents in the project. For example, some may analyze artifacts such as images or videos, or some may analyze numerical data. Some researchers specified the items used in their projects. These include cases, interviews, videos, reports, survey responses, judgments or decisions, information on a violation that happened or is alleged to have happened, indicators such as the percentage of children that are malnourished, and additional specific responses. In our next questionnaire item, we asked the respondents to define what an item meant in their study. The respondents gave varying definitions. To some, data items are a primary document or a video. For others, data items are “cases” or “index value per year and country.”

One respondent noted that, “the ‘item’ for me is every evidence or indication that can provide information on the question of participation of Congolese civil society to mechanisms of human rights at local, national and international level.” Another researcher stated that items are predominantly aggregates of human rights measures, with datasets often having more than 3000 data points to them. Furthermore, in response to what defines an item some respondents expressed that even the term ‘item’ was ambiguous so they could not specify (e.g., “it could be any piece of information” one respondent noted,”), some noted that items in they analyzed in their researches included cases, academic analysis, NGO reports, government reports and human rights organisation reports.

We asked the researchers to indicate whether they analyzed primary data, secondary data, or both primary and secondary data in their human rights research projects. 21 of 77 respondents (27.3%) noted that they analyzed primary data only, while 19 (24.7%) analyzed only secondary data. 37 (48%) respondents analyzed both primary and secondary data in their research projects.

We also asked the researchers to select if they considered their data sources to be public or private. From the response to this question, 67 of 77 respondents (87%) selected that their data sources were private, 36 of 77 respondents (46.8%) selected public, and 15 of 77 respondents (19.5%) selected semi-public. Table 1 shows the examples of data sources reported by the respondents, which they categorized into public, private and semi-public.

<table>
<thead>
<tr>
<th>Public data sources</th>
<th>Private data sources</th>
<th>Semi-public data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online availability</td>
<td>Interviews and witness testimonials (individual)</td>
<td>Academic literature/publications</td>
</tr>
</tbody>
</table>
A variety of software programs have been developed to facilitate the qualitative data analysis process. Both the interview and questionnaire results suggest that the use of qualitative data analysis software is not common in this particular research community. The interviewees acknowledged the lengthy process of reading, annotating and analyzing the textual data. For example, one interviewee said that he/she had to read all of a specific country’s reports to judge the country’s human rights violation situation based on the reading. Despite acknowledging the time required for these efforts, no interviewee used any QDA software (e.g., NVivo) to facilitate the process. In fact, many interviewees were not even familiar with QDA software. Instead, they used traditional and low-technical methods to analyze textual data, such as reading through PDFs, Word documents, or working almost exclusively with print documents. One interviewee explained that he/she used Word processing and spreadsheet software to store excerpts that had been extracted from the textual documents. Several interviewees created their own databases for storing and accessing their documents, and one interviewee created their own open source software program for researchers to share audio/video interviews within the research team.

In the questionnaire results, 69 of 74 respondents expressed that they used qualitative data analysis for their projects. Yet only 21 of these respondents (28.4%) acknowledged that they used software programs in analyzing their qualitative data. The researchers gave varied reasons for not using the qualitative analysis software programs. Majority of the respondents (31.4%) reported that were not aware such software existed, 11 of 51 people (21.6%) selected that they were aware that such software existed, but did not have the time or resources to learn the program, while 6 people (11.7%) recorded they preferred analyzing documents in print. In addition 18 of 51 people selected ’other’ reasons for not using qualitative data analysis tools. These ‘other’ reasons given were that they were either unaware of its existence, saw no value in its use, or that their analysis required statistical software.

Eight researchers who used QDA software used NVivo, accounting for 47%. Two researchers used OpenEvsys and Martus Bulletin System (MBS) respectively; one used CDS/ISIS, Palantir, and Stories respectively, while eight researchers used other software programs. The other qualitative analysis programs listed by the researchers were ATLAS.ti, Excel’s non-statistical functions, R, and a respondent noted that they developed their own qualitative analysis tool.

The interview study suggested that a collaborative or cooperative data analysis process is common. One reoccurring theme is that the data analysis process is staged from judging/filtering textual data to synthesizing and publishing the results, and multiple people are involved in different stages. Also, those who perform the initial reading/judging/filtering step are often undergraduate or graduate students as well as volunteers with certain background that meets the need of the research (e.g., cultural or language background).

Several interviewees used collaborative software programs such as Dropbox and Basecamp to aid the sharing and management of their data or documents for data analysis. One interviewee whose work was done mainly through ethnographic studies explained that he/she used Facebook to communicate with people from the studied countries and the analytical process was “a collaborative interpretation (process)”. Another interviewee discussed how the online project management software Basecamp was used for collaborating their interviews of victims of human rights violations. In this project, there were multiple working groups performing these interviews and the interviewers’ reflections were highly regarded. The interviewee said, “Building a reflexology into methodology is so crucial, especially in this project based research”. As a result, every interviewer and videographer wrote a reflection blog within 24 hours of interviewing an individual. These reflection blogs were stored on the research team’s private Basecamp site, and were shared among the research team (which

<table>
<thead>
<tr>
<th>Data Analysis Practices in Human Rights Research</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Analysis Software Programs</td>
<td>The interviewees reported the use of quantitative and qualitative analysis techniques, with the majority of the interviewees focused on the qualitative analysis approach; however, for the questionnaire, about 47% of the 74 researchers who responded to this question reported that they used both qualitative and quantitative approaches in analyzing their human rights violation data, while 5 of 74 respondents (6.7%) and 34 of 74 (46%) researchers used only statistical and qualitative methods respectively.</td>
</tr>
</tbody>
</table>

### Data Analysis Processes – Collaborative or Individual

The interview study suggested that a collaborative or cooperative data analysis process is common. One reoccurring theme is that the data analysis process is staged from judging/filtering textual data to synthesizing and publishing the results, and multiple people are involved in different stages. Also, those who perform the initial reading/judging/filtering step are often undergraduate or graduate students as well as volunteers with certain background that meets the need of the research (e.g., cultural or language background).

Several interviewees used collaborative software programs such as Dropbox and Basecamp to aid the sharing and management of their data or documents for data analysis. One interviewee whose work was done mainly through ethnographic studies explained that he/she used Facebook to communicate with people from the studied countries and the analytical process was “a collaborative interpretation (process)”. Another interviewee discussed how the online project management software Basecamp was used for collaborating their interviews of victims of human rights violations. In this project, there were multiple working groups performing these interviews and the interviewers’ reflections were highly regarded. The interviewee said, “Building a reflexology into methodology is so crucial, especially in this project based research”. As a result, every interviewer and videographer wrote a reflection blog within 24 hours of interviewing an individual. These reflection blogs were stored on the research team’s private Basecamp site, and were shared among the research team (which

<table>
<thead>
<tr>
<th>Categories of data sources reported by respondent</th>
<th>Table 1. Categories of data sources reported by respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open datasets (government and other institutions)</td>
<td>Subscription</td>
</tr>
<tr>
<td>Judgments/legal decisions</td>
<td>Researcher rights/ownership</td>
</tr>
<tr>
<td>Government and institution-published data (UN, World Bank, IMF)</td>
<td>Unpublished/internal information</td>
</tr>
<tr>
<td>Published literature/articles</td>
<td>Personal collection</td>
</tr>
<tr>
<td>Public reports (including news sources)</td>
<td>Government publications</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Categories of data sources reported by respondent</th>
<th>Table 1. Categories of data sources reported by respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open datasets (government and other institutions)</td>
<td>Subscription</td>
</tr>
<tr>
<td>Judgments/legal decisions</td>
<td>Researcher rights/ownership</td>
</tr>
<tr>
<td>Government and institution-published data (UN, World Bank, IMF)</td>
<td>Unpublished/internal information</td>
</tr>
<tr>
<td>Published literature/articles</td>
<td>Personal collection</td>
</tr>
<tr>
<td>Public reports (including news sources)</td>
<td>Government publications</td>
</tr>
</tbody>
</table>
included several hundred members in this particularly large project).

When asked “who analyzes the dataset in qualitative/non-statistical analysis of human rights violation data?” in the questionnaire, 35 of 73 respondents (48%) reported that they did the qualitative analysis of human rights violation data by themselves. 35 noted that they analyzed the data with others, and to illustrate who the “others” were in their projects, they listed colleague/peer/co-researcher, "team" (unspecified), research assistants, respondents, and organizational staff as exemplification of people that they analyzed their qualitative data with. Specifically, eight of thirty-five people (22.9%) reported colleagues, 7 people (20%) entered research assistants, 6 people (17.1%) noted research team members, and 4 of 35 (11.4%) noted co-authors.

**DISCUSSION**

**Primary Data or Secondary Data?**

Interestingly, we learned from both the interviews and questionnaire results that a researcher’s understanding of the difference between primary and secondary data could be obscure. We speculate that this is due to the wide variety of primary and secondary data that the researchers deal with in their practices. Working on an Inter-American course project, one interviewee examines advisory opinions and supervisory decisions by analyzing government archives and the archives of human rights organizations, in addition to the oral history life stories of human rights activists. We consider these data as primary data since they were firsthand records of advisory opinions and supervisory decisions. In a different project, the interviewee analyzed the Columbian newspapers published in the last fifty years in order to understand patterns of violence. These newspapers appear to be secondary data to us for the research focus, as they are not from direct observers of the violence. Nonetheless, the interviewee referred to all of the data he/she has worked with as primary data. Another interviewee analyzes the human rights reports from U.S. Department of State (<http://www.state.gov/j/drl/rls/hrrpt/>) which are secondary data to us, but when the interviewer asked, “is it fair for say that your data are mainly secondary data?”, the interviewee said, “I don’t know… it’s something I struggled with you know. …we do quite a bit of reliability testing to make sure that the coding guidelines are tight. See that’s why I don’t know if that’s secondary…we’re using secondary information, but the measure is very original…” Two other interviewees who also do statistical modeling with human rights violation data worked with primary data such as court reports, surveillance cards, and victim interviews. To convert the textual primary data into numeric data for statistics, they also recruited people to read through these primary data and enter certain information to a database based on a form. The information is then converted to numeric scale, and they use the information for quantitative analysis and statistical modeling.

The unclear perception of the idea of primary and/or secondary data was also evident in the questionnaire data. In the examples of primary and secondary data that the respondents gave, similar examples were identified as both primary and secondary sources of data. For instance, the examples of primary data listed by the respondents include court cases, judgments, case reports, UN reports, case law judgments, interviews with people who have had their rights compromised, legal documents, victim and witness testimony, health records. And the listed secondary data are government reports, human rights NGO reports, Political Terror Scale and Cingranelli-Richards datasets, reports of other organizations, books, journal articles, newspapers, field reports, legislative review, incident reports, transcripts, legal literature, interviews with present and former clients, legal aid lawyers, judges, and court officials, existing data sets, demographic data, HR reports, legal documents, etc.

As shown from these two lists, there is vagueness and inconsistency in what the researchers consider to be primary and secondary sources of data, and this is noteworthy.

Another interesting point related to this issue arose from the interview with a researcher who researches how countries treat poverty issues through studying constitutions. Despite that his/her main approach is quantitative methods with secondary data or public documentations, she/he commented, “I do know that our biggest frustration is that getting data, getting information from those places…Really for people who work in poverty getting disaggregated data, very on the ground data, instead of just averages, is crucially important”.

**Terms that Cross or Do Not Cross Disciplines**

One thing we learned from this interview study is that it is important to consider whether certain terminologies cross or do not cross disciplines when doing research about the research practices, including data collection and analysis techniques, of other researchers. Some terms such as “coding” and “qualitative” may not cross disciplines because there are inconsistent understandings of a term’s meaning. We found that although the interviewer and the interviewee were talking about the coding process, it actually could mean something different to the interviewee. In an interview with a political science professor, coding meant judging the severity of a country’s human rights violation situation based on the textual data, and giving a numeric code to rank the country (e.g., in the coding scale of 1 to 3, 1 could mean no concern on human rights violation, 2 some concerns/issues, and 3 severe situation in the country). We did not realize this different interpretation of coding until the interviewee showed us the coding schema and the related instruction guide. They are essentially evaluation criteria of different aspects of human rights violation and explanations of what each numeric code means. With this documentation the evaluators (the interviewee referred to them as coders) hopefully have shared understanding on the meaning of each numeric code.
and their results are comparable then (e.g., the rating of 3 should mean the same level of severity among the evaluators). Whereas in our interview with a history professor who analyzes news articles and interviews, coding meant open-theme reading and interpreting textual data and extracting segments as evidence to support the interpretation.

Also, although the qualitative analysis is similar to thematic analysis according to our knowledge of research methods, the term “thematic analysis” does not seem to cross disciplines well. One interviewee explained how he/she reads the course case documents and generate themes to interpret his/her readings but did not refer to this as thematic analysis – “I take the leading principles from these cases, and I would then summarize where the law is going by reading from these documents...So it’s an intuitive reasoning based on all my experiences...” The different interpretation of coding between the interviewer and the interviewee was also captured in the same interview: at one point, the interviewer tried to make the connections between the qualitative analysis that she is familiar with and this interviewee’s work, “if I come up with this (the results), I will need to follow some content analysis, or coding analysis...” and the interviewee responded, “we don’t do that". Further discussion revealed that the interviewee was only considering coding in the quantitative sense. When the interviewer explained the term thematic analysis, the interviewee acknowledged that it was very similar to what he/she has practiced in the research.

Similarly, not every discipline refers to qualitative analysis by using the word “qualitative”, even though that is the kind of analysis they are performing. One interview showed the interviewee’s awareness of this terminology issue across disciplines. When we asked an interviewee if he/she has done any work with qualitative analysis approach, the interviewee started with a description of a paper co-authored with his/her colleagues, and then paused, “I am not sure if this would count...Can you give me a definition of what you mean by qualitative? " We responded by asking how his/her research group generated the previously mentioned database based on written documents – constitutions. It turned out that it was an example of qualitative analysis: he/she had a “coding sheet” that detailed what kind of provisions can be considered as enforceable laws and what cannot, and he/she and the graduate student would then read the constitutions and identify and judge the provisions.

**Awareness of Data Availability, Data Pre-Processing, and Data Access**

Because of the sensitivity of their type of research, we were curious if there were additional requirements besides the ethics protocol that the researchers needed to satisfy in order to access the data, and if so, what they were. From the respondents’ answers, we identified the following requirements: confidentiality agreements; registration – e.g. “like CIRI dataset, I had to register and log in order to have an access”; general human-subject protocol, permissions to quote/limit of direct quotation; "official approval"; and legal adherence. However, many of the respondents noted that this question was not applicable or there were no specific requirements, such as “no barriers to data access”, and “no, as I am using publicly available data, there are no further restrictions”.

One interviewee who worked extensively in the process of obtaining primary human rights violation data commented that it was very important for the researchers to realize that the human rights violation data was often affected by hidden factors, like the witness’ availability and background, and whether the victim reports were filtered by the agency that took the victim’s statements. We were curious whether the researchers were aware of such processes, so we asked “ Some pre-collected data have been processed before being available to researchers (e.g. certain data are stripped for privacy concerns). Are you aware of such data pre-processing protocols in your research? How was this impacted your research? Please explain". A majority of the respondents (16 of 31) noted that they were not aware of any such preprocessing. While 12 respondents acknowledged their awareness of such processes, they expressed that this had no impact on their research. 13 of the 31 respondents said that they were aware and that it did have some impact on their research, for instance “I am aware that this happens, but I have not been aware of what has been stripped from which data sets. This means I haven't adjusted for these processes". And such processes could affect the selection of datasets and databases, as one respondent explains, “It makes me a much more critical user of some data sets. For example, I would not use PTS or Freedom House as their processes are not reliable and have validity issues as well", and/or make it difficult to find certain information.

**User-Centered QDA Software Development**

Our study also contributes to the development of the software programs that support such processes. As Peters and Wester (2007) stated, in spite of the growing attention for qualitative analysis, there is a problematic link between procedures of qualitative analysis and software programs that support it. To help novice qualitative researchers understand what tool to use for which stage of analysis, Peters and Wester (2007) emphasized the importance of providing specific and detailed instructions to link computer tools to research methodology in order to help researchers better understand the tools and support their research. They also advocate that more effort should be put into illustrating how computer programs may support the methodological quality of procedures used in various approaches of qualitative data analysis.

Taking a user-centered development perspective, it is critical to elicit the design requirements of such software programs through requirements analysis. One important component that helps designers construct the design requirements is the understanding of the existing practices
in qualitative data analysis and the important factors. There are studies that help us these aspects. For example, to understand the meanings attached to qualitative research practice and the perceived challenges posed by contemporary innovations in data management, access, and analysis through electronic archives, Broom, Cheshire and Emmison (2009) conducted six focus groups with 37 Australian qualitative researchers and revealed that the researcher has a special relationship with the data which prevents anyone else from analyzing them in their original context and which leaves the data ‘dismembered’ when archived. They also identified other important issues such as the concerns over research ethics and data ownership.

The literature in Qualitative Data Analysis (QDA) software has also identified important design requirements. For instance, one major concern is the potential loss of data, due to either technological obsolescence or technological failure. Cliggett (2013) pointed out that with advances in technology, QDA software programs must have “robust exporting options for durable formats” (Cliggett, 2013, p. 7). The author suggested that an important feature to have in QDA software is the ability to anonymize individuals and key identifiers while preserving the context and relationships, and the software must strive to preserve metadata in a variety of file formats, so future researchers using the same data can consistently “identify relationships between interviewees, images and spatial data, or regional relationships such as village residence and agricultural fields identified in maps, or topical themes addressed in different data types, such as interviews and field notes.” (Cliggett, 2013, p. 8). This emphasis on assisting future qualitative human rights researchers is also highlighted by Wesley (2014).

Prior studies have identified important analytic tasks that occur in qualitative research. Gilbert (2014) identifies four broad steps, with more specific analytic tasks, that are essential to the qualitative data analysis process: organizing data, exploring data, interpreting and reflecting, and integrating data. Gilbert also highlights the work of Lewins and Silver (2007), who devised a list of required QDA software functionalities based on the following analytic tasks: project planning and management; analytic memo writing; reading, marking and commenting on data; searching data; developing coding schemes as well as the actual coding, retrieval of coded sections, and recoding; data organization, linking, and mapping; searching the data and codes; and reporting features (Lewins and Silver, P. 9).

In our view, it is also important to understand the researchers’ experiences with the existing software programs in their qualitative data analysis, such as the features they disliked, the features that would have helped, etc. So we asked for the researchers’ experiences with the existing QDA software, their expectations of a new QDA program, as well as the concerns regarding the software.

**Experiences with the Existing QDA Software**

The researchers that responded to our questionnaire used the software to map relationships between different human rights violation (12 of 18), to compare data from many different sources (10 of 18), to organize and share human rights violation data with other (9 of 18), to collect and input human rights violation data (8 of 18), to analyze complex datasets (7 of 18) and to securely store human rights violation data (7 of 18). Also, 5 of them explained that they used it because it saved time and/or costs associated with pre-processing data. The researchers also provided other reasons for using the qualitative software programs, e.g., it helped them explore new areas of rights and procedures, to analyze my interview data, to reduce biases in during interpretation, or to share their data and results with their project partners and for presentations to their donors.

We also asked our respondents about the features that were available in the qualitative data analysis software programs they had used. Only 16 researchers responded to this question. The features that at least 8 researchers noted were: the ability to search through textual data by keyword matching, the ability to search by concepts, the ability to generate reports about the analysis results, the ability to sort analysis results by different criteria such as date, author, and source, and the ability to allow user to annotate the documents without tampering the original data. Features like the ability to search through annotations, to share analysis results with other users, to visualize analysis results, to search by sentiments, to identify trends and of violations with algorithms, and to save search histories were noted by a few researchers. Although it is expected that identifying victims and/or perpetrators from the data would be highly desired by the researchers, only 1 respondent noted that the qualitative data analysis software program provided such feature. In addition to the features we provided in the multiple-choice question, the respondents noted two additional features: to map to a geographic location, and to map against other kinetic and non-kinetic factors.

**Expectation of a New QDA Software Program**

With regards to the expectations of qualitative data analysis software, the interviewees presented a wide range of characteristics that would be useful for their research purposes. Two interviewees would like a program that would recognize patterns and meaning in the text being analyzed. One suggested, “create a more effective, searchable database with all of the documents held by the Inter-American Court, and again, that would prompt the user to look for certain types of patterns and certain types of information that could be organic, be changing. I wouldn't want it to just be a Google search of a bunch of documents”. When asked to define a “pattern,” the interviewee provided an example of how the court’s decisions often do not mention violence against women. For example, certain cases “may in fact be cases of violence against women, or they may in fact contain within them a
story about patterns of violence against women that are not being treated by the court as such, so the court is talking about this individual being disappeared or murdered.

Another interviewee was also interested in a program that would locate patterns, stating that such a program would “allow me to find patterns that I wouldn't be able to find otherwise, and also quickly, which is great, and of large interview collections”.

Interestingly, one interviewee would like to see a program that not only assists in the analysis of data, but also the management of data, in order to assist in the collaborative efforts of qualitative data analysis. The interviewee, who supervises students analyzing U.S. Department of State Human Rights Reports, stated, “one of the things we struggled with was how to have the students manage the massive amounts of data...and be able to find things quickly.” Other requested features included: offering multilanguage support, reducing the workload of collecting disaggregated survey data, enabling transcription of the audio files, and enabling adding and searching detailed metadata.

In the questionnaire, sixty-three people responded to the question concerning their expectations of a qualitative data analysis software program. Out of which they identified that it is important for the software program to have the ability to search through textual data by keyword matching (85.7%), to allow user to annotate the documents without tampering the original data (74.6%), to search by concepts (73%), to generate reports about the analysis results (66.7%), to visualize analysis results (66.7%), to sort analysis results by different criteria such as date, author, and source (66.7%), to search through annotations (61.9%), to identify trends and patterns of violations with algorithms (60.3%), to share analysis results (57.1%), to refine analysis results (57.1%), to share documents with other users (53.9%), to save search histories (49.2%), to relate self-generated results (46%), to search by sentiments (30.1%), and to identify victims and/or perpetrators with algorithms (12.7%).

From our questionnaire results, we see that there is a correspondence with the features of the qualitative data analysis tools that the researchers reported to be available in the programs that they have used, and the features that they think should be included in qualitative analysis programs. For instance, it appears that the ability to identify human rights violation victims was noted by only one researcher as a feature in the existing analysis software, and as an identified expected feature by the respondents it also had the smallest percentage of the responses compared to the other expected features. Another example is the ability to search through textual data. It was well recognized both as a function provided by the existing software, and as a feature that should be included in qualitative software programs. In addition, the ability for the software to be able to generate reports and search by concepts seem to be quite important too, as more than half of the respondents reported that they have used this feature, and noted that this was a feature that should be included in the new software. This is in accord with the three reasons with the highest frequencies given by the respondents for using a software program for qualitative data analysis - to map relationships between different human rights violations, to compare data from many different sources, and to organize and share human rights violation data with other individuals. The important features identified by the researchers should enable them carry out these tasks.

There are features that a majority of the respondents expected from the new software but were not noted in the existing programs, suggesting that such features are additional user requirements that have not been met in the existing programs. Examples of such features are: to search through annotations, to identify trends and patterns of violations with algorithms, to visualize analysis results, and to share the analysis results.

Concerns regarding the QDA software

In the questionnaire, we asked, “What are the concerns or issues that you think should be addressed in qualitative/non-statistical analysis software?” This was a multiple-choice item with eight options provided plus an “other” choice to allow for textual input. Our questionnaire results reveal that the respondents have the concerns about using the qualitative analysis software programs in these aspects (N = 62): security of the data (75.8%), that the tool might not properly interpret the nuances/context of non-statistical data (58%), data security during the sharing (51.6%), the navigational complexity and data storage (46.8%), the inflexibility of technology in terms of importing and exporting data files (38.7%). In addition, 17 of 62 people (27.4%) had concerns related to the user access (different levels of access for different, and the data storage (after completion of the project). The respondents also noted some other concerns including the ability to share their data with the people they are researching, the cost of software, ownership, privacy issues, reduction of investigator bias, reliability of analysis, replication and internal as well as external validity.

CONCLUSION

As the importance of reporting human rights violations by governmental and non-governmental organizations rises, there is a quick emergence of primary and secondary sources with rich textual data to offer additional details about human rights violation events. The growing interest of preserving human rights data and making them publicly accessible has also increased the access to human rights violation data. While these trends in the generation and access of human rights violation data offer more data analysis opportunities to human rights researchers, the increasing amount of data also presents a challenge in the data analysis process. This challenge is particularly severe in qualitative data analysis, as the researchers often need to identify, extract, and interpret meaningful or relevant information from the datasets. Conversely, the advancement
of computational technologies for identifying and visualizing patterns from large amounts of textual data has offered a new direction for addressing the above challenge; however, it demands more than just the computational algorithms to develop user friendly tools that support the human rights researchers’ data analysis processes. A thorough understanding of the researchers’ data analysis and management practices is required.

Recognizing that there is lack of such understanding in the literature, we interviewed human rights researchers of different disciplines and research areas to understand their data analysis and management practices. We also surveyed the tools that researchers have used or were aware of for facilitating the qualitative analysis process. Our results provide insights about the human rights research data – data sources, types of data, and format of the data, and data analysis and management practices – data access issues, analysis methods and processes, and the use of software programs to facilitate data analysis. We also probed the researchers’ experiences with the existing software for qualitative data analysis, their expectations and concerns regarding such a software program.

ACKNOWLEDGEMENT
We thank all the participants in this study. This study is funded by Social Sciences and Humanities Research Council of Canada (SSHRC) through its 2011 “Digging into Data” grant.

REFERENCES