Abstract

A growing number of datasets collected information reflecting the behavior and characteristics of contentious and violent organizations. Most of these datasets are not arranged so that they can be easily combined for analytical purposes. In particular, there is a lack of common identifiers for organizations. This creates a great deal of duplication of effort as well as many potential stumbling blocks for those who would like to combine information from different datasets for new analyses or research questions. We describe an existing effort to overcome common challenges and link organizations called the Terrorist Organization (TORG) crosswalk. We further propose a new effort, a Shared Contentious Actor Registry (SCAR), to allow for a more comprehensive and concerted effort to meet the challenges in linking data.

Keywords: conflict, contentious politics, terrorism, data, integration

Corresponding author: vasil@albany.edu

The authors are listed alphabetically, equal authorship implied.
Introduction

Recent research has highlighted the importance of agency and choices in contentious politics (e.g., Cederman, Girardin & Gleditsch, 2009). The growing body of quantitative data reflecting organizations involved in contentious behavior such as protest, insurgencies, and terrorism opens new avenues for research on agency in conflict. Researchers have examined conditions that make non-state actors more likely to combine conventional military attacks and terrorist attacks (Findley & Young, 2012), or transitions from non-violent protest to use of violence, and vice-versa (e.g., Lichbach, 1987; Moore, 1998; Regan & Norton, 2005). Yet, despite the theoretical emphasis on agency, many empirical analyses remain limited to aggregate country-level measures that disregard the specific organizations. For example, the Democratic League of Kosovo headed a non-violent campaign against Serb dominated rule, while the Kosovo Liberation Army chose a violent strategy. The two organizations had an antagonistic relationship, and it would be misleading to conflate the two or treat this as a case of an actor shifting strategies. We posit that the lack of attention to specific actors follows from the practical problems in identifying and integrating data on organizational behavior found in disparate datasets. We provide an overview of these challenges, and demonstrate how a current set of organizational connectors – the Terrorist Organization (TORG) crosswalk – can help overcome these. We then outline how the TORG can provide a starting point for a more comprehensive Shared Contentious Actor Registry (SCAR).

Linking actors in contentious politics data

We first review different structures in existing contentious politics data. The first category identifies conflict periods, with start and end dates, and participant information. UCDP/PRIO Armed Conflict Dataset identifies post-1945 conflicts with more than 25 battle deaths per calendar
year (see Gleditsch et al., 2002; Themnér & Wallensteen, 2014). The Nonviolent and Violent Campaigns and Outcomes (NAVCO) data track nonviolent campaigns where non-state actors make maximalist claims on the state since 1900 (Chenoweth & Lewis, 2013). A second category reports individual violent events. The Global Terrorism Database (GTD) tracks terrorist attacks by date and location since 1970, the groups associated with each attack, the targets, and related event data (LaFree & Dugan, 2007). Datasets such as the Social Conflict in Africa Database (SCAD) track an even broader range of events, including protests, riots, strikes, inter-communal conflict, and government violence against civilians (Salehyan et al., 2012). Finally, a third category focuses on the attributes of conflicts and contentious groups, including background on the actors involved, conflict termination, and links to ethnic groups (see Cunningham, Gleditsch & Salehyan, 2009; Kreutz, 2010; Wucherpfennig et al., 2012). The Big Allied and Dangerous (BAAD) dataset provide data on violent organizations, including links to other groups and states (Asal & Rethemeyer, 2008). Similarly, Minorities at Risk Organizational Behavior (MAROB) provides background on ethno-political organizations in the Middle East (Asal, Pate & Wilkenfeld, 2008).

Periodic, event and attribute data can all help support testing a wide range of theories on contentious politics, but answers to new questions often require data from multiple sources. For example, assessing the overlap between conventional civil war and terrorist attacks needs to go beyond whether events overlap space and time and also consider whether the same actors carry out the events (e.g., Findley & Young, 2012). Likewise, studying transitions from violence to non-violent actions (and vice-versa) should distinguish between whether we see the same actors making different choices over time or whether transitions at the conflict level reflect the emergence of new organizations engaging in different tactics (e.g., Moore, 1998). Such analyses are often very
difficult to carry out as significant effort will be required to align disparate data formats and structures.

Working with country-level data is relatively easy, given work on standardized lists of countries and identifying codes (see Gleditsch & Ward, 2001; Russett, Singer & Small, 1968), but we lack an analogous system for organizations. This creates many practical problems (and thus potential for errors) when attempting to assess theories of conflict focusing on agency or organizations. Of course, there may be many differences in what each project considers the ‘relevant’ actors given its theoretical focus, but even when the scope of projects overlap, determining alignment can be challenging. It is common for groups to change ‘official’ names, adopt nicknames (or be given them by outsiders), and both journalistic and scholarly sources often use different names to refer to the ‘same’ group. This can be as prosaic as the varied use of full names, abbreviations or acronyms, but is compounded by enormous diversities in language, translation and transliteration. Shifting factions, partnerships, and mergers are also common. Some sources identify organizations with multiple factions as umbrella groups, while others consider some sub-groups to be distinct (e.g., political and military wings). Different groups often have the same (or quite similar) names, and sources do not always clearly distinguish ambiguous cases. Finally, reporting on group activities and interactions is often inconsistent. Events are sometimes misattributed, and it is not uncommon for journalists and scholars to use ‘older’ names to refer to successor groups. In these ways and more, accurately identifying specific organizations can prove exceptionally challenging.
**Terrorist organizations (TORG) crosswalk**

The TORG crosswalk was created to facilitate the coding of terrorist and insurgent organizations, and as such, demonstrates that an initial unifying framework for contentious actors is feasible, and may serve well as a starting point for a more expansive Shared Contentious Actor Registry (SCAR). The current crosswalk (v. 2014 2.0) includes the organizations and identifiers found in the most recent versions of the following: BAAD, GTD, MAROB, and UCDP/PRIO, as well as Minorities at Risk (MAR), and the START Terrorist Organization Profiles (TOPs). Each dataset provides clues for triangulating group-to-group matches, including names, countries, periods of activity, etc. A major challenge is attempting to reconcile – or at least encompass – different inclusion criteria. The TORG attempts to unite actors from GTD (i.e., perpetrators of terrorist acts), UCDP/PRIO (i.e., parties to armed conflicts), and MAROB (organizations said to represent minority interests). Some sense of this challenge can be inferred from Table I, which illustrates how these databases align up for a select number of organizations, with the locations and periods of activity and the ID codes assigned.
Table I. Differences in inclusion and naming criteria, coverage

<table>
<thead>
<tr>
<th>Organization</th>
<th>Notes</th>
<th>TORG</th>
<th>TOPs</th>
<th>GTD</th>
<th>UCDP/PRIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic Front for the Liberation of Rwanda (FDLR)</td>
<td>ALIR merged with EX-FAR in 2000 to form FDLR.</td>
<td></td>
<td></td>
<td>2372: COD (2007–2012)</td>
<td></td>
</tr>
<tr>
<td>Ex-FAR (Former Armed Forces of Rwanda)</td>
<td>TOPs: FDLR, INTERAHAMWE, and EX-FAR are AKAs for ALIR.</td>
<td></td>
<td>4342: COD</td>
<td></td>
<td>1380: RWA, BDI, TZA, COD, UGA (2001–2012)</td>
</tr>
<tr>
<td>Interahamwe</td>
<td>EX-FAR refers to ex-military – it is not an organization.</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peuple en Armes Pour la Liberation du Rwanda (PALIR)</td>
<td>PALIR is the political wing of ALIR</td>
<td></td>
<td>4342: COD (1994–)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**KEY:** Numerical id: ISO a-3 code of active locales (years of activity).

As a crosswalk connecting constituent datasets, TORG has no independent data on active locales, or years of activity.

A related challenge stems from the way datasets refer to distinct organizations. GTD and POICN each publish a single name for each group, whereas TOPs, UCDP/PRIIO and MAROB included multiple aliases associated to a primary group name. A prime example is the Iraqi militant group known variously as ‘Tawhid and Jihad,’ ‘Al-Qa'ida in Iraq,’ and the ‘Islamic State of Iraq and the Levant,’ each identified as aliases of a single organization by UCDP/PRIIO (Table II). Similarly, TOPs lists ‘Tawhid and Jihad’ and ‘Al-Qa'ida in Iraq’ as two (of eight total) alternative names. However, as an event-centric dataset, GTD identifies each group separately (temporal overlap across group names is likely an artifact of journalistic reporting on attack attribution). Finally, MAROB does not track any groups using any of these names. Overall, we have discovered nearly forty associated name variants. Unfortunately, naming differences across datasets are very
common – approximately 40% of TORG groups have multiple associated names associated to them (although this includes minor differences in spelling and punctuation).

Table II. Differences in inclusion and naming criteria, coverage

<table>
<thead>
<tr>
<th>ORGANIZATION</th>
<th>NOTES</th>
<th>TORG</th>
<th>TOPs</th>
<th>GTD</th>
<th>UCDP/PRIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Islamic State of Iraq and the Levant (ISI)</td>
<td>UCDP/PRIO: Tawhid and Jihad rebranded as ISI in 2006.</td>
<td>2698</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**KEY:** Numerical id: ISO a-3 code of active locales (years of activity).
As a crosswalk connecting constituent datasets, TORG has no independent data on active locales, or years of activity.

In other cases, disparate groups can have similar (or even identical) names, raising the risk that we conflate truly distinct organizations. For instance, the TORG includes at least five different groups called the ‘People's Revolutionary Army.’ While specific identities can be parsed by triangulating location and periods of activity (and occasionally via additional research), this often requires significant effort. A common identifier framework prevents duplication of outside efforts by integrating periodic dataset updates to the crosswalk – allowing others to focus on theory development and testing.

The current version TORG (v. 2014 2.0) identifies 2,750 distinct organizations and 6,125 names. Since many other data sources use actor codes from one of the core sources, the crosswalk will also link to other data sources through other data connectors such as the ACD2EPR data, which link organizations in UCDP/PRIO to the ethnic groups in the Ethnic Power Relations data.
through explicit claims or recruitment (see Wucherpfennig et al., 2012). Table III indicates the number of unique organizations in the constituent datasets as well as overlap across datasets. It is clear that many organizations (204) engage in both terrorist activities and conventional armed conflict, as classified by GTD and UCDP/PRIO, respectively. Just as relevant are the non-overlapping records. In particular, we can see that organizations in conventional civil war is a relatively small proportion of all the groups in TORG (204/2,750 = 7%), consistent with the idea that only stronger groups have realistic prospects to engage in conventional warfare. Conversely, a large proportion of groups involved in armed conflicts do not engage in terrorism or indirect targeting, including revolutionary groups and drug cartels (204/341 = 40%). Yet, the fact that most UCDP/PRIO groups also appear in in GTD suggests that terrorism (at least as defined by GTD) is common in most civil wars.

Table III. Records and overlap between datasets in TORG (v. 2014 2.0)

<table>
<thead>
<tr>
<th></th>
<th>TORG v.2014 2.0</th>
<th>TOPs</th>
<th>GTD</th>
<th>UCDP/PRIO</th>
<th>MAROB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total records</td>
<td>6,145</td>
<td>2,587</td>
<td>3,974</td>
<td>1,530</td>
<td>179</td>
</tr>
<tr>
<td>Unique groups</td>
<td>2,750</td>
<td>814</td>
<td>1,647</td>
<td>341</td>
<td>35</td>
</tr>
<tr>
<td>TOPs</td>
<td>814</td>
<td>520</td>
<td>99</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>GTD</td>
<td></td>
<td>1,647</td>
<td></td>
<td>204</td>
<td>31</td>
</tr>
<tr>
<td>UCDP/PRIO</td>
<td></td>
<td></td>
<td></td>
<td>341</td>
<td>16</td>
</tr>
<tr>
<td>MAROB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>35</td>
</tr>
</tbody>
</table>
Beyond TORG: a shared contentious actor registry

Although the TORG crosswalk can support many research questions, its focus is limited to terrorist organizations and groups engaging in violent conflict leading to 25 or more deaths per year. As such, it is not well-suited to inform research on conflict beyond violence, such as when actors refrain from violence, or substitution between violence and non-violent tactics. To broaden the focus, we propose a more comprehensive Shared Contentious Actor Registry (SCAR) to identify non-state organizations in both violent and nonviolent contentious politics. Just as the TORG data allow studying the relationship between terrorism and civil war, a broader framework of organizational identifiers would aid the study of contentious organizations and the integration of data beyond violent conflict. We hope to achieve this through an advisory board, tasked with creating a comprehensive online registry of organizational names and associated reference identifiers for existing relevant data sources.

Many of the problems of classifying groups arise from changing interactions over time, from ‘rebranding’ efforts, to factional splits and the formation of stable coalitions. Properly accounting for these will require explicit protocols for attributing characteristics, events, and relationships to factions and splinter groups over time. Since umbrella groups may be significant in planning and resources (leaving ‘implementation’ to member groups) there may be value in identifying these separately.\(^1\) Accurate accounting of variations in intra- and inter-group

\(^1\) In fact, where such collaboration is stable – and relies on centralized coordination – UCDP/PRIO identifies these as distinct actors. In such instances, care must be taken to prevent the double-counting of actions attributed to groups (member or umbrella) across datasets.
relationships will require an efficient means of capturing longitudinal relational data,² capable of supporting automated matching with external datasets. Ultimately, we hope to support queries for individual groups, identifiers, locales and date ranges via a web-based interface.

Conclusion

The growing number of datasets on contentious political organizations has the potential to transform the study of political conflict, but this is hampered by the lack of a common organizational framework. Cooperative efforts that help users combine and take advantage of existing coding efforts can help prevent duplication of efforts and facilitate new opportunities for research. We offer the TORG crosswalk as both proof-of-concept and as a starting point for a more ambitious and inclusive ‘Rosetta Stone’ of contentious politics.

Acknowledgements

We thank Idean Salehyan as well the anonymous reviewers for their helpful suggestions. Gleditsch is grateful for support from the Research Council of Norway (213535/F10) and the European Research Council (313373). Support for this research was also provided by the Science and Technology Directorate of the US Department of Homeland Security (grant number 2008ST061ST0004) through the National Consortium for the Study of Terrorism and Responses to Terrorism (START).

---

² The UCDP project identifies splinter organizations and alliances, as well as ‘former’ naming variants (see www.pcr.uu.se/digitalAssets/55/55112_NonStateActorCodebook.pdf). However, the format makes it difficult to integrate this with other data sources.
References


VICTOR ASAL, b. 1965, PhD in Political Science (University of Maryland, College Park, 2003); Director of the Center for Policy Research, University at Albany, State University of New York (SUNY) (2003–).

KEN COUSINS, b. 1964, PhD in Political Science (University of Maryland, College Park 2006); Research Scientist, University at Albany, SUNY (2009–).