

Lessons from the Issue Correlates of War (ICOW) Project

Paul R Hensel

Department of Political Science, University of North Texas

Sara McLaughlin Mitchell

Department of Political Science, University of Iowa

Abstract: We present lessons and best practices for conflict data collection from the experiences of the Issue Correlates of War (ICOW) project. Common problems for conflict data collection include the development of a search strategy for potential events, the consultation of a broad range of sources, and recognition of the limitations of these sources. More general best practices address the development of detailed instructions for coders, detailed descriptions for data users, and strategies for managing research assistants and preserving project documentation.

Keywords: ICOW project, conflict data, data collection, data management, contentious issues

The decades since J David Singer and Melvin Small started the Correlates of War project have seen exciting developments in data collection. Conflict data sets now cover everything from threats of military force to politicide, and these have been supplemented by data on numerous political, social, and economic characteristics of political actors. However, these advances have not resulted in a set of best practices for data collection. Scholars starting new data collection efforts often start from scratch, and may waste time and resources that could have been used more productively if they had been guided by the experiences of earlier projects.

This article compiles a list of best practices from the experiences of the Issue Correlates of War (ICOW) research project. Many of the lessons learned by ICOW researchers have clear applicability for other scholars collecting data sets on interactions between or within states.

The ICOW project

ICOW collects data on contentious issues between nation-states, with a goal of understanding both conflict and negotiation processes over these issues (Hensel 2001; Hensel et al. 2008). ICOW began in 1997 as an effort to study contentious issues without restricting data collection to issues that became militarized. This allows scholars to study how issues are managed, with militarization being a topic for analysis rather a case selection mechanism.¹ This approach also allows for the study of peaceful issue management.

Three types of issues have been collected by ICOW, defined by explicit contention between official government representatives over *territorial* sovereignty, *maritime* zones, or the usage of international *ivers*. The project began with territorial issues, but the ultimate goal was to include additional issue types to allow systematic comparison of how different issues are managed. In 2002 ICOW expanded to include river and maritime issues; other issues may be collected in the future.

Analyzing data in the Western Hemisphere, Europe, and the Middle East, ICOW research finds that less than half of all contentious issues ever become militarized, thus studies of armed

¹ Issue militarization is related to the COW project's Militarized Interstate Dispute (MID) data set, but the correspondence is not perfect. Not all MID's that are coded as involving attempts to revise the territorial status quo qualify under ICOW coding rules, often because they involve support for secession or other efforts where the threatening state does not directly claim the territory for itself.

conflict issues are missing many exclusively diplomatic cases. Bilateral and third-party efforts to settle issues peacefully outnumber armed conflicts over the issues (Hensel et al. 2008). ICOW data sets have been used to study the conditions under which issues are likely to become militarized, as well as the conditions under which claimants undertake peaceful settlement attempts and the success of such attempts. ICOW data are also used to study many other research questions such as the impact of territorial conflict on trade, diversionary uses of force over issue claims, and the success of international organizations and courts as conflict managers.

Common problems faced by conflict data sets

Identifying potential cases

The most important problem in collecting conflict data is identifying potential cases. Unlike projects coding details of a known population of treaties or political characteristics of states, conflict data sets face a nearly infinite set of possible disputes to investigate. It is likely that some conflict events will never be uncovered, but the data collection strategy must be designed to identify as many events as possible.

One best practice is the use of as many different sources as possible. Any single source will have quirks or limitations in event coverage. In our experience, focusing on only one source would have missed many of the territorial, river, or maritime claims that were ultimately identified. We use a number of standard news sources (the *London Times*, *New York Times*, *Keesing's*, *Facts on File*, and Lexis-Nexis). These sources often cover individual events that may not show up in books, but academic books and journal articles offer greater context for events that they cover and may be the only source to cover many events before World War II. We recommend consulting books about the history and foreign relations of each country and region. Reference books are also helpful, including gazetteers or geographical dictionaries, as well as subject-specific books on armed conflicts, borders, maritime zones, and similar topics.

Once the appropriate set of sources has been determined, the research team must do as systematic a search as possible. For ICOW, this involved putting together a list of all international borders (including colonies or dependencies), adjoining maritime zones, or international rivers. We did not limit our data collection to such cases, but these lists offer a

useful starting point. It is also important to consult histories of each country or region to help identify conflicts against adversaries that do not appear in these lists.

It is vital for researchers to be alert during this systematic search, as additional cases may be mentioned briefly by a source that is discussing a different topic. Our researchers identified many new cases based on short passages in sources covering other events. Some of these new cases will not qualify for inclusion in the dataset, but we strongly recommend following up on any brief reference to be sure.

Limitations of sources

Once a candidate list of events is identified, several new problems may arise. One common issue is a lack of detail. Some obscure cases are mentioned in a few sentences in one or two sources, with little sense of exactly what happened. For example, the territorial claim between Canada and Denmark over Hans Island also involves disagreement over resources in the nearby maritime space. Researchers identified only a few news stories, making it difficult to determine the events surrounding the claim.

Another issue is the potential for some events to be kept secret. One example is the occurrence of secret negotiations which are only revealed some years later. Even where journalists indicate that there are reports of secret talks being held, the nature and timing of these talks may not be revealed until archives are opened decades later.

A third issue is uneven coverage of certain events. Journalists and historians are more likely to report "newsworthy" events than routine events. The ICOW project records the occurrence and success of efforts to settle contentious issues peacefully. While the press often reports militarized incidents or the onset of negotiations, followup stories about whether or not agreements were reached are less common, and information on compliance with agreements is even moreso. Some areas of the developing world are also less likely to receive coverage in English sources.

With any of these problems, our advice is to consult as many sources as possible in search of corroborating evidence. Where possible, identify multiple sources including newspapers, books, journal articles, and policy reports. Consult sources from multiple perspectives, such as books and newspapers from the perspective of each involved country.

Interoperability with other data sets

A final consideration is making sure that one's conflict data set allows interoperability with other data sets. Most conflict scholars identify countries using the COW project's interstate system membership list and country codes. This makes it easy to merge one's conflict data with existing data sets on democracy, alliances, capabilities, and other variables.

Unfortunately, such a list does not exist for all types of actors. Other scholars in this special issue work to develop a standardized list of terrorist or rebel groups. There is less progress in developing a definitive list of international organizations or non-state actors that become conflict participants or mediators. The ICOW project developed a list of the non-state actors that have attempted to settle issues, ranging from the UN Security Council to the Vatican and regional institutions. We released this list on the ICOW web site, but data interoperability would benefit from a standardized list of non-state actors.

General advice for data collectors

Detailed instructions for coders

J. David Singer often used the phrase "You live and die by your coding rules" to emphasize the importance of clear, explicit coding rules for data collection. It is desirable to remove individual judgment from the research process as much as possible. This should increase the reliability of the data collection effort, reduce coding errors, and make it easy for future scholars to determine why a particular case was coded as it was.

ICOW created a general coding manual that addresses issues that arise in all of our data sets, as well as separate coding manuals for details specific to the territorial, river, and maritime claims data sets. Following these instructions, there should be little confusion about how to code most situations that come up in data collection. ICOW also provides supplementary information on historical state names, so that researchers are aware of changes over time (e.g. Rhodesia becoming Zimbabwe)².

² The codebook and codesheet are available at <<http://www.icow.org>>.

One situation where clear and explicit coding rules matter occurs where sources use sloppy terminology. Journalists or historians may have personal conceptions of "war" that differ from the data set being collected. "Border dispute" frequently refers to any problem that crosses a border, even if this problem concerns immigration or trade rather than a territorial claim. "Mediation" is often used for third party activities ranging from fact-finding missions or shuttle diplomacy to arbitration. The coding rules must make clear what qualifies as a war, border dispute (or territorial claim), or mediation effort.

No coding manual can perfectly anticipate every situation that arises. It is important to instruct researchers to explain and justify coding decisions. Code sheets should have a "coding notes" section and researchers should be instructed to put as much detail into this section as possible. This allows them to explain their reasoning, and if later evidence becomes available, to consider re-coding that case.

Clear descriptions for users

It is also important to consider the eventual users of any data set. While the data collectors themselves understand how the data set is constructed and know any limitations of the variables, most data sets are intended for use by scholars beyond the original research team. Every data set should include a document that explains these matters to minimize the risk of confusion or misuse of the data by future scholars.

Managing research assistants

Most data collection projects involve the participation of research assistants beyond the principal investigators. ICOW has employed approximately fifty different research assistants for periods ranging from one month to three years. One unfortunate lesson is that there is wide variation in the motivation and background knowledge of research assistants. Furthermore, each time a new assistant starts work, potential research time is lost to training, and they will work more slowly until they become more experienced. To minimize these risks, we recommend hiring research assistants for extended periods where possible. ICOW has been far more productive when the

same research assistants were available for several consecutive years than when new assistants were hired every semester.

It is important to assign the research assistants appropriate tasks for their experience level and to monitor their work. We have used new research assistants for tasks like building chronologies of potential claims, monitoring and checking their work very carefully, and reserved the initial coding of cases for more experienced assistants or the co-PIs. To ensure inter-coder reliability, ICOW co-PIs review and approve every case before it enters the final dataset.

Finally, it is helpful to have the PIs do initial data collection by themselves before any research assistants are hired for the project. This ensures that the coding procedures are based on personal experience in coding the data, and that the PIs are able to answer any questions that arise.

Preserving documentation

Data collectors need to develop a reliable plan for preserving the information they collect. This allows scholars to go back to the original materials whenever questions arise from data users. There have been several times where consulting new sources led us to reconsider the coding of earlier cases, and having the original source material allowed us to reevaluate it in light of the new information. Earlier research projects did not always maintain documentation, with the first two versions of the COW Militarized Interstate Dispute data being a prominent example: while cases occurring from 1993-2010 (coded in the MID3 and MID4 efforts) are well documented, for 1816-1992 events, it can be difficult to determine why cases were coded a particular way. We recommend saving documentation electronically and having an off-site back-up to preserve the material for future usage.

Realistic time expectations

Our final advice is to be realistic about time tables. ICOW began in 1997 by focusing on South American territorial claims. The full data set of territorial claims across the globe, 1816-2001, was not released until 2013, and the river and maritime claims data sets are only completed for some regions of the world so far. There are many reasons, including the PIs' other professional

commitments, variation in the number and quality of research assistants, and the scope of the collected data. The addition of river and maritime issues and our collection of all cases for the entire ICOW temporal period slowed the public release of data. A focus on several decades at a time or data collection for one issue would have allowed faster data releases, yet hampered our ability to research certain questions. Our research approach made long-term sense, but it produced short-term delays. Future data projects should consider the implications of their research strategies for the amount of time that will be required.

Conclusion

ICOW project researchers encountered numerous issues in data collection. This article summarizes many of these issues and the ways that ICOW has addressed them, in the hope that this will help future data collectors avoid these problems. Readers wishing more details or examples may view the ICOW coding manuals and may download all ICOW data and user's guides at <<http://www.icow.org>>.

Acknowledgements

This article was presented at the workshop "Best Practices in Data Collection" at the 2013 International Studies Association conference. The authors thank the other workshop participants and the journal's reviewers for their comments. This project was supported by the National Science Foundation awards SES-0960567, SES-0214417, and SES-0079421; all responsibility for this article lies with the authors.

References

Hensel, Paul R (2008) Contentious issues and world politics: Territorial claims in the Americas, 1816-1992. *International Studies Quarterly* 45(1): 81-109.

Hensel, Paul R; Sara McLaughlin Mitchell, Thomas E Sowers II & Clayton L Thyne (2008) Bones of contention: Comparing territorial, maritime, and river issues. *Journal of Conflict Resolution* 52(1): 117-143.

PAUL R. HENSEL, b. 1970, PhD in Political Science (University of Illinois, 1996); Professor of Political Science, University of North Texas (2008-); research interests include territorial claims, international rivers, international conflict and conflict management. Recent articles in *International Interactions*, *International Studies Quarterly*, and *International Negotiation*.

SARA McLAUGHLIN MITCHELL, b. 1969, PhD in Political Science (Michigan State University, 1997); Professor of Political Science, University of Iowa (2004–); research interests include international conflict, democratic peace, international courts, conflict management, and contentious issues. Recent articles in *Journal of Conflict Resolution*, *International Interactions*, and *International Studies Perspectives*.