

ICEWS Events and Aggregations

Prepared by:
Jennifer Lautenschlager
Lockheed Martin
Advanced Technology Laboratories

Table of Contents

1. ICEWS Events and Related Concepts	3
1.1 Event Types	3
1.2 Actors	4
1.2.1 Named Actors	4
1.2.2 Composite Actors	5
1.2.3 Unspecified Actor	6
1.3 Agents	6
1.4 Sectors	7
1.5 Locations	9
2. ICEWS Monadic Event Aggregations	10
2.1 Defining how to filter events for aggregation purposes	10
2.2 Defining how to aggregate filtered event sets	13
2.3 The Size of Aggregate Event Data	14
2.4 Considerations when combining aggregations	14
2.5 Nuances of Aggregate Event Data	15
3. ICEWS Dyadic Event Aggregations	17
4. Provincial Event Aggregations	17
Appendix A: Composite Sector-Based Actors	18
Appendix B: Examples of Aggregation Definitions	21
Appendix C: How Monadic Events Sets are initially filtered	24
Appendix D: How Filtered Event Sets (monadic and dyadic) are currently aggregated	29
Appendix E: How Dyadic Events Sets are initially filtered	31
Appendix F: Province-Level Aggregation List	34
Appendix G: Glossary	38

Event aggregations in ICEWS is the means through which ICEWS event data is transformed into meaningful numerical values calculated over some time interval (e.g., monthly, weekly) and associated with a country¹. Event aggregations are used in iCAST as potential model parameter inputs. In order to understand how event aggregations work, there are several underlying concepts that must first be understood. With this groundwork in place, an explanation of the aggregation process can then be made.

1. ICEWS Events and Related Concepts

Events in ICEWS represent *who did what to whom, when, and where*. They are extracted through automatic means from the text of various news stories. One example of an event would be: Sheikh Khaled Al Jarrah Al Sabah engaged in negotiation with King Muhammad VI on June 27, 2001 in Rabat, Morocco.

1.1 Event Types

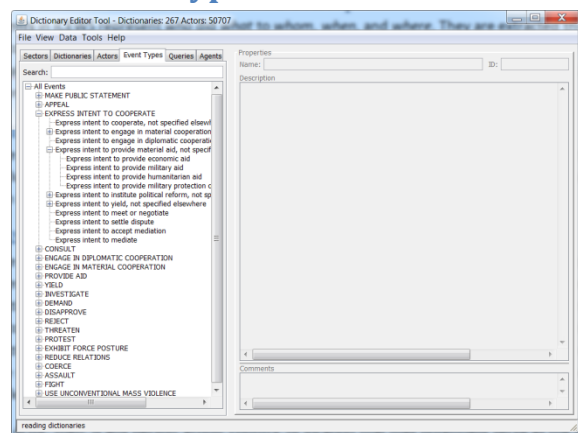


Figure 1 - A depiction of the event type hierarchy, by category.

The action part of events is known as the **Event Type** in ICEWS. There are 312 ICEWS event types, including the 20 category definitions under which they fall; a graphical depiction of some event types can be seen in Figure 1. ICEWS event types were originally derived directly from the CAMEO Conflict and Mediation Event Observations Codebook² and, in particular, its Verb Codebook. The names of the event types are typically descriptive enough for the end users, though more details can be found in this CAMEO Codebook.

In addition to each event type having a unique name and code (typically of system-level interest only) ascribed to them, they also have a numerical value ranging from -10 to +10. This number is alternately known as the **Goldstein value**, the **intensity value**, or simply the **event intensity**. The numbers originally came from the CAMEO Scale³ values of the associated CAMEO codes, which were in turn motivated by the Goldstein scale for WEIS event coding⁴. The number is used to represent the amount of *hostility* or *cooperation* implied by the event type, where negative numbers represent hostile actions and positive numbers represent cooperative actions; -10 represents the most hostile of hostile events, while +10 represents the most cooperative of cooperative events. Values of 0 are interpreted as being neutral.

The intensity value represents a convenient, albeit somewhat simplistic, way in which a single event can be translated into a numerical representation. When considering the intensity value of an event or set of events, there are a few considerations to keep in mind:

¹ Currently aggregations are associated a country, though conceptually they could be associated with any sort of geographical region, or even a set of actors or some other facet of events.

² <http://eventdata.psu.edu/cameo.dir/CAMEO.CDB.09b5.pdf>

³ <http://eventdata.psu.edu/cameo.dir/CAMEO.SCALE.txt>

⁴ <http://www.chsbs.cmich.edu/fattah/courses/empirical/jgscale.htm>

- The CAMEO scale from which intensity values were derived was created by a political scientist.
- The scale represents a neutral point of view. What this means is that the US would consider a military agreement between China and North Korea to be a hostile action, as military agreements in general are defined as being cooperative, it would be classified as a cooperative event. It is up to the analyst or other users to interpret the intensity values in the appropriate context.
- The scale does not take into account the concept of magnitude. That is to say, the killing of 1000 persons is classified as no more hostile than the killing of a single person.⁵

1.2 Actors

The “*who*” and “*whom*” parts (also known in ICEWS parlance as the **Source** and **Target**, respectively) of events are known as **Actors** in ICEWS. There are two general types of actors as described below.

1.2.1 Named Actors

Named actors are formally defined in the **ICEWS Actor Dictionary** in a process performed by system maintainers (as opposed to ICEWS end users). The actor dictionary is stored as a directory of XML text files that may be modified in a text editor, through automated ICEWS utilities, or most commonly in the graphical **ICEWS Dictionary Editor**. This *actor dictionary* is periodically ingested into the ICEWS core data repository for system usage; when changes are made to the XML files, but these files have not yet been ingested, none of the changes will yet be reflected in the ICEWS system.

A named actor, then, is any person, group, country, or location that may potentially serve as the source or target of an event. Actors have a variety of information associated with them:

- **Name:** The common name by which the actor is identified. The name is displayed in various displays in the ICEWS user interface.
- **Patterns:** Text that represents various ways in which the actor may be identified in news stories. Typically, all actors must have at least one pattern associated with them.⁶ It is common for most actors to have multiple patterns associated with them, used to denote alternate spellings or name variants (e.g., formal versus informal group names, acronyms for groups, persons commonly identified by title and last name).
- **Memberships:** Memberships are used to associate actors with their known affiliations, either in the context of a specific timeframe or in a more general sense. One type of affiliation is a membership or leadership position in groups identified by other actors (e.g., World Bank, General Motors) or by more general *sectors* (e.g., the military, dissidents), which are described

⁵ In practice, as events are extracted from news stories, and as large-scale actions such as mass murders tend to invite more written news stories than smaller actions such as the murder of a single individual, there tend to be more ICEWS events for real world actions of greater magnitude. This has provided a limited means through which the impact of an event’s magnitude can be assessed, in the absence of a scale that takes magnitude into effect: through quantity of events.

⁶ The two most common exceptions to this rule are: a) actors that represent regions with which an actor may be associated, even though the region itself can never appear as the source or target of an event; or, b) organizations that correspond

in the next section. Another type of affiliation is a general group identity, such as ‘Muslims’ or ‘Bedouin people’. Memberships can have *roles* associated with them, to indicate the actor’s role within the organization or sector. For example, an actor may be a member of a political party, or they may be a leader of a political party. Though roles are persisted in the actor dictionary XML files, they are not currently used for any purpose in the ICEWS system itself.

- **Type:** Actors are classified as individuals, groups, or locations. There are certain system rules in ICEWS that make use of this information. For example, it is a rule that actors may be members of other actors that are of type group or location, but they may not be members of other individuals.
- **Description:** Text descriptions may optionally be given to provide more free-form information about an actor, or to cite sources.
- **Comments:** Similarly, free-form text comments may be used to record information about the evolution of the actor information, for example, noting when it was originally created or modified, or explaining why certain information was recorded in the way it was. Automated systems that create actors may also record information here.
- **ID:** A special identifier is used within the ICEWS system to uniquely identify the actor.
- **Gender, Date of Birth, Date of Death:** These attributes may optionally be recorded for reference, though they are not currently used by the ICEWS system.

The two images below are screenshots from the dictionary editor when used to edit the actor Kevin Rudd of Australia.

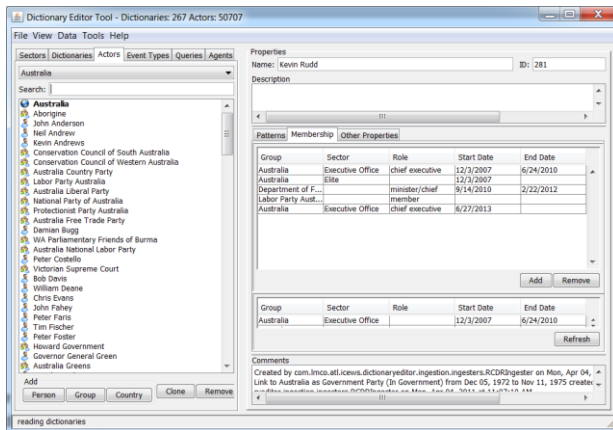


Figure 2 - The dictionary editor used to edit information about Kevin Rudd of Australia. Here, the actor’s memberships are being viewed. They consist primarily of different government positions held over the course of time.

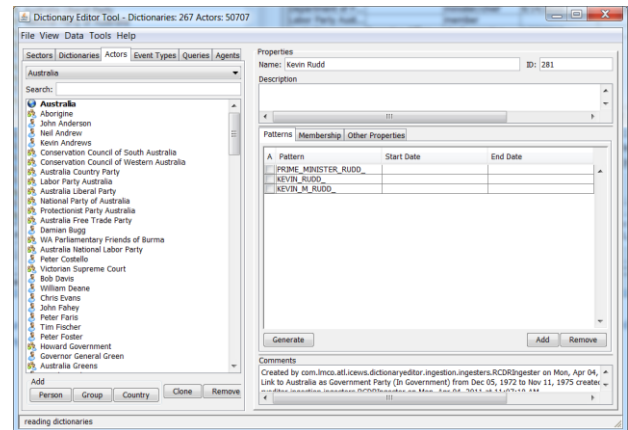


Figure 3 - The dictionary editor used to edit information about Kevin Rudd of Australia. Here, the patterns of the actor are displayed, which represents the various ways in which this person may be referenced in news stories.

1.2.2 Composite Actors

Composite Actors are not listed in the Actor Dictionary but are instead formed by combining an actor with either an agent or a sector. The former is known as a *composite agent-based actor* and is described

in the following section. The latter is known as a *composite sector-based actor* and is a more complex topic explained in Appendix A.

1.2.3 Unspecified Actor

There is a special actor, named 'Unspecified Actor', which is used as a placeholder when the target of an event is not known. This makes the event what is known as a *monadic event*, where there is a clearly identified source actor but no identified target actor. These events are almost exclusively protest events, where the specific target of the protest is either not known, or is an abstract concept.

1.3 Agents

There are some individuals or groups that are common in many, perhaps all, countries. Examples include the military, students, or protesters. Though such groups and individuals could be listed as actors in every country in which they may occur, this would quickly become tedious. There would be the 'Egyptian police', the 'Iranian police', the 'Brazilian polices', and so on and so forth.

The concept of an **Agent** in ICEWS is the solution to this dilemma. An agent is similar to an actor, in that it can serve as the source or target of an event. Unlike an actor, when initially defined in the Actor Dictionary, an agent has no country affiliation. It is only at the point in time where an event is automatically coded that the country or group affiliation is associated with an agent. Because of this, a single agent such as 'Police' can be used in the context of any country, but need only be defined in the Actor Dictionary once⁷. The country or group affiliation is done through a *paired actor reference*, combined with the *paired agent reference*, combining to form what is known as a *composite agent-based actor*.

As with actors, agents have a variety of information that is associated with them and stored in the Actor Dictionary:

- **Name:** The common name by which the agent is identified. The name is displayed in various displays in the ICEWS user interface, combined with its paired actor reference, as 'Paired Agent Reference (Paired Actor Reference)'. Examples include 'Police (Egypt)', 'Rebel (Aum Shinrikyo)', and 'Student (Baghdad University)'.
- **Patterns:** Text that represents various ways in which the agent may be identified in news stories. This is analogous to the patterns that actors have.
- **Memberships:** Memberships are used to associate actors with their known affiliations, and are analogous to memberships for actors with one distinction: memberships for agents never have time constraints placed on them.
- **Type:** Agents are classified as individuals or groups. There are no agents that represent locations.

⁷ For those who remember your grammar classes from grade school, an actor is typically a *proper noun* while an agent is generally an *improper noun*. There are, however, some notable exceptions to the latter rule. Some agents are proper nouns that occur in the context of many countries, to the extent that it is easier to define them once as an agent than repeatedly with distinct country affiliations. In general, political parties (e.g., Democratic Party, People's Party) and common organization names (e.g. The Times) are good candidates to be made into agents.

- **Description, Comments, and ID:** These concepts are analogous to the ones for actors.
- **CAMEO Code, SAE Code:** These are special codes retained for the interest of social and political scientist modelers, and are of no interest to end users.
- **Restricted to Country Actors:** By default, an agent can have as its paired reference actor any actor representing a group, location, or country. Agents can never be paired with individuals. However, at times it is useful to specify that an agent only makes sense when it is paired with an actor representing a country. For example, since a military is generally something that only countries claim, the military agent is defined so that it can only be paired with an actor which is a country. This is done by setting its 'Restricted to Country Actors' value to True.
- **Agent Identified by Sectors:** This is an advanced concept which is not essential to understand for the scope of this document. Those who are interested, however, may refer to Appendix A for an explanation. Be sure to understand the concept of sectors first.

1.4 Sectors

Sectors in ICEWS are general affiliations which can be applied to an actor via its defined memberships. Some sectors, known as **national sectors**, can only be used when defined in the context of another actor, which must either be a country actor or some actor (e.g., person or group) affiliated with a country through its memberships. For example, 'Military' is a national sector, as in ICEWS the concept of identifying an actor as being involved with the military must occur within the context of a country. Other sectors, known as **unaffiliated sectors**, should never be used in the context of another actor, and instead are independently applied to an actor. For example, 'Muslim' is an unaffiliated sector, as in ICEWS one specifies that an actor is Muslim without the need to ascribe any particular national identity.

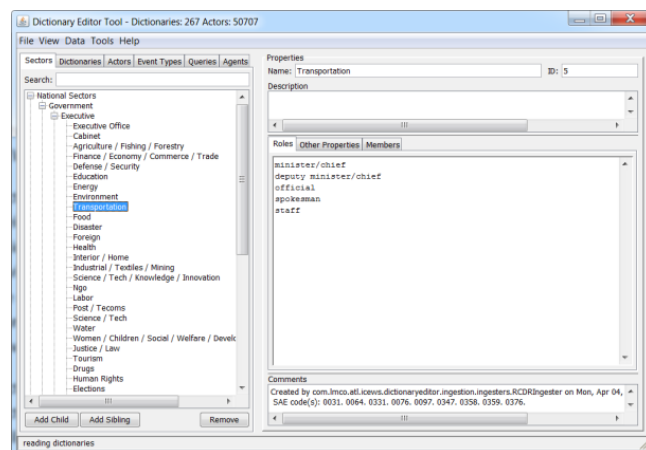


Figure 4 - The dictionary editor showing the hierarchy of national sectors in the executive branch of government.

Sectors are arranged in a hierarchical fashion, where all child sectors inherit the sector affiliations of their ancestors. For example, in Figure 4 the Transportation sector derives from the Executive sector, which derives from the Government sector, which is a national sector. An actor who is identified as having a membership to the Transportation sector of some country – for example, an actor serving as the country's Minister of Transportation, would have an *explicit* membership stated to the Transportation sector. They would also have *implicit* memberships to the Executive sector and the Government sector.

Sectors, then, have a certain transitive property about them, as when Sector A has a membership to Sector B (by nature of being a descendent of Sector B), then any actor affiliated with Sector A is also affiliated with Sector B. The same transitive property is applied to groups that an actor has a

membership to. For example, if an actor A has a membership to some group B, which in turn has a membership to some sector C, then by the transitive property of sector membership actor A also has an implied membership to sector C.

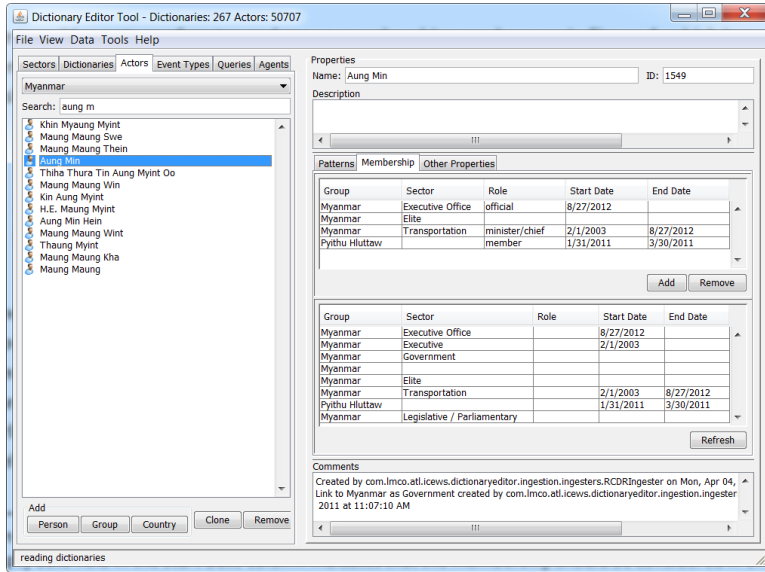


Figure 5 - Dictionary Editor showing the explicit and implicit membership for Aung Min of Myanmar.

The transitive nature of sector and group memberships can be seen in Figure 5, which is a screenshot of membership information for Aung Min of Myanmar. Here, there are four memberships explicitly stated, as shown in the top panel of the membership tab:

1. He has been an Executive Office official for Myanmar from 8/27/2012 to present⁸.
2. He is always considered to be an Elite of Myanmar.
3. He was the Minister of Transportation for Myanmar from 2/1/2003 to 8/27/2012.

4. He was a member of Pyithu Hluttaw (the House of Representatives for Myanmar) from 1/31/2011 to 3/30/2011.

Note that, as in this example, an actor may have multiple sector or actor memberships that are valid for a particular date. In this example, on 3/1/2011 Aung Min had an active membership to the Elite sector, to the Transportation sector, and to the Pyithu Hluttaw.

Whereas the top panel shows only *explicit* memberships that have been defined, the bottom panel shows both explicit and *implicit* memberships. In this example, the implicit sector membership to the Executive sector and the Government sector are derived through the explicit sector memberships to the Transportation sector and the Executive Office sector. The implicit membership to the Legislative / Parliamentary sector is derived through the explicit actor membership to the Pyithu Hluttaw (which has itself an explicit sector membership to the Legislative / Parliamentary sector).

Realize also that by giving an actor a membership to a national sector, you are also giving them an affiliation to some country, that being the country in whose context the national sector is specified. Typically, an actor is affiliated with a single country during the duration of their lifespan. Some actors may be affiliated with multiple countries, either at different points in time (for example, a person who

⁸ An empty date field in the End Date column indicates that the membership continues into the present, where an empty date field in the Start Date column indicates that the membership should be considered in effect for any point in time up until the specified End Date. If a membership has neither Start nor End Date specified, then it is assumed to be in effect at all times.

was involved in the government of Sudan who later was involved with the government of South Sudan when the country established its independence) or simultaneously (for example, a person who is involved in the government of some country, such as the Netherlands, and also in the government of some constituent country, such as Aruba).

Sectors are defined in the ICEWS actor dictionary and are relatively static, seldom if ever changing since their initial definition. In general, their meaning is fairly straightforward, though a few interesting points are worth noting:

- There are national and unaffiliated sectors with similar or even identical names and similar interpretations. One example is the Rebel sectors, which exist as both a national and international sector. These two sectors are actually distinct from one other, with the following difference: the national Rebel sector is *always* specified in the context of some country, while the unaffiliated Rebel sector is *never* specified in the context of some country. Which Rebel sector should be used for a particular actor is a subtle distinction. If the actor has a strong national identity, then the national Rebel sector should be used in the conjunction with the country in question. If, on the other hand, the actor has no known or no strong national identity, then the unaffiliated Rebel sector should be used.
- There is a special 'Unidentified Sector' that is used as an affiliation for the special Unspecified Actor. This enables it to be used in standard sector-based aggregations.
- There is a special 'Agent-Based Sector' that is used to identify actors that are composite agent-based actors. This enables them to be used in standard sector-based aggregations.
- As the only way to give an actor a national identity is through a national sector or group membership, it is sometimes desirable to use a sector to express this identity without an emphasis on the sector itself. This is done to ensure that references made to a politician, for example, before they came into office (at which point they would have national sector membership in the government area), are still made in a manner such that the politician maintains their national identity. Two national sectors are commonly used for this purpose. For more prominent individuals (e.g., world leaders, widely recognized politicians) the Elite national sector is used. For less prominent individuals the General Population national sector is used. It is not uncommon to see one or the other of these sectors assigned as a membership for an actor, with no date range specified, in order to preserve this national identity at any time the actor might be referenced in a news story.

1.5 Locations

The “*where*” part of an event is known in ICEWS as its **Location**. A location may be a country, province, district, city, location within a city, or region of multiple countries. If a location is a city or district, its province will also be known. As with all of event generation, locations are automatically associated with events through an automated process. Event locations only factor into provincial event aggregations.

2. ICEWS Monadic Event Aggregations

ICEWS monadic event aggregations (most often referred to as simply event aggregations, with the default assumption being that they are of the monadic type) provide the means to associate numerical data derived from event data with a given country for some time interval. Conceptually there are two steps to calculating an event aggregation:

1. Determine a set of events to use in the aggregation, based on the source actor, target actor, event type, and event date of all ICEWS events. This is sometimes known as *filtering* the events.
2. Calculate a numerical value for the event set, either by counting the events or by performing mathematical computations on the events' intensity values. This is sometimes known as *aggregating* the events.

It can be helpful, then, to consider the task of defining event aggregations as two distinct steps which can be defined *independently* of one another. Indeed, this is how aggregations are currently defined in ICEWS, as two separate pieces of information: how to filter the events, and how to aggregate the events.

2.1 Defining how to filter events for aggregation purposes

Filtering events is done by specifying the following properties of events:

- **Event Date:** This is implicitly defined based on the time interval at which events are to be aggregated. For ICEWS, a monthly time interval is currently used; this means that aggregations for January 2001 filter events occurring between 1/1/2001 and 1/31/2001, and events for February 2001 filter events occurring between 2/1/2001 and 2/28/2001, and so forth. Other time intervals, such as weekly or daily intervals, follow an equivalent principle.
- **Source or Target Country Affiliation:** Just as aggregations are always expressed in the context of a time interval, aggregations are also always expressed in the context of a country. Sometimes, it is specified that this is based on the source actor affiliation, whereby for a 'China January 2001' aggregation value, only events whose source actor has an affiliation with China *at the time of the event* are considered. Other times, it is specified based on the country affiliation of the target actor. Often the source and target country affiliations are required to be the same, which means that only events internal to a country are included in the aggregation. The alternative is where one actor (either source or target) is affiliated with the country in question while the other actor is associated either with a different country or no country at all.
- **Source or Target Sector Affiliation:** A final way of filtering events relies on examining the sectors affiliated with an actor *at the time of the event* to determine which events are included in the filtered sets. For example, an aggregation might specify that only events where the source actor of the event is affiliated with the government be included. Suppose a particular actor involved in several events over a number of years held a government office from 2004-2008, at which point they left government service. An event involving that actor as its source on 6/28/2007 would be included in such an aggregation, because the actor was affiliated with the government *at that point in time*. Conversely, an event involving that actor as its source on 6/28/2009 would **not** be included in such an aggregation, as they had no government affiliation at the time of the event.

Sector affiliations are one of the most powerful methods used in defining event aggregations. There are a few nuances to bear in mind when defining sector affiliation requirements for source or target actors:

- If no sector affiliations are defined, then any event that falls in the correct event date range, and have the appropriate source or target country affiliations, will be included in the set of events used for the aggregation.
- Source and target sector affiliations can either be inclusive (requiring that an actor be affiliated with the sector in question) or exclusive (requiring that an actor **not** be affiliated with the sector in question).
- There is a special 'National Sector' which is used to indicate that the source (or target) actor must be a country-level actor or (as of July 2014) an autonomous region that is not expressly identified as being a separatist region. This is needed as countries and such autonomous regions

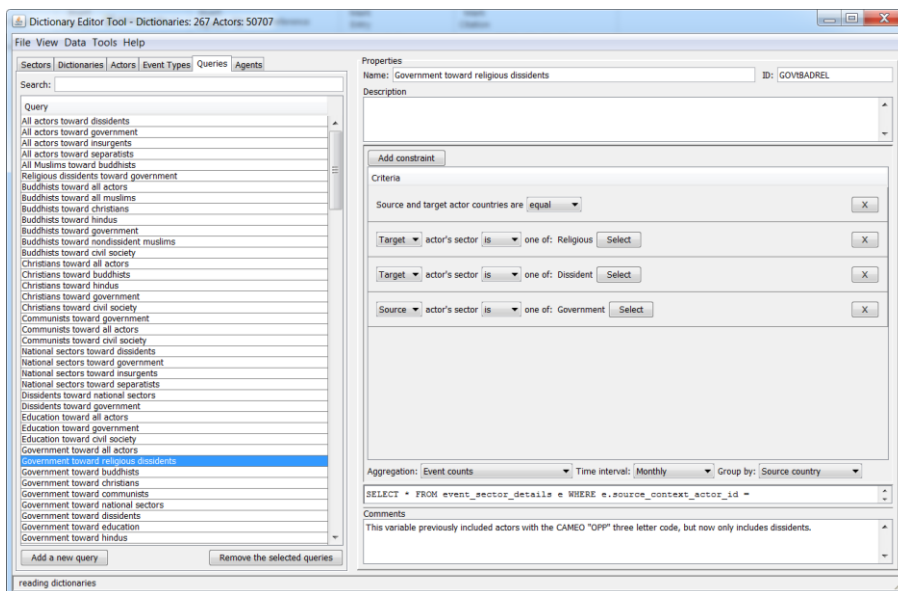


Figure 6 - Dictionary Editor being used to specify an event aggregation.

are the two types of actors which, by design, have no sector affiliation. Without this special 'National Sector' there would be no way to specify that an event must specifically involve either the country itself (or one of its autonomous regions), and not simply a group or individual within the country.

The way to filter events, as described above, is specified in the Actor Dictionary through the use of the Dictionary Editor. This is done in the Queries section of the Dictionary Editor, with the information stored in the `queries.xml` file. There is a 1-to-n relationship between event queries defined in the Dictionary Editor and actual event aggregations, where the value *n* is defined by the number of ways that there are to aggregate a set of filtered events into a numerical value.

The specification of an event query (also referred to as a *dyadic pair*) consists of the following pieces of information:

- **Name:** A descriptive name by which the event filtering portion of the query is identified.
- **ID:** A unique identifier by which the event filtering portion of the query is identified. Typically, there is a very specific naming convention used in query IDs, which are a short-hand method of specifying what the query represents. As with the name, however, there is **no requirement that the ID of a query accurately reflect what it represents**. Though certainly in the case of implied

standards such an agreement is highly desirable, **the ICEWS system in no way uses the ID itself to filter the events used in the aggregation.** It is instead the Criteria of the aggregation that define the filtering of events.

- **Constraints** (also known as **Criteria**): Constraints, or criteria, are the means through which the country and sector affiliations of the query are enforced. There are two types of constraints, as follows:
 - **Country constraints:** All queries have one and only one country constraint, which must be satisfied for an event to be included in the filtered set of events for an aggregation. This constraint specifies whether the source and target country affiliations must be the same or different, and corresponds to the Source or Target Country Affiliation property described previously for filtering events.
 - **Sector constraints:** Queries can have one or more sector constraints, or they may have no sector constraints at all. Regardless of the number, any event included in the filtered set of events for an aggregation must satisfy **all** sector constraints. Conceptually, this corresponds to the Source or Target Sector Affiliation property described earlier in filtering events. Sector constraints are described in further detail following this list of event query properties.
- **Description:** As with actors, agents, and sectors, queries can have a textual description that provides further information about the query definition, intended for human users of the aggregation.
- **Comments:** This is analogous in purpose to the Comments associated with Actors, Agents, and Sectors.
- **Time Interval:** This specifies how the events are temporally grouped for aggregation purposes. The time interval must be one that is supported by the ICEWS system, which currently consists of Monthly and Weekly intervals.
- **Group By:** As aggregations are always expressed in the context of a country, the 'Group By' value simply specified whether the source, or target, country affiliation should be used. Note that in cases where there is a country constraint requiring that the source and target country affiliation be equal, the two 'Group By' options of 'source' and 'target' are equivalent.

Sector constraints are specified in the context of source actors, target actors, or both source and target actors. As mentioned previously, the sector constraints can be either inclusive, requiring that the constraint be met for the event to be included in the filter set of events for the aggregation, or exclusive, requiring that the constraint **not** be met. Finally, one or more sectors (including the special National Sector mentioned previously) are associated with a single sector constraint. In the case of an inclusive sector constraint, only **one** of the listed sectors must be matched for the constraint to be satisfied⁹. In the case of an exclusive sector constraint, the actor(s) in question must not match **any** of the sectors specified.

⁹ This implies that if you wish to require multiple sectors be matched for the inclusion of an event in a filtered set of events, you must use multiple sector constraints.

In order to best understand how to define how to filter events for aggregation purposes, refer to the annotated examples in Appendix B, along with the list of filters in Appendix C.

2.2 Defining how to aggregate filtered event sets

Once a filtered set of events is calculated for a given country and time interval, based on a defined event query, it is necessary to derive a single numerical value from the event set. There is, however, one final filtering of events that is done before calculating this value. This final filtering of events is done based on the event's event type, and the same set of additional event type filters is applied to every single set of events to aggregate.

The way that this final stage of filtering, along with the actual aggregation into a numerical value, is defined is not through the Dictionary Editor. Instead, it is defined in the ICEWS system code, with the intention to move it into the Dictionary Editor at some future point in time for greater configurability. Until this is done, it requires a code change to add a new way of post-filtering and aggregating event sets.

The post-filtering of event sets based on event type can be done in a number of ways:

- **Through a range of event intensities:** As all event types have an associated numerical event intensity, a range of intensity values can be used to post-filter the events. Only events whose event type's intensity value fall within the specified range will pass this additional post-filter.
- **Through a set of one or more categories:** As described previously, there are 312 event types which fall into one of twenty different categories. By defining a set of one or more categories, only events whose event types match (or, in the exclusive case, do not match) the event type categories will pass the additional post-filter.

Once the final post-filtered set of events is determined at last, the events are aggregated into a final numerical value. This can be done by one of the following methods:

- **Event count:** In this case, a simple count of events which passed the final post-filter is recorded. This will result in a non-negative integer.
- **Event days:** Counts the number of days in a month in which at least one (and possibly multiple) events occurred. This value will range from 0 to 31.
- **Sum of event intensity:** In this case, the event intensities of all the post-filtered events are summed together, resulting in a decimal value that may be positive or negative.
- **Average of event intensity:** This case is similar to the sum of event intensity, except instead of summing all of the event intensities, they are averaged.
- **Standard deviation of event intensity:** This case is similar to the average of event intensity, except of calculating the average, the standard deviation is calculated.

For a current description of how filtered event sets are aggregated, refer to Appendix D.

2.3 The Size of Aggregate Event Data

The way in which events are aggregated creates a combinatorial explosion of data. To understand, consider the current state of ICEWS event aggregations, which involve:

- **256 dyadic pairs** of source and/or target actor constraints defined in the Actor Dictionary
- **44 ways** to post-filter and aggregate each of these dyadic pairs
- **167 countries** (corresponding to the countries of interest to iCAST)
- **Many months** of aggregations for each of the 167 countries (representing monthly aggregations from January 2001 to present)

With 248 dyadic pairs, and 40-44 ways (depending on dyadic pair) to post-filter and aggregate them, there are a total of 10,752 different aggregation variables tracked in the ICEWS system.¹⁰ Given that each of these variables has a value for each of 167 countries and 164+ months, there are nearly 300 million aggregate values currently available in ICEWS, any of which may be used in one or more iCAST models.

In 2015, weekly aggregations have been added to support the ME-ICEWS effort. This increases the amount of data by roughly a factor of 4.

2.4 Considerations when combining aggregations

In addition to using aggregate values directly in models, values may be combined to come up with new data. For example, by combining an aggregation that looks at events with an event type in the 'Fight' category with one that looks at events with an event type in the 'Assault' category, you end up with an aggregation of events that involve either the 'Fight' or 'Assault' category.

Care should be taken, however, when combining event aggregations to avoid the unintentional 'double-counting' of events. For example, consider the following two aggregations:

1. Event counts where the source actor is a government-affiliated actor
2. Event counts where the source actor is an actor in the 'Social' sector

Because an actor can hold multiple sector affiliations at a given point in time, it is possible that the same event might be counted in both aggregations. Though 'double-counting' events is not necessarily inherently wrong, it should be understood when the potential exists for double-counting, in case such an occurrence is undesired.

In general, aggregations with different post-filtering techniques, which are based on the event type of the events, are easier to analyze for potential double-counting circumstances. This is because an event has one and only one event type. If one count is of high hostility events with an intensity value ≤ -8 ,

¹⁰ Actually there are some 600 fewer, as the monadic events unspecified actors are only generated for protest events. This means that some aggregation combinations, such as GOVtUNKassaultct, will always be zero, as assault events will never be present in the system with an unspecified actor as a target.

and another count is of medium hostility events with an intensity value > -8 and ≤ -4 , it is guaranteed that combining the aggregations will not lead to a double-counting of events.

With sector constraints, however, unless one aggregation has an 'is not a member of sector' constraint on the same sector for which the other aggregation has an 'is a member of sector' constraint, guaranteeing that no event will be potentially double-counted is much more difficult. For example, if one aggregation has a source sector constraint of 'is not in the government sector', and the other aggregation has a source sector constraint of 'is in the government sector', there will be no potential double-counting of events; this is because no single event could possibly satisfy both constraints simultaneously. However, if one aggregation has a source sector constraint of 'is in the government sector' and the other has a source sector constraint of 'is in the social sector', it is possible for the same event to count toward both aggregations, as it is possible for the source actor to simultaneously hold affiliations toward the government sector and some social-derived sector, such as 'Business'.

2.5 Nuances of Aggregate Event Data

For any event aggregation that is used by an iCAST model, ICEWS tracks the post-filtered set of events that are used in calculating the aggregations¹¹. Sometimes on examining these post-filtered sets of events, there are events which someone might assume should be included in the aggregation, and yet are not.

A subtle source of excluded events comes from inconsistencies, oddities, or actual errors within the ICEWS dictionaries, and in particular in the memberships of named actors. As an example, consider an aggregation involving Laurent Nkunda of the Democratic Republic of Congo, who is defined in the Actor Dictionary with sector membership to Criminals / Gangs since 1/22/2009, and a membership to the National Congress for the Defence of the People (aka, the CNDP) without any date constraints on the membership. The CNDP is, in turn, given a sector membership to the Organized Violent sector from 12/1/2006 to 1/1/2009.

¹¹ Currently there are roughly 500 of the 7760 event aggregations used by iCAST.

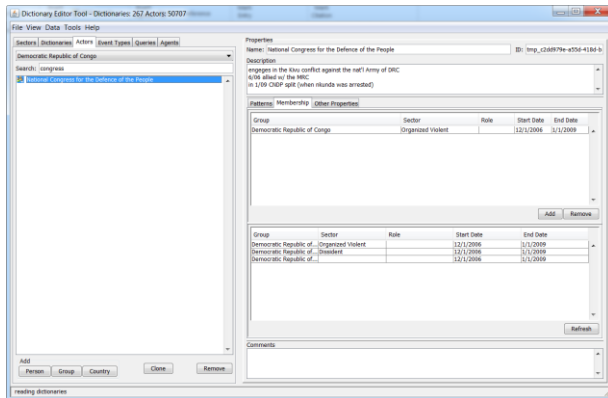


Figure 7 - Dictionary Editor view of the sector memberships for the Congolese National Congress for the Defence of the People.

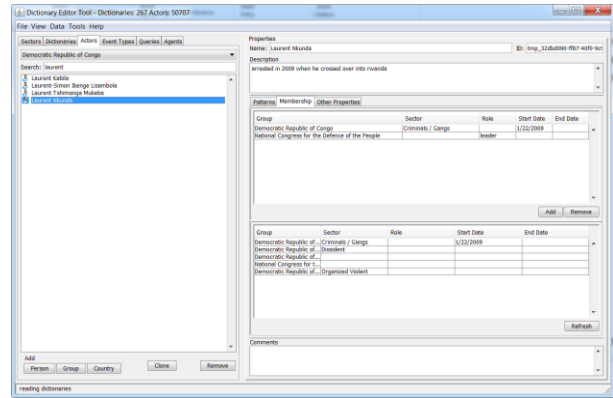


Figure 8 - Dictionary Editor view of the sector and group memberships for Laurent Nkunda of the DRC.

Now consider an aggregation that is concerned with events between the government and dissidents. This would be specified with sector constraints, where the source actor is required to be a member of the government sector and the target actor is required to be a member of the dissident sector.

Then consider an event on 6/30/2004 between the Congolese government and Laurent Nkunda. At first glance, it may seem that this event would fall into the filtered event set for the described aggregations for June 2004. After all, Laurent Nkunda is always a member of the CNDP, and the CNDP's only sector membership is to the Organized Violent sector, which is itself derived from the Dissident Sector. From the transitive nature of sector memberships, it seems as if Laurent Nkunda would also have a membership to the Dissident sector, thus causing the event to pass the query filter.

The nuance here lies with the date. Laurent Nkunda is always a member of the CNDP, but the CNDP only has a membership to the Organized Violent sector from 12/1/2006-1/1/2009. The event in question occurred in 2004. At this point in time, the CNDP actually had **no** sector memberships active, so no sector memberships could be transitively associated with Laurent Nkunda. The event therefore does not pass the required sector constraint.

This highlights an inconsistency in the Actor Dictionary. The reason that the CNDP has its sector membership defined for a set time period is that the time corresponds to when the organization was actually in existence. A more consistent dictionary would have Laurent Nkunda a member of the organization for this same date range. Given the scope of the actor dictionary, such inconsistencies can and will occur.¹²

¹² Note that if, for the sake of argument, Laurent Nkunda was made a member of the CNDP only for the dates of its existence, then when he was involved in an event in 2004, he would have **no** sector memberships. Though he would still show up in filtered event sets where there no constraints placed on the source (or target, depending) actor, he would not show up in any aggregations that did specify constraints. This is because sector constraints must be satisfied by the actor based on the sector affiliations *at the time of the event*. If he should be considered as a dissident prior to the CNDP's inception in 2006, he would need an explicit sector membership for that time frame.

3. ICEWS Dyadic Event Aggregations

In the summer of 2014, the concept of (directed) dyadic event aggregations were added. Here, the interactions between source and target countries are specified by source and target country affiliations. Originally any interaction involving actors within a country were considered. As of December 2014, further restrictions were added to enable sector-based dyadic aggregations, similar to how sector-based monadic aggregations work. In this sense, dyadic event aggregations perform very similarly to regular monadic event aggregations in terms of how events are aggregated (through one of the five aggregation methods defined in Section 2.2).

Dyadic event aggregations are directed. That is, the two countries specifying an aggregation refer to the source and target countries, in order. This means that a particular type of dyadic aggregation for China to Japan will be different than the same type of dyadic aggregation for Japan to China.

As sector affiliations are used in dyadic event aggregations, the types of aggregations per source-target pairing are the same as with the monadic data, resulting in roughly 10,752 different aggregation variables tracked. As there are 28,056 possible country pairings, leading to roughly 301 million possible aggregate values a year; however, this is a very sparsely populated matrix as the vast majority of values are 0.

4. Provincial Event Aggregations

Starting in 2015, provincial event aggregations *for select countries* were added, to support province-level modeling of select Middle East countries. Also included are the Occupied Palestinian Territories, broken down into the provinces of West Bank and Gaza Strip. These aggregations are similar to country-level aggregations, except the events are aggregated within sub-country regions, not countries. Only the ALLtALL dyad is supported for provincial aggregations.

The list of province-level aggregations supported is:

- **Egypt:** All 27 current governorates as of October 2014. This doesn't include:
 - *Helwan*, as it only existed between 2008 and 2011 and has since been re-incorporated into Cairo Governorate.
 - *6th of October*, as it only existed from 2008 to 2011 and has since been re-incorporated into Giza Governorate. Plus it's a silly name.
- **Iraq:** 18 of the 19 current governorates as of October 2014. This does not include:
 - *Halabja*, as it wasn't established until 2014
- **Jordan:** All 12 current governorates as of October 2014.
- **Libya:** All 22 current districts (established in 2007) as of 2014.
- **Saudi Arabia:** All 13 current provinces as of 2014.
- **Syria:** All 14 current governorates as of 2014.
- **Yemen:** 20 of the 21 current governorates, plus the municipality of Amanat Al Asimah (which is mostly the capital Sana'a) as of 2014. This does not include:
 - Soqatra Governorate, which was just established independently in December 2013
- **Occupied Palestinian Territories:** Both territories as of October 2014.

Appendix F lists all provinces for the eight countries above, for full details.

Appendix A: Composite Sector-Based Actors

Composite actors, which are not listed in the ICEWS Actor Dictionary, are created dynamically by the ICEWS system by combining a named actor with either an agent or a sector. Agent-based composite actors are described in section 1.3. Sector-based composite actors are described in this appendix.

Sector-based composite actors are used primarily for organizational display purposes. In various ICEWS user interfaces, lists of dozens or even hundreds of actors are presented. As an alternative to displaying them as a flat list, the actors can be displayed in a hierarchical view, with their associated sector and group actor memberships forming the hierarchy. The sectors are paired with country actors to create a country-sector hierarchy that parallels the national sector hierarchy. Actors are placed at one or more locations within the hierarchy, depending on the temporal context in which they are viewed.

Introducing the organizational element of sector-based actors into user interfaces can lead, at times, to odd visualizations. Consider the case where there is a Military national sector, a Military agent, and a named actor for the country’s military, such as Orangeland National Armed Forces. The Military agent and actor would, of course, be given membership to the Military sector. When viewed in a hierarchical visualization of actors, it would look similar to the depiction below:

- Government (Orangeland)
 - Military (Orangeland)
 - Military (Orangeland) ← sector-based composite actor
 - Orangeland National Armed Forces ← sector-based composite actor
 - General So-and-So ← agent-based composite actor
 - Orangeland National Armed Forces ← named actor
 - General So-and-So ← named actor

There is obvious visual redundancy in this visualization, as conceptually speaking the three military-oriented actors represent the same entity. Moreover, there will never be events associated with the sector-based actor of “Military (Orangeland)”. There are ways, however, to effectively ‘condense’ these three actors into a single actor.

In the Dictionary Editor, when editing agents, the “Agent Identified By Sectors” checkbox can be used to merge a sector-based composite actor and an agent-based composite actor. By selecting this checkbox, **there is only a single sector assignment for the agent**, instead of creating a new agent-based composite actor, the appropriate sector-based composite actor will be used. For example, in Figure 9 the Military agent is being edited. It has been associated with the national Military sector (not shown as it’s on

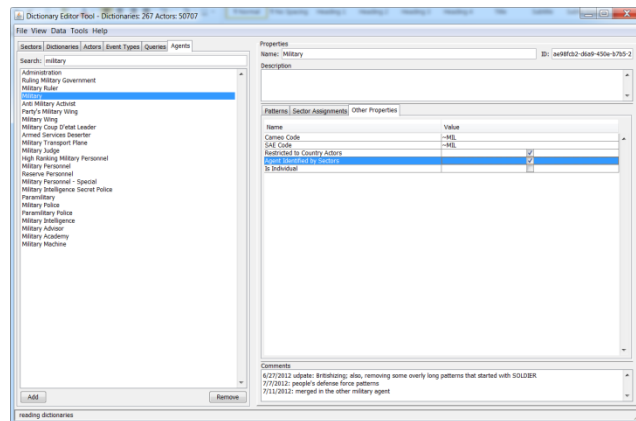


Figure 9 - Dictionary Editor for the Military agent, showing the selection of the Agent Identified by Sectors checkbox.

if

another tab), and the Agent Identified by Sectors checkbox is selected. Wherever an agent-based composite actor for “Military (country)” would normally be used, the sector-based composite actor for “Military (country)” will be used instead. This would lead to a hierarchy within the user interface that looks similar to the depiction below:

- Government (Orangeland)
 - Military (Orangeland)
 - Orangeland National Armed Forces
 - General So-and-So
- ← sector-based composite actor
 ← merged composite actors
 ← named actor
 ← named actor

Just as agent-based composite actors and sector-based composite actors can be merged, so can named actors and sector-based composite actors. Consider the hypothetical actor of Orangeland National Armed Forces, which is assumed for this example to be equivalent to the Military (Orangeland) actor. Within the Dictionary Editor, when assigning a sector that is not date-constrained to an actor, there is the option to ‘Use actor as proxy for the selected sector’ available through a checkbox at the bottom of the “Select a sector” window¹³, as shown in the figure below.

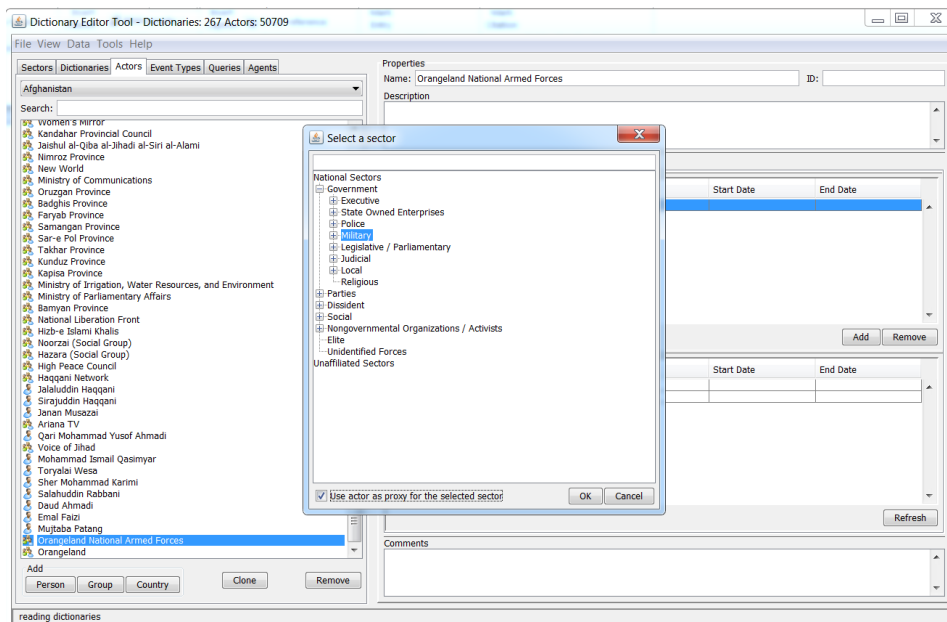


Figure 10 - Dictionary Editor where an actor is defined to use as a proxy for the selected sector.

If both of these modifications were made simultaneously the hierarchy shown in the user interface where actors were listed would look as follows:

¹³ Though nothing in the system prevents someone from using the ‘Use actor as proxy for the selector sector’ checkbox for two actors involving the same sector within the same country, this would be an error condition and will lead to unspecified, but likely erroneous (or at least confusing), results.

- Government (Orangeland)
 - Orangeland National Armed Forces
 - General So-and-So
- ← sector-based composite actor
 - ← merged composite actors and named actor
 - ← named actor

Appendix B: Examples of Aggregation Definitions

One of the simpler aggregations is one that looks at events involving government-affiliated actors as the source, and dissident-affiliated actors as the target, that considers events internal to the country only. This query is illustrated in the screenshot below:

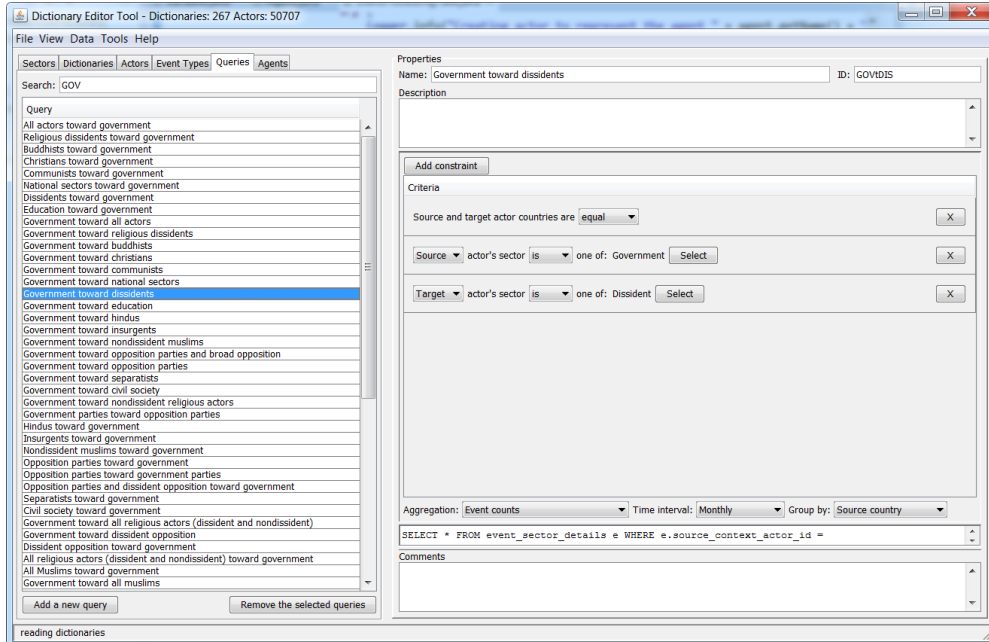


Figure 11 - Dictionary Editor showing a query that considers government-affiliated actors as the source and dissident-affiliated actors as the target, and internal events only.

There are three constraints defined for the query. A country constraint specifying that the source and target countries must be equal restricts the filtered event set to events internal to a country only (i.e., involving a single country's actors as both the source and target). Two sector constraints are used, the first specifying that the source actors have to be government affiliated and the second specifying that the target actors have to be dissident-affiliated. Since sectors are hierarchical, selecting the Government sector will include all of its sub-sectors: Military, Police, etc.

A slight modification to the specification, affecting the source sector, can add to the filtered event set ones that involve not just government-affiliated actors, but the actor representing the country itself, as the source. This modification is shown in Figure 12:

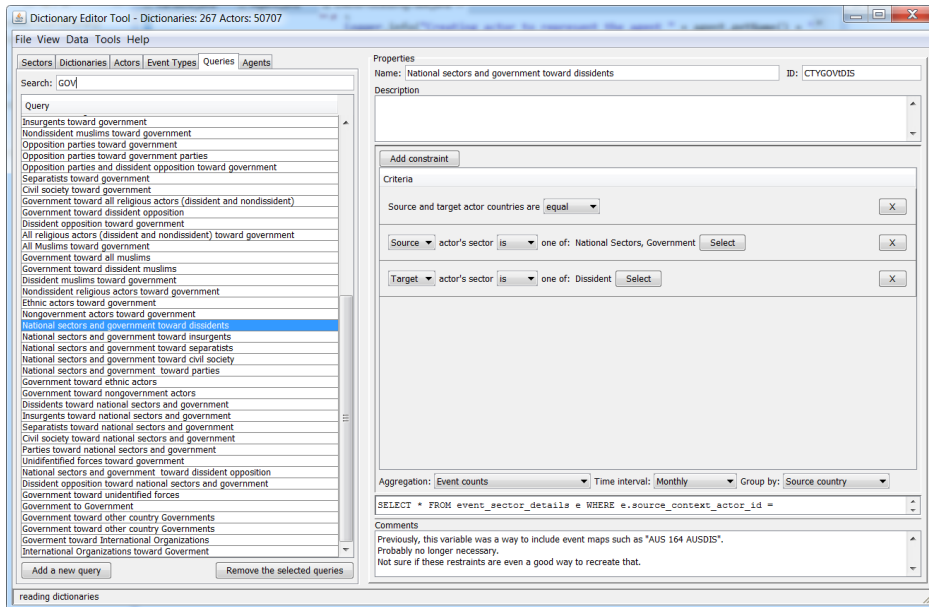


Figure 12 - Dictionary Editor showing a modified aggregation query where the source actor must either be government-affiliated OR the country itself.

Since the lack of sector constraints for the source (or target) means that any event satisfying the country constraint will be included in the filtered event set, simply removing the sector constraint on the source actor as shown in the screenshot below will create a query for all events internal to a country with dissident-affiliated actors as the target:

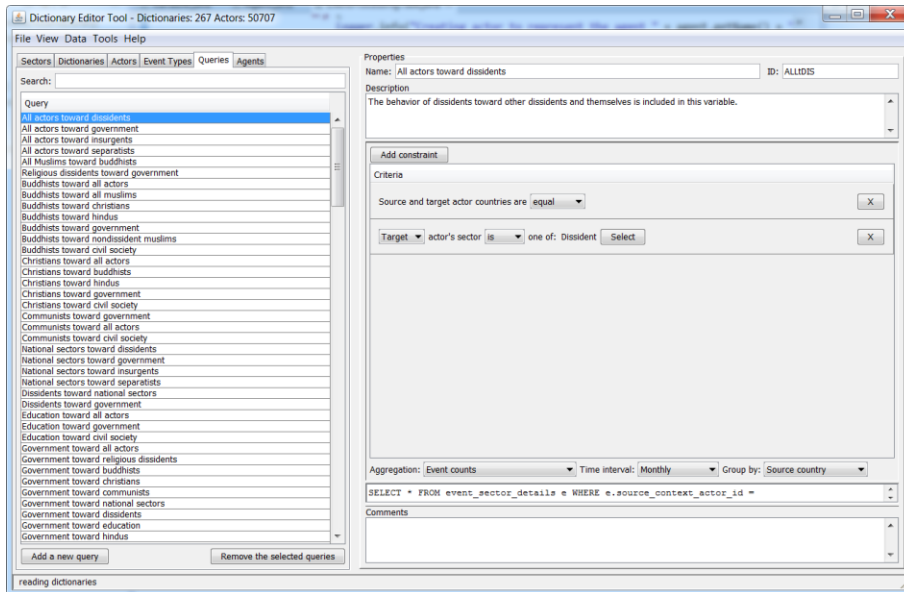


Figure 13 - Dictionary Editor showing an aggregation query for internal events against dissident-affiliated actors.

For aggregations that are focused on international, as opposed to internal, events, use a country constraint where it is specified that the source and target countries must not be equal. An aggregation query that considers events between the governments of two different countries is shown below:

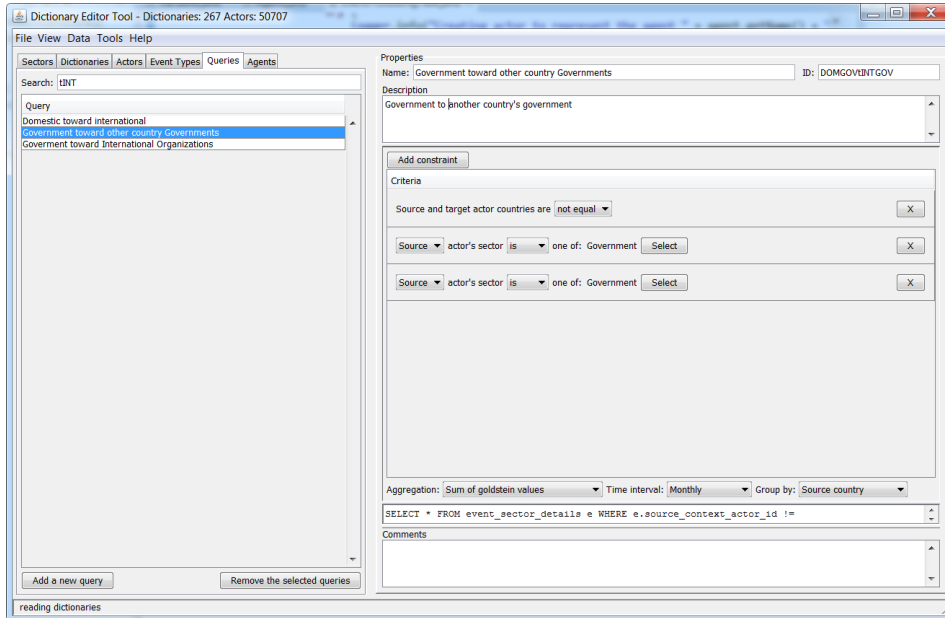


Figure 14 - Dictionary Editor showing an international-focused aggregation query.

Combining multiple sector constraints can create some complex aggregations. The aggregation query below is looking at internal events between Muslims and Christians, but excludes Muslims associated with the government or with dissident groups as well as Christians who are involved with the government.

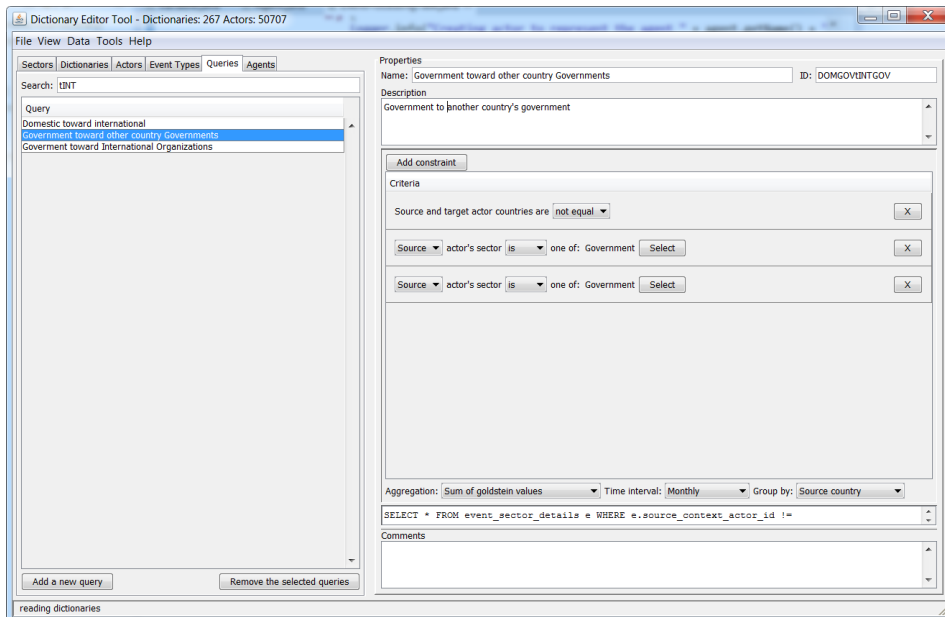


Figure 15 - Dictionary Editor showing a complex aggregation query.

Appendix C: How Monadic Events Sets are initially filtered

Events are initially filtered by dyadic pairs of source and target sectors. The canonical source of this data is in the `queries.xml` used in the ICEWS Dictionary Editor; when this list differs from that file, then that file is the more current list. There are currently 256 dyadic pairs defined, as listed below.

Name	Description
AGTtALL	Agent-based actors toward all actors
AGTtALLMUS	Agent-based actors toward all Muslims
AGTtBADREL	Agent-based actors toward religious dissidents
AGTtBUD	Agent-based actors toward Buddhists
AGTtCHR	Agent-based actors toward Christians
AGTtETH	Agent-based actors toward ethnic actors
AGTtGOV	Agent-based actors toward government
AGTtHIN	Agent-based actors toward Hindu
AGTtMUS	Agent-based actors toward nondissident muslims
AGTtNOTGOV	Agent-based actors toward nongovernment actors
AGTtREL	Agent-based actors toward religious actors
AGTtSOC	Agent-based actors against civil society
AGTtSOCREL	Agent-based actors toward nondissident religious actors
ALLMUSStAGT	All Muslims toward agent-based actors
ALLMUSStALL	All Muslims toward all actors
ALLMUSStBUD	All Muslims toward buddhists
ALLMUSStCHR	All Muslims toward christians
ALLMUSStGOV	All Muslims toward government
ALLMUSStHIN	All Muslims toward hindus
ALLMUSStSOC	All Muslims toward civil society
ALLMUSStUNK	All Muslims toward unspecified actors
ALLtAGT	All actors toward agent-based actors
ALLtALL	All actors towards all actors
ALLtDIS	All actors toward dissidents
ALLtGOV	All actors toward government
ALLtINS	All actors toward insurgents
ALLtSEP	All actors toward separatists
ALLtUNK	All actors toward unspecified actors
BADRELtAGT	Religious dissidents toward agent-based actors
BADRELtGOV	Religious dissidents toward government
BADRELtUNK	Religious dissidents toward unspecified actors
BUDtAGT	Buddhists toward agent-based actors
BUDtALL	Buddhists toward all actors
BUDtALLMUS	Buddhists toward all muslims
BUDtCHR	Buddhists toward christians
BUDtGOV	Buddhists toward government
BUDtHIN	Buddhists toward hindus

BUDtMUS	Buddhists toward nondissident muslims
BUDtSOC	Buddhists toward civil society
BUDtUNK	Buddhists toward unspecified actors
CHRtAGT	Christians toward agent-based actors
CHRtALL	Christians toward all actors
CHRtALLMUS	Christians toward all muslims
CHRtBUD	Christians toward buddhists
CHRtGOV	Christians toward government
CHRtHIN	Christians toward hindus
CHRtMUS	Christians toward nondissident muslims
CHRtSOC	Christians toward civil society
CHRtUNK	Christians toward unspecified actors
COMtALL	Communists toward all actors
COMtGOV	Communists toward government
COMtSOC	Communists toward civil society
CTYGOVtDIS	National sectors and government toward dissidents
CTYGOVtINS	National sectors and government toward insurgents
CTYGOVtPTY	National sectors and government toward parties
CTYGOVtSEP	National sectors and government toward separatists
CTYGOVtSOC	National sectors and government toward civil society
CTYGOVtUNK	National sectors and government toward unspecified actors
CTYtDIS	National sectors toward dissidents
CTYtGOV	National sectors toward government
CTYtINS	National sectors toward insurgents
CTYtPTY	National sectors toward parties
CTYtSEP	National sectors toward separatists
CTYtSOC	National sectors toward civil society
DISStCY	Dissidents toward national sectors
DISStCYGOV	Dissidents toward national sectors and government
DISStDIS	Dissidents towards Dissidents
DISStETH	Dissidents toward ethnic actors
DISStGOV	Dissidents toward government
DISStNOTDIS	Dissidents toward nondissident actors
DISStSOC	Dissidents toward civil society
DISStUAF	Dissidents toward unidentified forces
DOMGOVtINTGOV	Government toward other country Governments
DOMtINT	Domestic toward international
EDUtALL	Education toward all actors
EDUtETH	Education toward ethnic actors
EDUtGOV	Education toward government
EDUtSOC	Education toward civil society
ETHtAGT	Ethnic actors toward agent-based actors
ETHtDIS	Ethnic actors toward dissidents

ETHtEDU	Ethnic actors toward education
ETHtETH	Ethnic towards Ethnic
ETHtGOV	Ethnic actors toward government
ETHtINS	Ethnic actors toward insurgents
ETHtOPP	Ethnic actors toward dissident opposition
ETHtPTY	Ethnic actors toward parties
ETHtREL	Ethnic actors toward religious actors
ETHtSEP	Ethnic actors toward separatists
ETHtSOC	Ethnic actors toward civil society
ETHtUNK	Ethnic actors toward unspecified actors
GOVCTYtGOV	National sectors and government toward government
GOVPTYtOPPPTY	Government parties toward opposition parties
GOVtAGT	Government toward agent-based actors
GOVtALL	Government toward all actors
GOVtALLMUS	Government toward all muslims
GOVtBADREL	Government toward religious dissidents
GOVtBUD	Government toward buddhists
GOVtCHR	Government toward christians
GOVtCOM	Government toward communists
GOVtCTY	Government toward national sectors
GOVtDIS	Government toward dissidents
GOVtEDU	Government toward education
GOVtETH	Government toward ethnic actors
GOVtGOV	Government to Government
GOVtGOVCTY	Government toward national sectors and government
GOVtHIN	Government toward hindus
GOVtINS	Government toward insurgents
GOVtINTORG	Government toward International Organizations
GOVtJUD	Government actors toward Judicial actors
GOVtMED	Government actors toward the media
GOVtNOTGOV	Government toward nongovernment actors
GOVtOPP	Government toward opposition parties and broad opposition
GOVtOPPPTY	Government toward opposition parties
GOVtORGMUS	Government toward dissident muslims
GOVtPTY	Government actors toward political parties
GOVtSEP	Government toward separatists
GOVtSOC	Government toward civil society
GOVtSOCMUS	Government toward nondissident muslims
GOVtSOCREL	Government toward nondissident religious actors
GOVtUAF	Government toward unidentified forces
HINTAGT	Hindus toward agent-based actors
HINTALL	Hindus toward all actors
HINTALLMUS	Hindus toward all muslims

HINtBUD	Hindus toward buddhists
HINtCHR	Hindus toward christians
HINtGOV	Hindus toward government
HINtMUS	Hindus toward nondissident muslims
HINtSOC	Hindus toward civil society
HINtUNK	Hindus toward unspecified actors
INStALL	Insurgents toward all actors
INStCTY	Insurgents toward national sectors
INStCTYGOV	Insurgents toward national sectors and government
INStETH	Insurgents toward ethnic actors
INStGOV	Insurgents toward government
INStSOC	Insurgents toward civil society
INStUAF	Insurgents to unidentified forces
INStUNK	Insurgents toward unspecified actors
INTGOVtDOMGOV	Other country Governments toward Government
INTORGtGOV	International Organizations toward Government
intrapolparty	Parties and dissident opposition toward parties and dissident opposition
INTtDOM	International toward Domestic
MILCTYtMIL	National sectors and military toward military
MILtMIL	Military actors toward military actors
MILtMILCTY	Military toward national sectors and military
MUStAGT	Nondissident muslims toward agent-based actors
MUStALL	Nondissident muslims toward all actors
MUStBUD	Nondissident muslims toward buddhists
MUStCHR	Nondissident muslims toward christians
MUStGOV	Nondissident muslims toward government
MUStHIN	Nondissident muslims toward hindus
MUStSOC	Nondissident muslims toward civil society
MUStUNK	Nondissident muslims toward unspecified actors
nBADREltALL	Religious dissidents toward all actors
nBADREltBADREL	Religious dissidents toward religious dissidents
nBADREltSOCREL	Religious dissidents toward nondissident religious actors
nCTYGOVtOPP	National sectors and government toward dissident opposition
nDIStMILCOP	Dissidents toward military and police
nEXEtPTY	Executive office toward parties
nGOVtOPP	Government toward dissident opposition
nGOVtREL	Government toward all religious actors (dissident and nondissident)
nMILCOPtDIS	Military and police toward dissidents
nMILCOPtOPP	Military and police toward dissident opposition
nMILCOPtSOC	Military and police toward civil society
nOPPtALL	Dissident opposition toward all actors
nOPPtCTYGOV	Dissident opposition toward national sectors and government
nOPPtGOV	Dissident opposition toward government

nOPPtMILCOP	Dissident opposition toward military and police
nOPPtSOC	Dissident opposition toward civil society
nOPPtUAF	Dissident opposition toward unidentified forces
NOTDISTDIS	Nondissident actors toward dissidents
NOTGOVtAGT	Nongovernment actors toward agent-based actors
NOTGOVtALL	Nongovernment actors toward all actors
NOTGOVtGOV	Nongovernment actors toward government
NOTGOVtUNK	Nongovernment actors toward unspecified actors
NOTPTYtPTY	Nonparty actors toward parties
NOTSOctSOC	Non-civil society actors toward civil society
nPTYtALL	Parties toward all actors
nPTYtEXE	Parties toward executive office
nPTYtPTY	Parties toward parties
nRELtGOV	All religious actors (dissident and nondissident) toward government
nRELtREL	All religious actors (dissident and nondissident) toward all religious actors (dissident and nondissident)
nRELtSOC	All religious actors (dissident and nondissident) toward civil society
nSOCREltBADREL	Nondissident religious actors toward Religious dissidents
nSOctMILCOP	Civil society toward military and police
nSOctOPP	Civil society toward dissident opposition
nSOctREL	Civil society toward all religious actors (dissident and nondissident)
nUAFtOPP	Unidifentified forces toward dissident opposition
OPPPTYtGOV	Opposition parties toward government
OPPPTYtGOVPTY	Opposition parties toward government parties
OPPtALL	Opposition parties and dissident opposition toward all actors
OPPtCTY	Opposition toward the country in general (national sectors)
OPPtETH	Dissident opposition toward ethnic actors
OPPtGOV	Opposition parties and dissident opposition toward government
OPPtJUD	Opposition toward Judicial actors
OPPtMED	Opposition toward the media
ORGMUSTALL	Dissident muslims toward all actors
ORGMUSTGOV	Dissident muslims toward government
ORGMUSTSOC	Dissident muslims toward civil society
POLCTYtPOL	National sectors and politicians toward politicians
POLtPOL	Politician to politician
POLtPOLCTY	Politicians toward national sectors and politicians
PTYtCTY	Parties toward national sectors
PTYtCTYGOV	Parties toward national sectors and government
PTYtETH	Parties toward ethnic actors
PTYtGOV	Parties toward Government
PTYtJUD	Parties toward Judicial actors
PTYtMED	Parties toward Media
PTYtNOTPTY	Parties toward nonparty actors

PTYtUAF	Parties toward unidentified parties
religinfight	Nondissident religious actors and dissident muslims toward nondissident religious actors and dissident muslims
RELtAGT	Religious actors toward agent-based actors
RELtETH	Religious actors toward ethnic actors
RELtUNK	Religious actors toward unspecified actors
SEPtALL	Separatists toward all actors
SEPtCTY	Separatists toward national sectors
SEPtCTYGOV	Separatists toward national sectors and government
SEPtETH	Separatists toward ethnic actors
SEPtGOV	Separatists toward government
SEPtSOC	Separatists toward civil society
SEPtUAF	Separatists toward unidentified forces
SEPtUNK	Separatists toward unspecified actors
SOCRELtAGT	Nondissident religious actors toward agent-based actors
SOCRELtALL	Nondissident religious actors toward all actors
SOCRELtDIS	Nondissident religious actors toward dissidents
SOCRELtGOV	Nondissident religious actors toward government
SOCRELtINS	Nondissident religious actors toward insurgents
SOCRELtSEP	Nondissident religious actors toward separatists
SOCRELtSOC	Nondissident religious actors toward civil society
SOCRELtUNK	Nondissident religious actors toward unspecified actors
SOctAGT	Civil society toward agent-based actors
SOctALL	Civil society towards all actors
SOctALLMUS	Civil society toward all muslims
SOctBUD	Civil society toward buddhists
SOctCHR	Civil Society toward christians
SOctCTY	Civil society toward national sectors
SOctCTYGOV	Civil society toward national sectors and government
SOctDIS	Civil society toward dissidents
SOctETH	Civil society toward ethnic actors
SOctGOV	Civil society toward government
SOctHIN	Civil society toward hindus
SOctINS	Civil society toward insurgents
SOctJUD	Social actors toward Judicial actors
SOctMED	Social actors toward the media
SOctMUS	Civil society toward nondissident muslims
SOctNOTSOC	Civil society toward Non-civil society actors
SOctORGMUS	Civil society toward dissident muslims
SOctSEP	Civil society toward separatists
SOctSOC	Civil society toward civil society
SOctUAF	Civil society toward unidifentified forces
SOctUNK	Civil society toward unspecified actors

UAFtDIS	Unidified forces toward dissidents
UAFtGOV	Unidified forces toward government
UAFtINS	Unidified forces toward insurgents
UAFtPTY	Unidified forces toward parties
UAFtSEP	Unidified forces toward separatists
UAFtSOC	Unidified forces toward civil society

Appendix D: How Filtered Event Sets (monadic and dyadic) are currently aggregated

Where how event sets are initially filtered is defined in the `queries.xml` file, and editable in the Dictionary Editor, the post-filtering and aggregation of these event sets are defined in the ICEWS system code. There are 44 types of aggregations that are calculated for each aggregation query, as listed below.

Name	Description
eventstotal	Summing the intensity values of all events in the filtered event set
eventsct	Counting all events in the filtered event set
cooptotals	Summing the intensity values of all events in the filtered event set with a positive value
hosttotals	Summing the intensity values of all events in the filtered event set with a negative value
coopct	Counting all events in the filtered event set that have a positive intensity value
hostilityct	Counting all events in the filtered event set that have a negative intensity value
highhostilityct	Counting all events in the filtered event set that have an intensity value ≤ -8
highhostotals	Summing the intensity values of all events in the filtered event set with an intensity value ≤ -8
medhostilityct	Counting all events in the filtered event set that have an intensity value ≤ -4 and > -8
medhostotals	Summing the intensity values of all events in the filtered event set with an intensity value ≤ -4 and > -8
lowhostilityct	Counting all events in the filtered event set that have an intensity value < 0 and > -4
lowhostotals	Summing the intensity values of all events in the filtered event set with an intensity value < 0 and > -4
highcoopct	Counting all events in the filtered event set that have an intensity value ≥ 7
highcooptotals	Summing the intensity values of all events in the filtered event set with an intensity value ≥ 7
medcoopct	Counting all events in the filtered event set that have an intensity value ≥ 5 and < 7
medcooptotals	Summing the intensity values of all events in the filtered event set with an intensity value ≥ 5 and < 7
lowcoopct	Counting all events in the filtered event set that have an intensity value > 0 and < 5
lowcooptotals	Summing the intensity values of all events in the filtered event set with an intensity value > 0 and < 5
neutralct	Counting all events in the filtered event set that have an intensity value $= 0$
coopscaleav	Calculating the average intensity of all events in the filtered event set that have an intensity value > 0
hosscaleav	Calculating the average intensity of all events in the filtered event set that have an intensity value < 0
violencect	Counting all events in the filtered event set with an event type corresponding to the CAMEO categories of 17 (COERCE), 18 (ASSAULT), 19 (FIGHT), or 20 (USE CONVENTIONAL MASS VIOLENCE)
protestct	Counting all events in the filtered event set with an event type corresponding to the CAMEO categories of 14 (PROTEST)
matcoopct	Counting all events in the filtered event set with an event type corresponding to the CAMEO categories of 06 (ENGAGE IN MATERIAL COOPERATION), 07 (PROVIDE AID),

	08 (YIELD)
verbalcoopct	Counting all events in the filtered event set with a positive intensity value that are NOT of an event type corresponding to the CAMEO categories of 06 (ENGAGE IN MATERIAL COOPERATION), 07 (PROVIDE AID), 08 (YIELD)
matconflictct	Counting all events in the filtered event set with an event type corresponding to the CAMEO categories of 14 (PROTEST), 15 (EXHIBIT FORCE POSTURE), 16 (REDUCE RELATIONS), 17 (COERCE), 18 (ASSAULT), 19 (FIGHT), or 20 (USE CONVENTIONAL MASS VIOLENCE)
verbalconflictct	Counting all events in the filtered event set with a negative intensity value and an event type NOT corresponding to the CAMEO categories of 14 (PROTEST), 15 (EXHIBIT FORCE POSTURE), 16 (REDUCE RELATIONS), 17 (COERCE), 18 (ASSAULT), 19 (FIGHT), or 20 (USE CONVENTIONAL MASS VIOLENCE)
demandct	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 10 (DEMAND)
disapprovect	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 11 (DISAPPROVE)
rejectct	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 12 (REJECT)
threatenct	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 13 (THREATEN)
posturect	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 15 (EXHIBIT FORCE POSTURE)
reducereactionsct	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 16 (REDUCE RELATIONS)
coercect	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 17 (COERCE)
assaultct	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 18 (ASSAULT)
fightct	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 19 (FIGHT)
massviolencect	Counting all events in the filtered event set with an event type corresponding to the CAMEO category of 20 (USE CONVENTIONAL MASS VIOLENCE)
arrestct	Counting all events in the filtered event set with an event type corresponding to the CAMEO code of Arrest (CAMEO code 173)
hostscaled	Calculating the standard deviation of the intensities of all events in the filtered event set that have an intensity value < 0
coopscaled	Calculating the standard deviation of the intensities of all events in the filtered event set that have an intensity value > 0
eventscaled	Calculating the standard deviation of the intensities of all events in the filtered event set
protestdays	Counting the number of days on which one or more events in the filtered event set with an event type corresponding to the CAMEO categories of 14 (PROTEST) occurred. Will range from 0 to 31 (or max number of days in a specific month).
eventsscaleav	Calculating the average intensity of all events in the filtered event set
VEGroundTruth	Counting all events in the filtered event set with an event type indicating a violent event

Appendix E: How Dyadic Events Sets are initially filtered

Events are initially filtered by dyadic pairs of source and target sectors, as well as by source and target country. The canonical source of this data is a subset of the `queries.xml` used in the ICEWS Dictionary Editor; when this list differs from that file, then that file is the more current list. We have generated dyadic data for the same dyadic pairs as the monadic data, as given in Appendix C.

Appendix F: Province-Level Aggregation List

Country	Province	Events in Past Year	Events in Past 3 years
Egypt	Al Sharqia	101	192
Egypt	Alexandria	891	2802
Egypt	Aswan	130	312
Egypt	Asyut	91	363
Egypt	Beheira	97	243
Egypt	Beni Suef	9	42
Egypt	Cairo	19468	68462
Egypt	Dakahlia	266	522
Egypt	Damietta	41	104
Egypt	Faiyum	121	257
Egypt	Gharbia	79	297
Egypt	Giza	901	2095
Egypt	Ismailia	116	324
Egypt	Kafr el-Sheikh	12	41
Egypt	Luxor	101	357
Egypt	Matrouh	47	152
Egypt	Minya	287	413
Egypt	Monufia	5	28
Egypt	New Valley	6	35
Egypt	North Sinai	573	2096
Egypt	Port Said	101	1342
Egypt	Qalyubia	74	197
Egypt	Qena	67	310
Egypt	Red Sea	40	146
Egypt	Sohag	55	173
Egypt	South Sinai	217	510
Egypt	Suez	435	1553
Iraq	Al Anbar	2927	4353
Iraq	Al Qadisiyah	2	42
Iraq	Babil	77	170
Iraq	Baghdad	9383	27058
Iraq	Basra	155	634
Iraq	Dhi Qar	48	113
Iraq	Diyala	599	1804
Iraq	Dohuk	114	344
Iraq	Erbil	1693	4056
Iraq	Karbala'	144	545
Iraq	Kirkuk	502	1925
Iraq	Maysan	32	98

Iraq	Muthanna	3	21
Iraq	Najaf	150	353
Iraq	Nineveh	2465	3909
Iraq	Saladin	1539	2426
Iraq	Sulaymaniyah	633	1525
Iraq	Wasit	9	55
Jordan	Ajloun	16	37
Jordan	Amman	3327	13378
Jordan	Aqaba	147	320
Jordan	Balqa	25	70
Jordan	Irbid	89	441
Jordan	Jerash	12	51
Jordan	Karak	12	43
Jordan	Maan	1940	6593
Jordan	Madaba	11	63
Jordan	Mafraq	41	354
Jordan	Tafilah	12	190
Jordan	Zarqa	41	145
Libya	Al Wahat	96	153
Libya	Benghazi	2229	5741
Libya	Butnan	293	343
Libya	Derna	153	190
Libya	Ghat	4	9
Libya	Jabal al Akhdar	15	33
Libya	Jabal al Gharbi	97	477
Libya	Jafara	0	0
Libya	Jufra	0	0
Libya	Kufra	1	3
Libya	Marj	1	1
Libya	Misrata	152	394
Libya	Murqub	7	33
Libya	Murzuq	3	6
Libya	Nalut	17	31
Libya	Nuqat al Khams	4	17
Libya	Sabha	147	283
Libya	Sirte	97	329
Libya	Tripoli	4087	11046
Libya	Wadi al Hayaa	13	25
Libya	Wadi al Shatii	0	0
Libya	Zawiya	41	80
Saudi Arabia	Al Bahah	0	1
Saudi Arabia	Al Jawf	0	6
Saudi Arabia	Al Madinah	30	124

Saudi Arabia	Al Qasim	8	146
Saudi Arabia	Al Riyad	3095	9055
Saudi Arabia	Asir	0	0
Saudi Arabia	Eastern Province	71	732
Saudi Arabia	Ha'il	0	0
Saudi Arabia	Jizan	20	51
Saudi Arabia	Makkah	1400	6100
Saudi Arabia	Najran	3	32
Saudi Arabia	Northern Borders	1	5
Saudi Arabia	Tabuk	3	32
Syria	Al-Hasakah	230	1146
Syria	Al-Raqqah	378	677
Syria	Al-Suwayda	55	264
Syria	Aleppo	2516	10258
Syria	Damascus	6240	46522
Syria	Daraa	217	1237
Syria	Deir ez-Zor	188	678
Syria	Hama	261	1614
Syria	Homs	1219	7001
Syria	Idlib	579	3049
Syria	Latakia	649	1753
Syria	Quneitra	176	336
Syria	Rif Dimashq	1096	3038
Syria	Tartus	77	832
Yemen	Abyan	348	2634
Yemen	Adan	953	3170
Yemen	Al Bayda'	86	232
Yemen	Al Dali	71	151
Yemen	Al Hudaydah	84	234
Yemen	Al Jawf	11	21
Yemen	Al Mahrah	2	8
Yemen	Al Mahwit	4	9
Yemen	Amanat Al Asimah / Sana'a City	39	114
Yemen	Amran	38	41
Yemen	Dhamar	145	382
Yemen	Hadhramaut	595	1293
Yemen	Hajjah	44	143
Yemen	Ibb	0	0
Yemen	Lahij	49	114
Yemen	Ma'rib	107	507
Yemen	Raymah	0	3
Yemen	Sada	238	536

Yemen	Sana'a (Governorate)	5509	15308
Yemen	Shabwah	343	654
Yemen	Ta`izz	139	955
Occupied Palestinian Territories	Gaza Strip	18184	38741
Occupied Palestinian Territories	West Bank	7920	22116

Appendix G: Glossary

Actor. In an *Event*, the group, individual, or location that is involved in the “*who*” or “*whom*” role. *Actors* may serve as the *Source* or *Target* of an *Event*. In grammar terms, it would be a noun.

Actor Affiliation. An association between two actors, where one of the two actors must be a group or country. An individual who is a member of a group (e.g., Osama bin Laden as a member of Al Qaeda) would have an affiliation between themselves as an individual actor, and the group as a group actor.

Actor Dictionary. A set of XML files that define the *Actors*, *Agents*, *Event Types*, *Sectors*, and *Aggregation Queries* in use in the ICEWS system.

Actor Membership. An alternate term for *actor affiliation*.

Actor Pattern. A *Pattern* associated with a specific *Actor*.

Actor Type. *Actors* are classified as one of four types: *Individual*, *Group*, *Country*, or *Location*.

Affiliation. See *Country Affiliation* or *Sector Affiliation*.

Agent. A generic *individual* or *group* that can be combined with one of multiple *country* or *country-derived actors* into a *composite agent-based actor*. Examples of agents include Fishermen, Military, or Democratic Party. In grammar terms, an agent is an improper noun.

Agent-Based Composite Actor. See *Composite Agent-Based Actor*.

Agent-Based Sector. A special sector affiliation used to identify actors that are *Composite Agent-Based Actors*.

Agent Pattern. A *Pattern* associated with a specific *Agent*.

Aggregating. Taking a *country-affiliated filtered event set* that corresponds to a particular *time interval* and calculating a numeric value based on it.

Aggregation. See Event Aggregation.

Autonomous Region. A geopolitical region within a country that has a degree of autonomy, or has freedom from an external authority.

Aggregation Query. A specification for forming a *filtered event set* for aggregation purposes via a *country constraint* and zero or more *sector constraints*.

CAMEO. Conflict and Mediation Event Observations. A coding scheme for event data developed initially at Kansas State University.

CAMEO Category. In the *CAMEO* coding scheme, *CAMEO codes* are organized into one of twenty high-level CAMEO categories.

CAMEO Code. In the *CAMEO* coding scheme, CAMEO codes are used as a short-hand method to identify types of events. There are 312 CAMEO codes in the CAMEO version that *Event Types* in *ICEWS* are based on, and they are arranged in twenty *CAMEO Categories*.

Category. See *CAMEO Category*.

Composite Actor. An *Actor* that is not listed in the *Actor Dictionary* but is instead formed dynamically within the *ICEWS* system by combining a *Sector* with a *Country* to form a *composite sector-based actor*, or by combining an *Agent* with a *named group* or *country actor* to form a *composite agent-based actor*.

Composite Agent-Based Actor. A *composite actor* that is formed dynamically by the *ICEWS* system by combining an *agent* with a *group* or *country actor*.

Composite Sector-Based Actor. A *composite actor* that is formed dynamically by the *ICEWS* system by combining a *sector* with a *country*.

Constraint. Also known as a *criteria*, a restriction on some aspect of an event, such as its *source* or *target*, for the purpose of generating a *filtered event set*.

Country. A region that represents a distinct political entity, either an independent sovereign state or a non-sovereign state that is officially recognized by the US.

Country Affiliation. The country affiliation of an *actor* is the country (or, in some instances, the multiple countries) that the actor is closely affiliated with. For *aggregations*, the country affiliation is the affiliation of the *source* or *target* of all events in the *filtered event set* that underlies the aggregation; whether it is the source or target actor's affiliation is part of the aggregation definition.

Country-Based Constraint. A *constraint* placed on the *target* or *source* of an *event* based on the *country affiliation* of the *actor* in question.

Country-derived. An *actor* is said to be country-derived if it has a *membership* to a country (perhaps in context with a *sector*), or with some *actor* that is in turn country-derived.

Criteria. Another term for a *Constraint*.

Dictionary Editor. A graphical user interface that is used to edit the set of files making up an *Actor Dictionary*.

Dyadic Event Aggregation. The association of numerical data derived from event data between two countries for some time interval.

Dyadic Pair. The part of an *aggregation query* that specifies the *source* and *target* actors that are found in *events* of the *filtered event set*.

Event. Information about *who* did *what* to *whom*, *when*, and *where*. The “*who*” and “*whom*” are the *Source* and *Target* respectively, the “*what*” is an *Event Type*, the “*when*” is a date, and the “*where*” is a *Location*.

Event Aggregation. Event aggregations may be either *Dyadic Event Aggregations* or *Monadic Event Aggregations*. If not specified, the assumption is that the monadic form of event aggregations is being discussed.

Event Intensity. A value ranging from -10 to +10, used to express the level of cooperation or hostility exhibited in the *Event Type* it is associated with. Cooperative events have a positive value while hostile events have a negative value, with values toward the edges of the range expressing a greater degree of hostility or cooperation and following a roughly linear scale for other values. Event intensities are expressed from a neutral point of view.

Event Query. See *Aggregation Query*.

Event Set. A collection of one or more *events* that are considered as a group.

Event Type. In an *Event*, the “*what*” that occurs. It is an action that occurs between a *Source* and a *Target*. In grammar terms, it would be an action verb.

Explicit (sector) membership. A *sector membership* that is explicitly and directly defined with the actor in question within the *Actor Dictionary*.

Filtered Event Set. An *event set* that has been *filtered* such that it satisfies a set of *constraints*.

Filtering. Applying a set of *constraints* to an *event set*, typically as a precursor to *aggregating* them.

Goldstein value. An alternate term for *Event Intensity*.

Group. An *actor type* that corresponds to multiple individuals with a common and publicly professed identity.

iCAST. The part of the ICEWS system that is concerned with forecasting aspects of country stability.

ICEWS. Integrated Crisis Early Warning System, an analytical software system developed by Lockheed Martin.

Implicit (sector) membership. A *sector membership* which is not specified explicitly in the **Actor Dictionary**, but is instead inherited through the transitive nature of sector and actor memberships.

Individual. An *actor type* that corresponds to a single person.

Intensity value. An alternate term for *Event Intensity*.

Internal Event. An *event* that occurs in the context of a single country, such that the *country affiliations* of the *source* and *target* resolve to the same country.

International Event. An *event* where the *source* and *target country affiliations* differ.

International Sector. See *unaffiliated sector*.

iTRACE. The part of the ICEWS system that is concerned with historical event trends.

Location. The “*where*” part of an *event*. A location may be a country, province, district, city, location within a city, or region of multiple countries.

Membership. See Sector Membership.

Monadic Event. An event where the target (or perhaps in the future, the source) is unspecified, and represented by the Unspecified Actor.

Monadic Event Aggregation. The association of numerical data derived from event data with a given country for some time interval.

Named Actor. A non-composite actor that is listed in the Actor Dictionary. Named actors must have a proper name used to identify them.

National Sector. 1. Most generally, a *sector* that is classified as existing only within the context of a *country* or *country-derived actor*. 2. In *event queries*, the national sector is a special sector that indicates the actor must be a country (i.e., have an *actor type* of country).

Paired Actor Reference. In a *composite actor*, the *actor* with which the *sector* or *agent* is combined.

Paired Agent Reference. In a *composite agent-based actor*, the *agent* with which an *actor* is combined.

Pattern. A sequence of letters and underscores used to define how an actor or agent is represented in news stories. Though there are nuances beyond the scope of this document, they can be thought of as synonyms for an actor or agent, where spaces are replaced with underscores and the case (e.g., uppercase, lowercase) of letters do not matter.¹⁴

Query. See *Aggregation Query*.

SAE Code. Historically speaking, agents within the ICEWS system were based on an agent dictionary that had what were used as SAE Codes to represent them. These codes have been preserved for agents, for historical purposes.

Sector. A general role taken on by an actor or agent, sometimes in a time-constrained context. Examples of sectors include Government, Muslim, Dissident, and Refugees.

Sector Affiliation. An association between an *actor* or *agent* and a *sector*. Sector affiliations may either be explicit, in which case they are defined in the *Actor Dictionary*, or implicit, in which they are derived from the transitive nature of sector and actor affiliations.

Sector-Based Composite Actor. See *Composite Sector-Based Actor*

Sector-Based Constraint. A *constraint* placed on the *target* or *source* of an *event* based on the *sector affiliation* of the *actor* in question

Sector Membership. Used interchangeably with *sector affiliation*.

Source. In an *Event*, the *Actor* that is the instigator of the action.

Target. In an *Event*, the *Actor* that is the recipient of the action.

Time Interval. A unit of time, such as a month or a week. Time intervals may either be general (e.g., 'a week') or specific (e.g., 'the first week of 2006').

Unaffiliated Sector. A *sector* that is not constrained to the context of a specific country. Unaffiliated sectors are also known as *international sectors* due to their lack of country affiliation.

Unspecified Actor. A special actor, named 'Unspecified Actor', which is used as a placeholder when the specific target (or, in the future, perhaps the source) of an event is uncertain. See also *Monadic Event*.

Unspecified Sector. A special *international sector*, named 'Unspecified Sector', to which the *Unspecified Actor* (and only this actor) has an affiliation.

¹⁴ ICEWS patterns are a close variant of TABARI patterns, as explained in the TABARI manual: <http://eventdata.psu.edu/tabari.dir/tabari.manual.060228.pdf>