

ICEWS Dictionaries Read Me

The ICEWS dictionaries contain both named individuals or groups, known as actors, and generic individuals or groups, known as agents. Actors are known by a specific name, such as 'Free Syrian Army' or 'Goodluck Johnathan', while agents are known by a generic improper noun, such as 'insurgents' or 'students'. Both actors and agents have time-dependent affiliations with another actor (in the case of an individual being a member of an organization, for example), a country or other autonomous region, or with a general sector/role, such as 'Military' or 'Government'. Also included in the dictionaries are aliases that an actor or agent might be known by. In the case of actors, these are typically alternate spellings of a person's name, while for agents they are typically synonyms.

The data is in CSV (comma-separated value) format, with fields optionally enclosed by quotation marks and always separated by commas. There are three separate CSV files: one for the sector hierarchy, one for the agent dictionary, and one for the actor dictionary.

Sector Hierarchy

The sector hierarchy consists of a tree of sector names, stored in a CSV file, with one name per row. There is no header row for this file.

Blank fields are used to 'indent' sector names, to express the hierarchical structure. For example, in the snippet shown below, 'National Sectors' is the top-level sector name. A child of 'National Sectors' is the 'Government' sector, and its child, in turn, is called 'Executive'. There are two children of the 'Executive' sector shown: 'Executive Office', and 'Agriculture / Fishing / Forestry Ministry'.

National Sectors					
	Government				
		Executive			
			Executive Office		
			Agriculture / Fishing / Forestry Ministry		

There are no definitions or descriptions supplied for the sectors, though they carry the typical meaning that their English-language name would suggest.

Agent Dictionary

The first row of the CSV file is the header row, giving the field names. Subsequent rows contain data. Each subsequent row corresponds to a single agent in the dictionary.

Column Number	Column Name	Description
1	Agent Name	The name of the agent. This will be referenced in the ICEWS Event Data.
2	Agent Type	One of two values: 'Individual' or 'Group'. This simply expresses

		whether the agent is singular or plural in nature.
3	Sectors	One or more sectors which are affiliated with the agent. Multiple sectors are delimited with a ' ' symbol.
4	Aliases	One or more aliases, or synonyms, by which the agent may be known. Multiple aliases are delimited with a ' ' symbol. These aliases are, for the most part, expressed in English. However, a TABARI-like syntax is used for the purpose of pattern-matching and stemming rules in some cases. ¹

Actor Dictionary

The first row of the CSV file is the header row, giving the field names. Subsequent rows contain data. Each row contains information about an actor, though a single actor may have multiple rows that express data about it. More specifically, there is one row for each sector affiliation that an actor holds.

Column Number	Column Name	Description
1	Country	Name of the country or territory with which the actor is primarily affiliated, in most cases. There are special 'non-country' values that may also be listed: 'International', 'International Governmental Organizations', 'International Militarized Groups', 'Multi-National Corporations', 'Non-Governmental Organizations'. These country designations are for organizational purposes only. For true country affiliation, you will need to reference the sector affiliations of the actor.
2	Actor Name	The name of the actor, as used in the ICEWS system. In particular, this name can be used as the 'Source' or 'Target' of an ICEWS event.
3	Actor Type	One of two values: 'Individual' or 'Group'. This simply expresses whether the actor is singular or plural in nature.
4	Affiliation To	One of two possibilities: <ol style="list-style-type: none"> 1. A sector with which an actor is affiliated. As sectors are hierarchical, the actor will also be affiliated with all ancestor sectors, up to the root sector of the hierarchy. 2. Another group (actor) with which the actor is affiliated. For example, 'Abdul Salam Zaef' has an affiliation with 'Taliban', as he was a member of the Afghan Taliban. As another example, 'Abdul Haq Shafaq' has an affiliation with 'Faryab Province' as he was once governor of this Afghan province. When an actor is affiliated with a group, they inherit all sector affiliations of that group during the appropriate time ranges.
5	Affiliation Start Date	The date on which the affiliation began, inclusive. If the affiliation should be considered valid for all dates up until the 'Affiliation End Date', then a value of 'beginning of time' will be used.

¹ <http://eventdata.parusanalytics.com/tabari.dir/tabari.manual.0.7.6.pdf>, Section 5.2.2 (Stemming)

6	Affiliation End Date	The date on which the affiliation ends, inclusive. If the affiliation should be considered valid for all dates after the 'Affiliation Start Date', then a value of 'end of time' will be used.
7	Aliases	One or more aliases, or alternate spellings, by which the actor may be known. Multiple aliases are delimited with a ' ' symbol.

Additional Available Documentation

The following additional documents are available on the Dataverse Study page for this data.

Document	Description
ICEWS Events and Aggregations.pdf	Section 1 of this document explains the concepts related to ICEWS event data. Also of potential interest is the glossary in Appendix F.