

Documentation for the PLOVER gold standard file

PLOVER_GSR_CAMEO_readme.pdf

This collection of records is the first attempt to develop gold-standard records for the PLOVER event ontology. It is based on examples extracted from the CAMEO manual¹ with considerable editing, since CAMEO coded only actors and events. The records are coded with the version of PLOVER specification that was current as of the **dateCoded** field: PLOVER is still under development and some of these may change in the future. The file conforms to JSON standards and should be readable with any JSON utility.

It is important to note that the examples in the manual were originally human coded, and from the standpoint of automated coding, they are ideal codings and no known system would be likely to get all of them. There are **comments** are some of the more problematic cases.

The sentences used in the CAMEO manual were based on actual news stories around the turn of the century, but many have been edited, sometimes multiple times, and do not necessarily refer to actual historical events. The original CAMEO examples have been further edited so that all of the nation-states correspond to entities with names and codes in the current ISO-3166-alpha-3 entry in Wikipedia.² Ethnicity codes correspond to ISO-639-2.

Funding for PLOVER has been provided in part by the U. S. National Science Foundation award SBE-1539302, “RIDIR: Modernizing Political Event Data for Big Data Social Science Research.”

Additional comments on the coding

1. Only non-null JSON fields are included. All **date** fields are set to ‘2000-01-01.’ The current version does not include any **mode** or **context** fields but these are retained as placeholders.
2. City names are resolved as follows:
 - Cities in general go to **sector CVL**
 - Capital cities go to **sector GOV** if they are the subject or direct object, otherwise **sector CVL**
3. Some **eventLoc** fields have been filled in, usually on the basis of [human-identified] prepositional phrases. In some cases, the actor primary codes were inferred from

¹<http://eventdata.parusanalytics.com/data.dir/cameo.html>

²https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3

these locations.

Text Markup

The **textInfo/markup** field shows where the various text fields occur in the sentence. The following markers are used:

SRC	source actor and agent
TAR	target actor and agent
SRC/TAR:	Reciprocal source/target in the CAMEO system
EVT	event
LOC	location

This markup has been done through a combination of automatic translation from the original L^AT_EX highlighting in the CAMEO manual and additional manual editing. It sometimes gets a bit dodgy when there are multiple words in the phrases, but probably is close enough to be useful

Provenance

Location for this file and the data: <https://github.com/openeventdata/PLOVER>

The GSR file (and documentation) is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

For further information contact: schrodt735@gmail.com

Last update: December 16, 2016

Copyright ©2016 by the Open Event Data Alliance