

MID Source Data, Version 2.1

Codebook

July 30, 2009

Overview

This dataset includes partial data on the sources used to identify and code MIDs for version 2.1 of the MID dataset, which identified 2150 MIDs between 1816 and 1992. Users of the MID data frequently ask for reference information to obtain more information about particular MIDs. Until 2009, the COW project was unable to provide this information because it had not been compiled/recorded in a usable electronic form. This data set now provides some information on the original reference materials used to code MIDs through 1992 (some source information is still missing for reasons described below).

Sources for MIDs collected as part of the MID 3.0 data collection are kept separately.

Historical Background and Data Development

Data collection for the MID 2 data project occurred from the late 1970s through the mid-1990s. During this time, only limited information concerning sources was kept. For each MID, coders saved the information about sources only in the form of a short abbreviation, typically in the form of a periodical title (e.g., “New York Times” “Facts on File”) or the last name of an author (e.g., “Langer” “Stavrianos” “Sure”). For periodicals, more detailed information such as page numbers, edition, or publication dates was not recorded. For books used as references, titles were not kept in the database associated with the 2.1 data. In many cases, this means that the only identification of a source that we have is a last name; these names are not always clearly identifiable as corresponding to a particular source book.

Separately, the Correlates of War of project kept a physical card catalog of MID bibliography references. On each of these physical cards, full bibliography information was recorded. In 2008-2009, the COW project attempted to match the short source references for each MID with the bibliography references saved on physical cards. This match/merge successfully identified at least one source for 1853 MIDs, but no match was found for any reference for 297 MIDs. The matching process was complicated by several facts:

1. Many names on particular MIDs had no match in the bibliography/card catalog;
2. Some listed author last names were very common (e.g. “Smith”), and so it was impossible to identify a matching source when there was not a clear match in the bibliography;
3. Some listed author last names matched several bibliography references (in these cases, all possible matches are listed in the data);
4. The formatting of the 5000+ records in the data set was not consistent, in some cases multiple authors were identified, in other cases only one author, and it was unclear if multiple names listed were identifying separate sources, or only one.

Coders began by matching the short source references associated with MID numbers with bibliography entries, looking for exact matches as well as close matches (to deal with misspellings). In some cases coders looked up and identified further references where we thought it clear what book by a particular author was used. However, the data is still incomplete, with 297 of 2150 MIDs not having a matching source. In most of these cases, we have an author name, but this is it.

In addition to the database matching MID numbers with sources, a separate file is provided which includes two lists. The first list is of abbreviations used in the source data (e.g. “NYT” is “New York Times”). The second list is of references included in the MID bibliography card catalog that had no identifiable matches to specific MID numbers. These works may have been consulted as general references, may have been used to identify specific MIDs, or may have been consulted and recorded even though they did not identify any specific MIDs, but we do not know – that information is not available.

The data is organized by MID number. There may be several references listed for any given MID.

Future Updates

If users of this source data are able to identify sources that correspond to the short codes in the dataset, or identify pages or dates where information may be found, we encourage them to compile this information and email it to the COW project for updating of this source dataset.

Associated Files

MID2.1_SpecificSources.csv	The file containing source data is in comma-delimited format
MID2.1_GeneralSources.csv	The file (comma-delimited format) containing the full names for sources which are abbreviated in the source data file, and sources contained in the COW MID bibliography that do not correspond to specific MID numbers

Variables

MID	MID number
Source	Original source name as recorded in the MID v2.1 data collection. Many source names are abbreviated, so see “MID 2.1 General Sources.xls “ for full names
Author	Author name of work
Year	Year of publication
Title	Title of book
Publisher	Publisher of work
CallNum	Library of Congress call number

Notes	Remarks regarding a particular source
-------	---------------------------------------