

THE GDELТ HUMAN RIGHTS GLOBAL KNOWLEDGE GRAPH (HR-GKG) V1.0
EXECUTIVE SUMMARY
9/22/2014
<http://gdeltproject.org/>

INTRODUCTION

The GDELТ Human Rights Global Knowledge Graph (HR-GKG) is the first in a series of special collections expanding GDELТ beyond its traditional realm of mainstream news media. This particular special collection is designed to codify a cross-section of human rights documentation covering the last few decades, offering a rich and highly specialized dataset to explore global human rights issues through the quantitative lens of the GDELТ Global Knowledge Graph system, as well as the ability to further contextualize those findings through the broader news-based GDELТ ecosystem.

In total 110,552 documents were processed to create this specialized Global Knowledge Graph: 65,731 from Amnesty International, 5,255 from FIDH, 24,791 from Human Rights Watch, 709 from the International Criminal Court, 1,333 from the International Crisis Group, 5,976 from the United Nations Office of the High Commissioner for Human Rights, and 6,757 from the United States Department of State. Taken together, the reports, news releases, and alerts of these seven organizations offer a substantial cross-national view of human rights issues throughout the world.

Due to the complex URL structure of some of these sites, especially those of Amnesty International, Human Rights Watch, and FIDH, which use both static and dynamic CGI-based URL access pathways to PDFs or have nonstandard mirroring structures, there may be duplication in the collection with the same report indexed under several different URLs. Extensive manual investigation of the URL structure of each site was used in an attempt to minimize this duplication as much as possible. Many of the older reports were scanned and OCR'd from print copies by their respective agencies, leading to a variety of OCR errors that may impact the accuracy of the extracted metadata.

Each document was processed using the GDELТ Global Knowledge Graph system ¹ in the new GKG 2.0 file format. ² The following materials were processed:

- **Amnesty International.** ³ Processed the complete online library from January 1960 to present including all English-language Reports, Press Materials, Urgent Actions, Event, and "Other" documents for all countries and issues. Supplemented with a search of the URL structure of the site as recorded by the Internet Archive.
- **FIDH (International Federation for Human Rights).** ⁴ Conducted a search of the URL structure of the site as recorded by the Internet Archive for all PDF reports and processed the 5,255 resulting matches.
- **Human Rights Watch.** ⁵ Processed the complete news archive 1997 to present (archive starts in 1991, but 1997 is the first year with a substantial number of articles). Processed the entire report archive 1990 to present (some earlier reports are no longer web-accessible and thus

¹ <http://gdeltproject.org/#computing>

² http://data.gdeltproject.org/documentation/GDELТ-Global_Knowledge_Graph_Codebook-V2.pdf

³ <http://www.amnesty.org/>

⁴ <http://www.fidh.org/en/>

⁵ <http://hrw.org/>

were excluded). Supplemented with a search of the URL structure of the site as recorded by the Internet Archive.

- **International Criminal Court.**⁶ Conducted a search of the URL structure of the site as recorded by the Internet Archive for all PDF reports and processed the 708 resulting matches.
- **International Crisis Group.**⁷ Conducted a search of the URL structure of the site as recorded by the Internet Archive for all PDF reports and processed the 1,333 resulting matches.
- **United Nations Office of the High Commissioner for Human Rights.**⁸ Conducted a search of the URL structure of the site as recorded by the Internet Archive for all PDF reports and processed the 5,976 resulting matches.
- **US State Department Country Reports on Terrorism Report.**⁹ Processed the entire web-accessible collection 2013-2000.
- **US State Department Human Rights Reports.**¹⁰ Processed the entire web-accessible collection 2013-1999.
- **US State Department International Religious Freedom Report.**¹¹ Processed the entire web-accessible collection 2013-2001.
- **US State Department Trafficking in Persons Report.**¹² Processed the entire web-accessible collection 2014-2001.

The GKG 2.0 files for this collection are available for download at the following links. NOTE that as with all GKG files, these do NOT contain any of the text of the articles, only the codified metadata computed from them. At this time the collection is available only through download of the GKG 2.0 codified files, but will soon be available through the GDELT Analysis Service as it is upgraded to the GKG 2.0 format. For more information on working with the GKG 2.0 file format, please see the GKG 2.0 codebook.¹³

- http://data.gdeltproject.org/gkgv2_specialcollections/HR-GKG.AMNESTY.gkgv2.csv.zip
- http://data.gdeltproject.org/gkgv2_specialcollections/HR-GKG.CRISISGROUP.gkgv2.csv.zip
- http://data.gdeltproject.org/gkgv2_specialcollections/HR-GKG.FIDH.gkgv2.csv.zip
- http://data.gdeltproject.org/gkgv2_specialcollections/HR-GKG.HRW.gkgv2.csv.zip
- http://data.gdeltproject.org/gkgv2_specialcollections/HR-GKG.ICC.gkgv2.csv.zip
- http://data.gdeltproject.org/gkgv2_specialcollections/HR-GKG.OHCHR.gkgv2.csv.zip
- http://data.gdeltproject.org/gkgv2_specialcollections/HR-GKG.USSTATE.gkgv2.csv.zip

⁶ <http://icc-cpi.int/Pages/default.aspx>

⁷ <http://www.crisisgroup.org/>

⁸ <http://www.ohchr.org/EN/Pages/WelcomePage.aspx>

⁹ <http://www.state.gov/j/ct/rls/crt/>

¹⁰ <http://www.state.gov/j/drl/rls/hrrpt/>

¹¹ <http://www.state.gov/j/drl/rls/irf/>

¹² <http://www.state.gov/j/tip/rls/tiprpt/index.htm>

¹³ http://data.gdeltproject.org/documentation/GDELT-Global_Knowledge_Graph_Codebook-V2.pdf