

# United Nations Peacekeeping Missions: A Data Extract Project

Debbie Rabina<sup>1</sup> and Julian Todd<sup>2</sup>

<sup>1</sup> Pratt Institute, School of Information and Library Science, New York

<sup>2</sup> ScraperWiki, Liverpool

## Abstract

This paper describes a project to uncover data about the United Nations peacekeeping operations (PKO). The goals of the project are twofold: first, to uncover the raw numbers available about UN peacekeeping missions, in order to reveal information about the number of troops allocated by different countries to the various missions, as well as numbers about the funding of these missions; second, to describe the process of scraping data and place the practice in the broader context of social science research in the field of open data from intergovernmental organizations.

**Keywords:** data scraping, coding, open official information, United Nations peacekeeping operation, case-study

**Citation:** Rabina, D., & Todd, J. (2014). United Nations Peacekeeping Missions: A Data Extract Project. In *iConference 2014 Proceedings* (p. 583–591). doi:10.9776/14142.

**Copyright:** Copyright is held by the authors.

**Contact:** drabina@pratt.edu, julian@scraperwiki.com

## 1 Introduction

One of the expressions of globalization is dependence on data from multiple nations and international organizations that can be applied to examine and resolve international concerns and demonstrate the impact of their activities (Development Impact; OpenData; Zollie, 2013). Shared data, specifically shared international data, has been instrumental for global topics such as climate research, health, migration, and economics.

While international organizations have long been committed to activities for the betterment of global issues, it is only recently that they have become aware of the importance of sharing the data they produce with people outside their agencies.

Not all organizations and governments make their data available to the general public, whether due to policy, lack of awareness, or shortage of funds. Older data that predates the recent demand brought on by the popularity of the data age can be particularly difficult to come by. Researchers in need of data that is not available in existing datasets can mine documents, often available in formats such as Microsoft Word or PDF files, to extract the information they need and turn it into a dataset, a practice known as data scraping. Data scraping is an important contribution to the ability to share data.

This paper describes a partnership between a researcher and a programmer working together to uncover data about the United Nations peacekeeping operations (PKO). The goals of the project are twofold: first, to uncover the raw numbers available about UN peacekeeping missions, in order to reveal information about the number of troops allocated by different countries to the various missions, as well as numbers about the funding of these missions; second, to describe the process of scraping data and place the practice in the broader context of social science research in the field of open data from intergovernmental organizations.

## 2 The United Nations peacekeeping missions

Founded in 1945, the United Nations (UN) is an international organization that was established following the disbanding of the League of Nations and the end of World War II. The UN is governed by a charter and it is through this charter that peacekeeping is supported, specifically through Chapter VI, *Pacific*

*Settlement of Disputes*, and Chapter VII, *Action with Respect to Threats to the Peace, Breaches of the Peace, and Acts of Aggression* (United Nations, 1945).

Yet in spite of its favored standing in the role of peacekeeper, UN PKO activities have also been the subject of criticism and public scrutiny. Some critics see PKO, particularly long-term operations such as the missions to the Middle East, India-Pakistan, and Cyprus, as perpetuating existing conditions by lifting responsibility from conflicting countries (The Panel..., 2000).

### 3 UK PKO: Available data

The United Nations does not have an open data policy, and in this is not unique among governments and NGOs. Since data gathering and dissemination is not core to the responsibilities of such organizations, transforming information into data is often the responsibility of the user (Whitmore, 2012). There is a plentitude of information available to the public through their main statistical website (UN data), and in some cases these statistics are available for download as raw data, but availability varies.

The main UN peacekeeping portal (United Nations peacekeeping) provides statistics and basic facts, but does not offer raw data about the number of troops, number of troops per mission, funds allocated by country, funds allocated per mission, or funds allocated per country per mission. Such data is instrumental to researchers trying to answer questions regarding specific missions, or broader questions such as ways to achieve cost reduction or examine the benefits of interventions, relations of troops or expenditures to casualties, participation parity, or cost benefit analysis of PKO.

### 4 Data scraping

Data scraping began gaining popular usage in the 2000s, aided by the increase of data sources available on the web and the momentum built by open-source and open-access public information advocacy. Data sources are shared through data management systems (DMS), also referred to as data hubs, that allow gathering, sharing, and using data. In order for users to work with data for their intended purposes, a DMS should ideally allow the following capabilities: (1) load and update data from any source; (2) store datasets and index them for querying; (3) view, analyze, and update data in a tabular interface such as a spreadsheet; (4) visualize data, for example with charts or maps; (5) analyze data, for example with statistics and machine learning; (6) organize many people to enter or correct data (crowd-sourcing); (7) measure and ensure the quality of data and its provenance; (8) define data permission as open, private, or shared; (9) find datasets and organize them to help others find them; and (10) sell data, sharing processing costs between users (Irving & Pollock, 2012).

We identify these four uses/phases for scraped data as follows: The first use is in *journalistic statements*, the second use is to produce *tables* that cull data from various sources, the third use is for *software*, and the fourth use, finally, is in *analysis*. While these areas are not always distinct, the following examples can help explain each category of use.

*Journalistic statements:* A journalist may hypothesize that some nations use UN peacekeeping missions as training ground to prepare their soldiers for war, but would need to first uncover supporting evidence. Scraping data is often the only way to uncover such data from public sources, as will be demonstrated in the paper.

*Tables:* Scraped data can be used to gather data, store it on a spreadsheet, and create statistics from it. As suggested in an example by Paul Bradshaw (2010):

“you might use a screen scraper to gather information from a local police authority website, and store it in a lovely spreadsheet that you can then sort through, average, total up, filter and so on – when the alternative may have been to print off 80 PDFs and get out the highlighter pens, Post-Its and back-of-a-fag-packet calculations.”

*Software:* Part of the scraping process involves turning data in one format (e.g. text in a PDF file) into another format that is machine-readable, such as Python, and that allows users to run a computer program on it. The ScaperWiki provides a number of tutorials for programming scraped data (ScaperWiki Documentation).

*Analysis:* In a recent news story about immigration policy for the rich, researcher Dune Lawrence wanted to examine the success of an immigration policy that allows immigrants to buy U.S. citizenship. The researcher found that the INS does not make this data available, and had to cull data from many indirect sources to provide an analysis of the phenomena. Lawrence ended up using indirect data from the INS about the EB-5 program (an immigration program that allows wealthy investors to get a green card in exchange for funding American businesses), such as lawsuits filed or number of applications granted. Lawrence found that no evidence is collected that will be able to provide a picture of the true economic value of the EB-5 program (Investing in citizenship, 2013).

## 5 The scraping process

The scraping process typically consists of four steps: Step I, scraping; Step II, analysis; Step III, presentation; and Step IV, publishing and promoting. These help us progress from the initial idea that promoted the scraping, through recognizing its potential, all the way to pulling in and transforming the original raw data and determining a way to analyze it. For programmers it is akin to prospecting for diamonds, finding them in the earth, cutting a mine tunnel with bare hands, separating the diamond from the rock, roughing out its edges, and gluing it onto a steel washer for a ring to show the world.

Once the data sources have been identified, Step I in the process is the *scraping* itself, i.e., producing usable data. As indicated above, numeric information about UN Peacekeeping missions—particularly information regarding the funding of the missions — is not readily available from the UN, and what little is available is in the form of non-searchable PDF files.

A typical PDF can be viewed in Figure 1:

<b>UN Mission's Summary</b>					
<i>detailed by Country</i>					
Month of Report : <b>31-Jan-12</b>					
<i>Country</i>	<i>UN Mission</i>	<i>Description</i>	<i>M</i>	<i>F</i>	<i>Totals</i>
<b>Argentina</b>	MINURSO	Experts on Mission	3	0	3
	MINUSTAH	Individual Police	19	1	20
		Contingent Troop	688	33	721
	UNFICYP	Contingent Troop	251	14	265
			13	0	13
	UNMIL	Individual Police	13	0	13
	UNMISS	Individual Police	1	0	1
	UNOCI	Individual Police	2	1	3
	UNTSO	Experts on Mission	3	0	3
<b>Australia</b>	UNAMA	Experts on Mission	2	0	2

Figure 1: UN Missions summary, detailed by country, Source: ScaperWiki Data Blog

The table in Figure 1 shows that during the month of January 2012, the government of Argentina sent three of its citizens to the United Nations Mission for the Referendum in Western Sahara, 741 people to the United Nations Stabilization Mission in Haiti, 265 to the United Nations Peacekeeping Force in Cyprus, and so forth. For coding purposes, this format is relatively easy to work with, since it is consistent and the coder was able to complete it with about 160 lines of code. After getting the code working, we cleaned it up by removing all the leftover print statements until all that would be produced at runtime was a message appearing when a new month became available in the database. The email-generating code is on line 34, and it regularly sends an email as captured in Figure 2:

***Subject: UN peacekeeping statistics for 2012-01***

*Dear friend,*  
*There are 788 new records in the database for*

*[https://scraperwiki.com/scrapers/un\\_peacekeeping\\_statistics/](https://scraperwiki.com/scrapers/un_peacekeeping_statistics/)*

*after month 2011-12 to month 2012-01*

Figure 2: Sample e-mail alert for new document, Source: ScraperWiki Data Blog

The preceding step resulted in a table of over 86,000 records stretching back to January 2003, and now gives way to the second step: *analysis*.

The important columns in the table are shown in Figure 3:

```
month text,  
country text,  
mission text,  
people integer
```

Figure 3: Source: ScraperWiki Data Blog

This data, with the addition of SQL, allows the creation of hundreds of relevant timeline graphs. For example, one can answer questions such as: *What are the three top countries in terms of maximum deployment to any mission*, or: *To which missions do these three neighboring, sometimes-at-war, rival countries predominantly send their troops?* (detailed step-by-step examples for analyzing the data is available from Todd, 2012).

The data can be parsed down further to allow us to look at just the deployment of peacekeepers from India, Bangladesh, and Pakistan to MONUC (United Nations Organization Mission in the Democratic Republic of the Congo) over time, illustrated by Figure 4. As you can see, the pattern of deployment tends

to remain at a constant quota over many years, with sudden jumps, probably due to requirements on the ground. Pakistan appeared to supply both of these peacekeeping surges, once in 2003 and once in 2005, while Bangladesh surged at one and India surged at the other.



Figure 4: Source: ScraperWiki Data Blog

Following Step I, *scraping*, and Step II, *analysis*, we enter Step III, *presentation*.

Data scrapers are typically trained as programmers and feel inept when it comes to presentation. It is not uncommon that projects get abandoned at this stage, since programmers are intimidated by the challenges posed by design. While this is time-consuming, the job is not done until the datasets can be presented in a format comprehensible to non-programmers.

In the case of the UN peacekeeping mission, we used an interface for generating graphs of the queries that people might be interested in. It took two hard hacking sessions to get it into this form — twice as long as it took to write the original scraper — and the result can be seen in Figure 5:

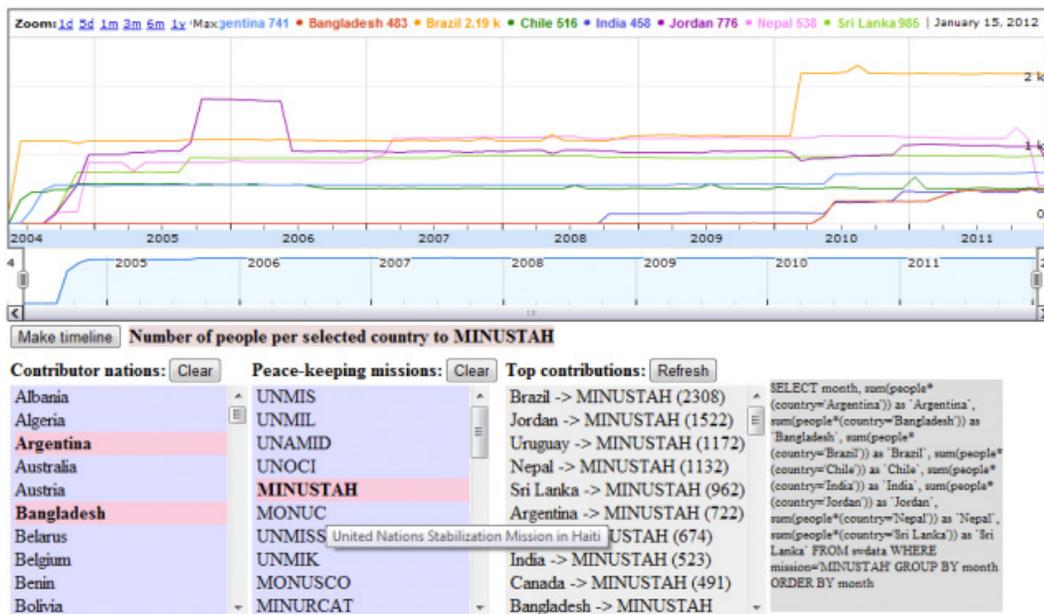


Figure 5: Source: ScraperWiki Data Blog

When the page initializes, there are three Ajax callbacks to the database to obtain the lists of countries, missions, and top contributions from countries to specific missions. Users can make multiple selections from the countries and the missions lists to create timeline graphs of numbers of people involved. If one selects only from the countries list, it shows the troop contributions from those countries to all UN missions. Users

can additionally select a single mission as well, and it will graph those country contributions to that specific mission. And it works in the other direction, for lists of missions by countries. The top contributors table helps identify the top countries (or missions), so you know which ones to select to make an interesting graph with content (e.g. there is no point in graphing the number of Italians deployed to Nepal, because there aren't any).

Where do the Italians go? You can find that out by selecting "Italy" from the "Contributor nations" column and clicking the "Refresh" button on the "Top contributions" column. And you can also click on "Make timeline" to discover that Italy never sent anyone anywhere until late 2006, when they suddenly started deploying two to three thousand peacekeepers to Lebanon. What happened then? Did something change in Italian politics around that point? That would be the subject of another project.

The fourth and final step in the process is *publishing and promoting* the data, by far the least straightforward part of the process. Typically the way in which the data is made public reflects the goals for which it was collected (see Section II). Journalists, academics, market researchers, lobbyists, and other potential interested parties will use the data to support their goals.

## 6 The case of the Democratic Republic of Congo

To further illustrate the process and utility of the scraping process, we focused on the UN peacekeeping mission in the Democratic Republic of Congo (DRC).

The UN established a peacekeeping mission in the Congo in 1999. The mission operated until 2010 under the name MONUC (Mission de l'Organisation des Nations Unies en République démocratique du Congo) and in 1999 the mission's name was changed to MONUSCO. The mission to Congo was authorized to operate until June 30, 2013. There are currently 20,000 uniformed personnel serving in the Congo with an approved budget of US\$1.4 billion for the 2012-2013 fiscal year (MONUSCO facts and figures). The number of troops in Congo increased by 440% from the original 3,700 of troops deployed when the mission was first established (Bellamy & Williams, 2011, 122).

To dig deeper down into a specific PKO, such as the MONUC, we will look more closely at deployment of peacekeepers from India, Bangladesh, and Pakistan to MONUC over time. Interest in these particular country deployments to Congo were done to test a journalistic hypothesis that these peacekeeping missions are a clever way for nations to get their troops battle-hardened before the inevitable conflict on their own territory. In other words, they also serve as war-training missions.

We created a SQL query (Figure 6):

**General graph of things**

Useful for quickly generating graphs from sql queries. *To do: timeline graph, layout (eg float right), multiple fields in output, create direct csv/html download links from api*

Scraper name:

API key:

Attach:

SQL:

```
SELECT month || '-15',
sum(people*(country='India')) as people_india,
sum(people*(country='Bangladesh')) as
people_bangladesh,
sum(people*(country='Pakistan')) as people_pakistan
FROM swdata
WHERE mission='MONUC'
GROUP BY month
ORDER BY month
```

[download apidocs](#)

Figure 6: Generating graphs from SQL queries, Source: Source: ScraperWiki Data Blog

This query creates a dynamic JavaScript timeline, presented as a bitmap image (Figure 7). As the timeline shows, the pattern of deployment tends to remain at a constant quota over many years, with sudden jumps, probably due to requirements on the ground. Pakistan appeared to supply both of these peacekeeping surges, once in 2003 and once in 2005, while Bangladesh surged at one and India surged at the other. Similarly, this process can be repeated for Bangladesh and Pakistan.



Figure 7: Deployment of peacekeepers from India, Bangladesh and Pakistan to MONUC over time, Source: ScrapperWiki Data Blog

Once the data is available, it is rendered usable in the presentation stage. This is often the most challenging and time-consuming part. In this particular case, creating the visual presentation (Figure 8) took twice as long as scraping the data. The result is an interactive table that allows users to select PKO and countries to check changes in deployment over time.

### UN peace-keeping stats

Sourced from [UN site](#) by application of [un\\_peacekeeping\\_statistics](#) scraper.

You can graph the peacekeeping contributions from different countries and to different missions, or graph the contribution from several countries. For more information, [see the blog](#).

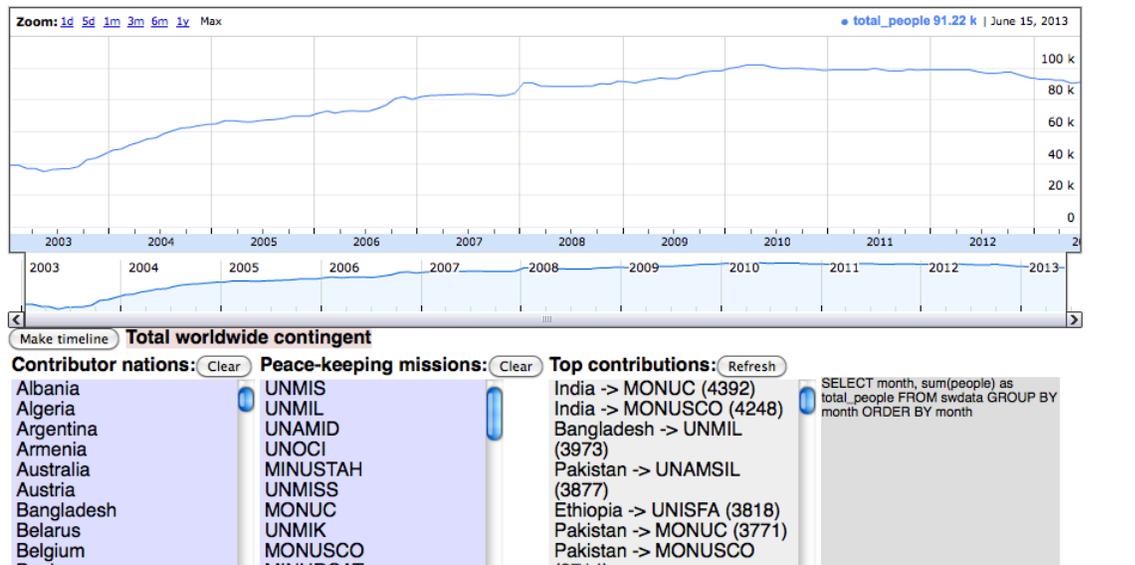


Figure 8: UNPKO statistics, Source: UN Peacekeeping Statistics

## 7 Conclusion

In this paper we described the uses and process of extracting data from United Nations documents. As awareness and interest in data increases we can expect to see more datasets available online in downloadable formats, although we can anticipate a backlog in making older data available. Data scraping allows users to convert statistics available in PDF or Word documents into useable data. This data can be used to answer questions that are quantitative in nature, even when the data creator (in our case the United Nations) does not made its data publicly available.

## 8 References

- Bellamy, Alex J., & Williams, Paul D. (2010). *Understanding peacekeeping*. Cambridge, UK: Polity.
- Bradshaw, Paul (2010) An introduction to data scraping with Scraperwiki. *Online Journalism Blog*. Retrieved from <http://onlinejournalismblog.com/2010/07/07/an-introduction-to-data-scraping-with-scraperwiki/>
- Cavallo, Alberto (2012). Scraped Data and Sticky Prices - MIT Sloan Working Paper. Available at <http://www.mit.edu/~afc/#papers>
- A comparative evaluation of united nations peacekeeping: Subcommittee on International Organizations Human Rights and Oversight, House Committee on Foreign Affairs, 110th Congress (2007). Development Impact. A World Bank blog. Retrieved from <http://blogs.worldbank.org/impacetevaluations/blog>
- Investing in citizenship: For the rich, a new road to the U.S. (2013). *All Things Considered*. Retrieved from <http://www.wamc.org/post/investing-citizenship-rich-new-road-us>
- Irving, F., & Pollock, R. (2012). From CMS to DMS: C is for content, D is for data. Retrieved from <http://blog.scraperwiki.com/2012/03/09/from-cms-to-dms-c-is-for-content-d-is-for-data/>
- MONUSCO facts and figures. Retrieved from <http://www.un.org/en/peacekeeping/missions/monusco/facts.shtml>
- Noble, S., Russell, C., & Wiseman, R. (2011). Mind the gap: Global data sharing. *IASSIST Quarterly*, Fall, 6.
- OpenData. The World Bank data blog. Retrieved from <http://blogs.worldbank.org/opendata/>
- The Panel on United Nations Peace Operations* (the Brahimi report). (2000). Retrieved from [http://www.un.org/peace/reports/peace\\_operations/](http://www.un.org/peace/reports/peace_operations/)
- Peacekeeping fact sheet*. (2012). Retrieved from <http://www.un.org/en/peacekeeping/resources/statistics/factsheet.shtml>
- ScraperWiki*. Retrieved from <https://scraperwiki.com/>
- ScraperWiki Data Blog. Retrieved from <http://tinyurl.com/l3dpyre>
- ScraperWiki Documentation*. Retrieved from <https://scraperwiki.com/docs/python/>
- Todd, J. (2012). The UN peacekeeping mission contributions mostly baked. Retrieved from <http://blog.scraperwiki.com/2012/02/22/the-un-peacekeeping-mission-contributions-mostly-baked/>
- UN at a glance*. Retrieved from <http://www.un.org/en/aboutun/index.shtml>
- UN data*. Retrieved from <http://data.un.org/>
- UN Peacekeeping statistics. Retrieved from [https://views.scraperwiki.com/run/un\\_peacekeeping\\_statistics\\_view/](https://views.scraperwiki.com/run/un_peacekeeping_statistics_view/)
- United Nations. (1945). *Charter of the United Nations*. Retrieved from <http://www.un.org/en/documents/charter/index.shtml>
- United Nations peacekeeping*. Retrieved from <http://www.un.org/en/peacekeeping/>
- United nations peacekeeping forces. (2012). In *Encyclopædia britannica*. doi:<http://www.britannica.com/EBchecked/topic/616454/United-Nations-Peacekeeping-Forces>;

- Uppsala University, Dept. of Peace and Conflict Research. (2012). *Frequently asked questions*. Retrieved June 7, 2012, from [http://www.pcr.uu.se/research/ucdp/faq/#How\\_are\\_UCDP\\_data\\_collected\\_](http://www.pcr.uu.se/research/ucdp/faq/#How_are_UCDP_data_collected_)
- Whitmore, A. (2012). Extracting knowledge from U.S. department of defense freedom of information act requests with social media. *Government Information Quarterly*, 29, 151-157.  
doi:10.1016/j.giq.2011.08.015
- Wolfram alpha*. Retrieved from <http://www.wolframalpha.com/>
- Zollie, A. (2013). The mismeasure of impact. Retrieved from [http://poptech.org/e3\\_andrew\\_zolli\\_1](http://poptech.org/e3_andrew_zolli_1)

## 9 Table of Figures

Figure 1: UN Missions summary, detailed by country, Source: ScraperWiki Data Blog.....	585
Figure 2: Sample e-mail alert for new document, Source: ScraperWiki Data Blog.....	586
Figure 3: Source: ScraperWiki Data Blog.....	586
Figure 4: Source: ScraperWiki Data Blog.....	587
Figure 5: Source: ScraperWiki Data Blog.....	587
Figure 6: Generating graphs from SQL queries, Source: Source: ScraperWiki Data Blog.....	588
Figure 7: Deployment of peacekeepers from India, Bangladesh and Pakistan to MONUC over time, Source: ScraperWiki Data Blog.....	589
Figure 8: UNPKO statistics, Source: UN Peacekeeping Statistics.....	589