

Data Profile - World Health Organization API

Kurt Wydrinski

11/22/2019

Source

During the formation of the United Nations in 1945, one of the key objectives for its ongoing mission was to foster cooperation between countries to address many social well-being issues, including health.¹ This objective was taken on as part of the World Health Organization (WHO) which was established on April 7, 1948.² The WHO continues this mission today with 194 countries cooperating at various levels to promote global health concerns. In doing so, the WHO regularly collects data from its members that are related to health issues. The data contains various time series regarding diseases, illnesses, economics, social demographics, etc. This data is collectively captured in the WHO's online database *Global Health Observatory (GHO)*.³

API Overview

The R package `WHO`, provides a simple API to access the GHO. It only provides two functions: `get_codes()` and `get_data()`. Inside the GHO, each time series that exists is identified by a label. Each label is a code that uniquely identifies the series. These labels are then used as a parameter to `get_data()` to retrieve the time series observations.

```
# install.packages("WHO")
library(WHO)
```

The code below uses the `extra` parameter to download all metadata available for the GHO codes.

```
who_codes <- get_codes(extra = TRUE)
glimpse(who_codes)
```

```
## Observations: 3,287
## Variables: 9
## $ label      <chr> "MDG_0000000001", "MDG_0000000003", "MDG_0000000005"...
## $ display    <chr> "Infant mortality rate (probability of dying between..."
## $ url        <chr> "https://www.who.int/data/gho/indicator-metadata-reg..."
## $ display_fr <chr> "Taux de mortalité des nourrissons (probabilité de d..."
## $ display_es <chr> "Tasa de mortalidad de menores de 1 año (probabilida..."
## $ definition_xml <chr> "http://apps.who.int/gho/indicatorregistryservice/pu..."
## $ category   <chr> "Mortality and global health estimates", "Sustainabl..."
## $ imr_id     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "16", NA...
## $ renderer_id <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
```

There are 3,287 indicators or time series in this dataset. The `category` variable is a WHO grouping of the indicators into 51 sets of indicators as follows:

```
who_codes %>%
  group_by(category) %>%
  summarise(indicator_count = n()) %>%
  knitr::kable()
```

¹ See <https://www.history.com/this-day-in-history/the-united-nations-is-born>.

² See <https://www.who.int/about/who-we-are/history>.

³ See <https://www.who.int/gho/about/en/>.

category	indicator_count
AMR GLASS Coordination	4
AMR GLASS Quality assurance	5
AMR GLASS Surveillance	3
Child health	37
Demographic and socioeconomic statistics	19
Essential health technologies	10
FINANCIAL PROTECTION	28
Global Observatory for eHealth (GOe)	92
Health Equity Monitor	64
Health financing	13
Health systems	381
Health workforce	24
HIV/AIDS and other STIs	37
Infectious diseases	31
Infrastructure	6
Injuries and violence	59
Insecticide resistance	9
International Health Regulations (2005) monitoring framework	13
Malaria	126
Medical equipment	8
Millennium Development Goals (MDGs)	15
Mortality and global health estimates	58
Neglected tropical diseases	9
Neglected tropical diseases	17
Neglected Tropical Diseases	17
Noncommunicable diseases	50
Noncommunicable diseases and mental health	2
Noncommunicable diseases CCS	82
Nutrition	27
Oral health	1
Public health and environment	177
RSUD: GOVERNANCE, POLICY AND FINANCING : PREVENTION	3
RSUD: GOVERNANCE, POLICY AND FINANCING: FINANCING	7
RSUD: GOVERNANCE, POLICY AND FINANCING: TREATMENT	12
RSUD: HUMAN RESOURCES	8
RSUD: INFORMATION SYSTEMS	6
RSUD: SERVICE ORGANIZATION AND DELIVERY: PHARMACOLOGICAL TREATMENT	11
RSUD: SERVICE ORGANIZATION AND DELIVERY: PREVENTION PROGRAMS AND PROVIDERS	5
RSUD: SERVICE ORGANIZATION AND DELIVERY: SCREENING AND BRIEF INTERVENTIONS	4
RSUD: SERVICE ORGANIZATION AND DELIVERY: SPECIAL PROGRAMMES AND SERVICES	12
RSUD: SERVICE ORGANIZATION AND DELIVERY: TREATMENT CAPACITY AND TREATMENT COVERAGE	2
RSUD: SERVICE ORGANIZATION AND DELIVERY: TREATMENT SECTORS AND PROVIDERS	8
RSUD: YOUTH	3
Substance use and mental health	1121
Sustainable development goals	28
Tobacco	3
Tuberculosis	68
Universal Health Coverage	16
Urban health	32
World Health Statistics	67
NA	447

Note the following about the above categories.

1. There are 447 indicators that do not have a category assigned by the WHO. These should be considered as a category called **Uncategorized** or **No category**.
2. The naming convention is not consistent across categories. For example, some use single words, phrases, all capital letters, irregular case, etc. Category names should be somewhat standardized while retaining the original meaning in the dataset.
3. A number of categories appear to be fragmented. For example, there are three permutations of *Neglected Tropical Diseases* that need to be combined into a single category.

The table below provides a list of the indicators in the *Demographic and socioeconomic statistics* category. However, inspection of the table should make it apparent that a number of these indicators are based on population.

```
who_codes %>%
  filter(category == "Demographic and socioeconomic statistics") %>%
  select(label, display) %>%
  knitr::kable()
```

label	display
WHS9_CBR	Crude birth rate (per 1000 population)
WHS9_CDR	Crude death rate (per 1000 population)
WHS9_CS	Cellular subscribers (per 100 population)
WHS10_1	Most recent census (year)
WHS10_2	Number of cause-of-death registration years available
WHS10_3	Number of national population surveys - child anthropometry
WHS10_4	Number of national population surveys - child mortality

label	display
WHS10_5	Number of national population surveys - maternal mortality
WHS10_6	Number of national population surveys - HIV prevalence
WHS10_7	Number of national population surveys - adult health
WHS10_8	Civil registration coverage of cause-of-death (%)
WHS10_9	Ill-defined causes in cause-of-death registration (%)
CCO_1	Poverty headcount ratio at \$1.25 a day (PPP) (% of population)
CCO_2	Human development index rank
CCO_3	Gender inequality index rank
ITU_IDI	ICT Development Index (IDI)
ITU_IDI_RANK	ICT Development Index (IDI) rank
ITU_ICT_1	Percentage of individuals using the Internet
ITU_ICT_2	Mobile-cellular telephone subscriptions per 100 inhabitants

A closer look may make it apparent that the data may not have a total population number. It is possible that WHS10_1 could have it by the display string.

```
who_codes %>%
  filter(label == "WHS10_1")
```

```
## # A tibble: 1 x 9
##   label display url      display_fr display_es definition_xml category imr_id
##   <chr> <chr>   <chr> <chr>      <chr>      <chr>      <chr>   <chr>
## 1 WHS1~ Most r~ http~ <NA>      <NA>      http://apps.w~ Demogra~ <NA>
## # ... with 1 more variable: renderer_id <chr>
```

Inspection of the indicator observation in the metadata shows that the `url` and `definition_xml` variables are URLs to further information about the indicator. `url` is an address for a web page on the GHO registry that explains what the indicator represents. The `definition_xml` provides the same information presented in the explanation page in XML format.

As the explanation page suggests, the WHS10_1 indicator does reflect the population count/census statistic. This may or may not be useful when used with other indicators. In looking at all indicators in the metadata, there are at least three others that are called Population.

```
who_codes %>%
  filter(str_detect(display, "^Population$")) %>%
  select(label, display, category)
```

```
## # A tibble: 3 x 3
##   label      display      category
##   <chr>      <chr>      <chr>
## 1 MALARIA_15279 Population Malaria
## 2 RS_1845      Population Injuries and violence
## 3 MEDS1_01_01  Population Health systems
```

API Risks

Based on the analysis above, the following risks appear to exist when using this API dataset.

1. The data is not tidy and needs to be made such before detailed analysis can be completed.
2. The structure of the data is inconsistent and needs detailed exploration when deciding on which indicators to use.

3. The `category` variable in the metadata (i.e. codes) is not very reliable. Careful considerations need to be made when leveraging this variable. The category should be used as the basis for a new variable which is a reliable identifier for logically-related indicators.

API Rewards

Based on the analysis above the following benefits seem to be gained by using this API and dataset.

1. The data can be accessed using the simple `WHO` package.
2. The data is accessible via the Internet via on-demand API calls.
3. The data contains international health and disease data along with related indicators.

Usage Prototype

The sections below provide a prototypical usage for this dataset and API. The actual use and visualization of the data will vary between projects. However, the sections below demonstrate common usage patterns for this data and API.

Prototype Assumptions / Requirements