Open standardized data, faster discovery

Unusable data

Easy-to-use data

Research
- Digitization
- Extraction
- Standardization
- Integration
- Annotation

US data:
1888-present
WHO dengue data: 1952-2010

www.tycho.pitt.edu

Wall street journal
Forbes magazine
Level 1 data
- Data from NEJM paper
- Imputed
- Population and rates
- 8 conditions
- 1916-2009
- GUI

Level 2 data
- All “consistent” data
- 50 conditions
- 1888-2014
- GUI or API

Level 3 data
- Everything
- Duplicates, unknown condition, etc.
- 58 conditions, many subcategories
- 1887-2014
- Upon request

In development
- Dengue from WHO for 99 countries
- Dengue from 8 SE Asia countries
- Chikungunya data from PAHO
- GUI
Reuse of Project Tycho v1 data

Cumulative number of registered users

3500+ users
7000 downloads
1 million API calls
272,000 page visits

50 Creative works by others:
- 18 peer-reviewed papers
- 9 other papers
- 4 theses
- 19 visualizations/software applications
  - Infographics in newspapers
  - R packages
  - Code on Github
  - Visualization applications
Scientific papers using v1 data

2014-3 | History of US immunization program. Nature

2015-5 | Measles immunomodulation. Science

2015-5 | Typhoid fever and health in US cities. Am Econ Rev

2015-6 | Resurgence of pertussis. BMC Med

2015-6 | Pertussis transmission. AJE

2015-6 | Polio transmission. PLOS Biol

2016-7 | Chickenpox and strep A transmission. SAGE Open Med


2017-1 | Temperature and salmonella transmission. Ann Am Assoc Geogr

2017-2 | Media reporting and disease outbreaks. Economica

2017-3 | State spending and disease incidence. Health Serv Res

2017-4 | Two-layered network model of measles. Sci Rep

2017-5 | Online ensemble model for parameter estimation. Stat Med

2016-10 | Measles transmission dynamics. PLOS One

2016-6 | Digital disease detection. PNAS

2016-2 | Measles transmission dynamics. PLOS Comp Bio

2015-12 | Measles and pertussis transmission. SAGE Open Med

2015-10 | Influenza and pneumonia transmission. Sci Rep

2013-11 | Impact of vaccination. NEJM
New in Project Tycho 2.0

NEW DATA ADDED
- US data updated until 31 December 2017
- Integration of dengue surveillance data with US data (no longer “in development”)
  - Annual/monthly aggregate counts from WHO online databases and from 8 partner countries in SE Asia

FINDABLE, INTEROPERABLE, REUSABLE (FAIR)
- Data grouped in 360 datasets, one for each condition-country combination
  - Digital object identifiers and metadata files
- Extensive standardization
  - Standards for condition, pathogen, location
  - Standard data format for all Project Tycho data
Findable: Digital object identifier, data indexed in online catalogues
Accessible: Process for data retrieval clear and no proprietary software
Interoperable: Standard data format, standard ontologies/vocabularies used to describe the data and metadata
Reusable: Rich metadata to help determine reuse options

FAIR -> Machine-interpretable -> Faster discovery
New website: www.tycho.pitt.edu
Continued access to previous versions

Not yet included in version 2.0!
More and better in version 2.0
Three ways to access the data

**DATASETS**

**Pre-Compiled**
All data for one condition in one country compiled in one dataset
Select from 360 datasets: 92 conditions for the US and dengue-related conditions for countries around the world
Each dataset has a digital object identifier and standard citation

**DATA QUERY**

**Compile Your Own**
Search data for any condition and location
Visualize and compare data for multiple conditions or locations
Download selected data in Project Tycho standard format

**DATA API**

**Application Programming Interface**
Enables data access for computer programs
Search data for any possible combination of countries, conditions, and other variables
Download selected data in Project Tycho standard format

START NOW!
## Pre-compiled US Datasets: 92

### Pre-Compiled US Datasets

**Datasets for 92 Conditions in the United States**

<table>
<thead>
<tr>
<th>Country name</th>
<th>Condition</th>
<th>DOI</th>
<th># Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States of America</td>
<td>Amebic dysentery</td>
<td>10.25337/T7/ptycho.v2.0/US.387754006</td>
<td>11708</td>
</tr>
<tr>
<td>United States of America</td>
<td>Anthrax</td>
<td>10.25337/T7/ptycho.v2.0/US.409498004</td>
<td>7100</td>
</tr>
<tr>
<td>United States of America</td>
<td>Babesiosis</td>
<td>10.25337/T7/ptycho.v2.0/US.21061004</td>
<td>3300</td>
</tr>
<tr>
<td>United States of America</td>
<td>Bacillary dysentery</td>
<td>10.25337/T7/ptycho.v2.0/US.274081004</td>
<td>11094</td>
</tr>
<tr>
<td>United States of America</td>
<td>Brucellosis</td>
<td>10.25337/T7/ptycho.v2.0/US.75702008</td>
<td>20002</td>
</tr>
<tr>
<td>United States of America</td>
<td>Chlamydia trachomatis infection</td>
<td>10.25337/T7/ptycho.v2.0/US.240569008</td>
<td>7049</td>
</tr>
<tr>
<td>United States of America</td>
<td>Chlamydial infection</td>
<td>10.25337/T7/ptycho.v2.0/US.105629000</td>
<td>74097</td>
</tr>
<tr>
<td>United States of America</td>
<td>Cholera</td>
<td>10.25337/T7/ptycho.v2.0/US.63650001</td>
<td>101</td>
</tr>
<tr>
<td>United States of America</td>
<td>Coccidioidomycosis</td>
<td>10.25337/T7/ptycho.v2.0/US.60826002</td>
<td>10678</td>
</tr>
<tr>
<td>United States of America</td>
<td>Cryptosporidiosis</td>
<td>10.25337/T7/ptycho.v2.0/US.240370009</td>
<td>54082</td>
</tr>
</tbody>
</table>

*Showing 1 to 10 of 92 entries*
Pre-compiled dengue datasets: 271

<table>
<thead>
<tr>
<th>Country name</th>
<th>Condition</th>
<th>DOI</th>
<th># Records</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Samoa</td>
<td>Dengue</td>
<td>10.25337/T7/ptchc.v2.0/AS.38362002</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>American Samoa</td>
<td>Dengue hemorrhagic fever</td>
<td>10.25337/T7/ptchc.v2.0/AS.20927009</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>American Samoa</td>
<td>Dengue without warning signs</td>
<td>10.25337/T7/ptchc.v2.0/AS.722862003</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Anguilla</td>
<td>Dengue</td>
<td>10.25337/T7/ptchc.v2.0/AS.38362002</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Anguilla</td>
<td>Dengue hemorrhagic fever</td>
<td>10.25337/T7/ptchc.v2.0/AS.20927009</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Anguilla</td>
<td>Dengue without warning signs</td>
<td>10.25337/T7/ptchc.v2.0/AS.722862003</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>Dengue</td>
<td>10.25337/T7/ptchc.v2.0/AG.38362002</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>Dengue hemorrhagic fever</td>
<td>10.25337/T7/ptchc.v2.0/AG.20927009</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>Dengue without warning signs</td>
<td>10.25337/T7/ptchc.v2.0/AG.722862003</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>Dengue</td>
<td>10.25337/T7/ptchc.v2.0/AR.38362002</td>
<td>129</td>
<td></td>
</tr>
</tbody>
</table>

Showing 1 to 10 of 271 entries
Dataset landing page

**DOI**
10.25337/T7/ptycho.v2.0/KH.38362002

**DOWNLOAD LINKS**
Dataset contains: 8889 records
Download data and readme in Project Tycho format: KH.38362002.zip
Download metadata in DATS format: KH.38362002.json
Download metadata in DataCite XML format: KH.38362002.xml

**CONDITION AND PATHOGEN INFORMATION**
SNOMED-CT name: Dengue
SNOMED-CT code: 38362002
Pathogen name: Dengue virus
Pathogen code: 12637

**SPACE AND TIME INFORMATION**
Locations included: CAMBODIA
Apollo Location Service: CAMBODIA
Earliest year: 1980
Latest year: 2011

**CITATION**
Van Panhuis W., Cross A., Burke D., Choisy M., Counts of Dengue reported in CAMBODIA: 1980-2011 (version 2.0, April 1, 2018): Project Tycho data release, DOI: 10.25337/T7/ptycho.v2.0/KH.38362002
1. Explore data content
2. Download a custom datasets
3. Visualize data time series
GUI – Explore data content

CONDITION:
- Acquired immune deficiency syndrome
- Active tuberculosis: 11480
- Acute hepatitis C: 671
- Acute nonparalytic poliomyelitis
- Acute paralytic poliomyelitis: 194
- Acute poliomyelitis: 137581
- Acute type A viral hepatitis: 4105
- Acute type B viral hepatitis: 4127
- Amebic dysentery: 11708
- Anthrax: 7100
- Aseptic meningitis: 41828
- Babesiosis: 3300
- Bacillary dysentery: 11094
- Brucellosis: 20002
- Campylobacteriosis: 4486
- Chlamydia trachomatis infection
- Chlamydial infection: 74097

COUNTRIES:
- BRAZIL: 2068
- BRUNEI DARUSSALAM: 31
- CAMBODIA: 11932
- CANADA: 6
- CAYMAN ISLANDS: 75
- CHILE: 40
- CHINA: 7060
- COLOMBIA: 159
- COOK ISLANDS: 74
- COSTA RICA: 97
- CUBA: 831
- CURAÇAO: 25
- DOMINICA: 116
- DOMINICAN REPUBLIC: 149
- ECUADOR: 101
- EL SALVADOR: 5184
- FIJI: 102
### GUI – Explore data content

#### CONDITION:
- Dengue: 8889
- Dengue hemorrhagic fever: 1543
- Dengue without warning signs: 1500

#### COUNTRIES:
- BOLIVIA (PLURINATIONAL STATE OF): 120
- BRAZIL: 2068
- BRUNEI DARUSSALAM: 31
- CAMBODIA: 94 [11838 in admin level below]
- CANADA: 6
- CAYMAN ISLANDS: 75
- CHILE: 40
- CHINA: 7060
- COLOMBIA: 159
- COOK ISLANDS: 74
- COSTA RICA: 97
- CUBA: 831
- CURAÇAO: 25
- DOMINICA: 116
- DOMINICAN REPUBLIC: 149
- ECUADOR: 101
- EL SALVADOR: 5184
- FIJI: 102

#### SOURCE NAME:
- Cambodia Dengue Surveillance System: 3677
- World Health Organization DengueNet Database: 3677
- World Health Organization Western Pacific Region

#### ADMIN LEVEL 1:
- BÂNTÉAY MÉANCHEY: 507
- BÄTDÂMBÂNG: 515
- KAÕH KÔNG: 506
- KRÂCHÉH: 487
- KRÕNG KEP: 485
GUI – Explore data content

CONDITION:
- Dengue: 8889
- Dengue hemorrhagic fever: 1543
- Dengue without warning signs: 1500

COUNTRIES:
- BOLIVIA (PLURINATIONAL STATE OF): 120
- BRAZIL: 2068
- BRUNEI DARUSSALAM: 31
- CAMBODIA: 94 [11838 in admin level below]
- CANADA: 6
- CAYMAN ISLANDS: 75
- CHILE: 40
- CHINA: 7060
- COLOMBIA: 159
- COOK ISLANDS: 74
- COSTA RICA: 97
- CUBA: 831
- CURAÇAO: 25
- DOMINICA: 116
- DOMINICAN REPUBLIC: 149
- ECUADOR: 101
- EL SALVADOR: 5184
- FIJI: 102

SOURCE NAME:
- Cambodia Dengue Surveillance System: 3677
- World Health Organization DengueNet Database:
- World Health Organization Western Pacific Region:
- World Health Organization Western Pacific Region:

ADMIN LEVEL 1:
- BÂNTÉAY MÉANCHEY: 507
- BÂTDÂMBÂNG: 515
- KAÔH KONGLONG: 506
- KRÂCHÉH: 487
- KRÂNG KÈP: 485
### GUI – Explore data content

#### COUNTRIES:
- CAMBODIA: 0 [507 in admin level below]

#### PATHOGEN:
- Dengue virus: 507

#### SUBPOPULATION:
- None specified: 507

#### PLACE OF ACQUISITION:
- NA: 507

#### DIAGNOSTIC CERTAINTY:
- NA: 507

#### FATALITIES:
- No: 399
- Yes: 108

#### AGE RANGE:
- 0-130: 507

#### START YEAR:
- 1998
- 1999
- 2000
- 2001
- 2002
- 2003

#### END YEAR:
- 1998
- 1999
- 2000
- 2001
- 2002

#### ADMIN LEVEL 1:
- BÁNTÉÁY MÉANCHEY: 507
- BATDÁMBAANG: 515
- KAÔH KÔNG: 506
- KRÂCHEH: 487
- KHÂNG KÔ: 465

#### ADMIN LEVEL 2:
- No options available based on your set filters

#### PLOT PREVIEW:
To preview a plot please select:
- At least 1 condition
- At least 1 source name
- The type of interval
- The type of cumulative count
GUI – Download custom data

CONDITION:
- Meningococcal meningitis: 126493
- Mumps: 165242
- Murine typhus: 2786
- Ornithosis: 824
- Pellagra: 29738
- Pertussis: 297794
- Pneumonia: 198788
- Post-infectious encephalitis: 12918
- Primary encephalitis: 26152
- Rocky Mountain spotted fever: 71024
- Rubella: 72676
- Salmonella infection: 71824
- Scarlet fever: 351882
- Shigellosis: 63790
- Smallpox: 260066
- Smallpox without rash: 19
- Spotted fever group rickettsial disease:

COUNTRIES:
- UNITED STATES OF AMERICA: 0 [71024

SOURCE NAME:
- US Nationally Notifiable Disease Surveillance

ADMIN LEVEL 1:
- No options available based on your set filter

INTERVAL TYPE:
- Week: 71021

ADMIN LEVEL 2:
- No options available based on your set filter
GUI – Download custom data

CONDITION:
- Pneumonia: 5423
- Post-infectious encephalitis: 509
- Primary encephalitis: 1062
- Rocky Mountain spotted fever: 4615
- Rubella: 2129
- Salmonella infection: 3224
- Scarlet fever: 10613
- Shigellosis: 3050
- Smallpox: 10088
- Spotted fever group rickettsial disease: 369
- Streptococcal sore throat: 157
- Syphilis: 2
- Tetanus: 1387
- Toxic shock syndrome: 346
- Tuberculosis: 10532
- Tularemia: 2919

COUNTRIES:
- UNITED STATES OF AMERICA: 0 [4615 in admin level below]

SOURCE NAME:
- US Nationally Notifiable Disease Surveillance System: 4615

INTERVAL TYPE:
- Week: 4615

CUMULATIVE:
- No: 1193
- Yes: 3422

ADMIN LEVEL 1:
- ALABAMA: 2384 [4 in admin level below]
- ALASKA: 29
- ARIZONA: 873
- ARKANSAS: 2227
- CALIFORNIA: 4615

ADMIN LEVEL 2:
- No options available based on your set filters

CITY:
- No options available based on your set filters

RESET ALL FILTERS  DOWNLOAD  DETAILED PLOT
GUI – Visualize data

PLOT PREVIEW:
To preview a plot please select:

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least 1 condition</td>
<td>At least 1 condition</td>
</tr>
<tr>
<td>At least 1 source name</td>
<td>At least 1 source</td>
</tr>
<tr>
<td>The type of interval</td>
<td>Type of interval: weekly, monthly, yearly</td>
</tr>
<tr>
<td>The type of cumulative count</td>
<td>Cumulative or not</td>
</tr>
<tr>
<td>The records selected is greater than 0</td>
<td>Number of records &gt; 0</td>
</tr>
</tbody>
</table>

Need to select sufficient information to get a time series of equivalent counts, e.g., not mixing up different age groups, cumulative and non-cumulative counts, etc.
GUI – Visualize data

**CONDITION:**
- Dengue: 15692
- Dengue without warning signs: 39

**COUNTRIES:**
- SINGAPORE: 314
- SOLOMON ISLANDS: 21
- SRI LANKA: 2011
- SURINAME: 61
- TAIWAN, PROVINCE OF CHINA: 140 [15552 in admin]
- TIMOR-LESTE: 105
- TOKELAU: 18
- TONGA: 37
- TRINIDAD AND TOBAGO: 61
- TURKS AND CAICOS ISLANDS: 4
- TUVALU: 31
- UNITED STATES OF AMERICA: 4
- URUGUAY: 24
- VANUATU: 50
- VENEZUELA (BOLIVARIAN REP): 50
- VIET NAM: 12366
GUI – Visualize data

SOURCE NAME:
Thailand Dengue Surveillance System: 64

INTERVAL TYPE:
Month: 648

ADMIN LEVEL 1:
- Phuket: 216
- Prachin Buri: 216
- Prachuap Khiri Khan: 216
- Ranong: 216
- Ratchaburi: 216

ADMIN LEVEL 2:
No options available based on your set filters

CUMULATIVE:
No: 648

CITY:
No options available based on your set filters
GUI – Visualize data

PLOT PREVIEW:

(Jan, 1993 - Dec, 2010)

# of cases

Admin 1: PHUKET
Admin 1: PRACHUAP KHIRI KHAN
Admin 1: RATCHABURI
GUI – Visualize data
GUI – Visualize data
<table>
<thead>
<tr>
<th><strong>ColumnName</strong></th>
<th><strong>ColumnType</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>ConditionName</td>
<td>CityName</td>
</tr>
<tr>
<td>ConditionSNOMED</td>
<td>PeriodStartDate*</td>
</tr>
<tr>
<td>PathogenName</td>
<td>PeriodEndDate*</td>
</tr>
<tr>
<td>PathogenTaxonID</td>
<td>PartOfCumulativeCountSeries*</td>
</tr>
<tr>
<td>Fatalities*</td>
<td>AgeRange*</td>
</tr>
<tr>
<td>CountryName*</td>
<td>Subpopulation*</td>
</tr>
<tr>
<td>CountryISO</td>
<td>PlaceOfAquisition</td>
</tr>
<tr>
<td>Admin1Name</td>
<td>DiagnosisCertainty</td>
</tr>
<tr>
<td>Admin1ISO</td>
<td>SourceName*</td>
</tr>
<tr>
<td>Admin2Name</td>
<td>CountValue*</td>
</tr>
</tbody>
</table>
Available data by country and condition
Available data by country and year

Year

Countries
Next steps

- Expand data: online surveillance data, data from research/country partners, age-specific data, climate/demography/vaccination data

- Improve online graphical user interface: selecting data, visualization and analytics

- Integration of research data (level 1) with data from health agencies: Augmented datasets
  - Represent data collection/pre-processing methods
  - Data identifiers and metadata
  - User interface and download options

- Data contributions by others (repository function)
- Sustainable business model
Acknowledgements

Current Faculty, Staff, and Students

Marc Choisy, Institute de la Recherche pour le Développement; Derek Cummings, University of Florida; Ernesto Marques, University of Pittsburgh; Leah Goeke, Kelly Carey, Matt Loiacacono, Austin Chen, Lingshu Xue, Tenley Brownwright, Hannah Polglase, Christian Garcia, Angel Paternina, Danielle Seigneur, Jessica Salerno, Vivian Lin, Rhonda Toth, Scott Chadwell

Partners

Laos Ministry of Health; Taiwan Ministry of Health; Thailand Ministry of Public Health; Cambodia Ministry of Health, Vietnam National Institute for Hygiene and Epidemiology; Philippines Ministry of Health; Singapore National Environment Agency; Malaysia Ministry of Health; Brazil Ministry of Health; U.S. Department of Health & Human Services; US Centers for Disease Control; World Health Organization;

Funders
Take some Project Tycho stickers!

Help improve the discovery of open disease data!

www.tycho.pitt.edu