TYCHO
Open Access to Public Health Data

CSTE Annual Meeting, 15 June 2011

Don Burke, MD
Dean, Graduate School of Public Health
University of Pittsburgh
Tycho Brahe

1546 – 1601

Danish nobleman who made accurate and comprehensive observations of the positions of the stars and planets. After his death, Tycho’s assistant Johannes Kepler used these data to derive the laws of planetary motion.
A project to digitize and render computable all the data in the US weekly National Notifiable Disease Surveillance System, since the beginning of disease reporting, and provide open access to these data

- 1888 to present (more than 120 years)
- 50 states and 1500 cities and towns
- 55 reportable infectious diseases
- 6,300 weekly reports
- 100 million cases and 4 million deaths
Outline

1. Background and vision
2. Development of the Tycho database
3. Applications of the Tycho database
4. Website
5. Future direction
USE OF DATA FOR IMPROVED DECISION-MAKING

Scientific Journals

Expert Opinion

Resource allocation
Disease eradication
Pandemic alert
School closures
Vaccination
Health regulations
Quarantine
Medication stockpiling

Public Health Data

Modeling & Simulation

Media
Some problems with current surveillance data systems:

1. Historical data are only available as hard copy or difficult to find PDFs

2. Modern reports are available on-line, but data are not down-loadable in computable form

3. Data are often available only in very aggregated format (eg state level), with loss of spatial and temporal granular detail

4. There is no single source for comprehensive, accurate, and computable public health disease surveillance data
Endangered public health data: USA and world-wide
Current US NNDSS data online

1) Select year and week
   - 1996-current

2) Select table

3) View but not download
Outline

1. Background and vision

2. Development of the Tycho database

3. Applications of the Tycho database

4. Website

5. Future direction
Sources of US Surveillance Reports

Name of Report

- Weekly Abstract of Sanitary Report
- Abstract of Sanitary Report
- Public Health Report
- Morbidity Mortality Weekly Report

Current availability

- PubMed Central
- Hathi Trust Digital Library
- CDC MMWR
Mortality reports for cities: 1887-1947

Weekly Abstract of Sanitary Reports: 1888 July

| Cities              | Week ended | Estimated population | Total deaths from all causes | Cholera | Yellow fever | Small-pox | Typhus fever | Enteric fever | Scarlet fever | Diphtheria |
|---------------------|------------|----------------------|-----------------------------|---------|--------------|-----------|--------------|---------------|---------------|------------|------------|
City morbidity reports: 1906-1953

<table>
<thead>
<tr>
<th>Place</th>
<th>Cases</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akron, Ohio</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Alameda, Calif.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Albany, N. Y.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Alexandria, Va.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Alliance, Ohio</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Alto, Ill.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Asheville, N. C.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Atlantic City, N. J</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Athens, Mass.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ashland, Ohio.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Buffalo, N. Y.</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Burlington, Vt.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Butte, Mont.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cambridge, Mass.</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Charleston, S. C.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Charlotte, N. C.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chelsea, Mass.</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Jamestown, N. Y.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Jersey City, N. J.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Kalamazoo, Mich.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Kansas City, Kans.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Kansas City, Mo.</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Keene, N. J.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Lackawanna, N. Y.</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>La Faye, Ill.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lawrence, Mass.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Lexington, Ky.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lima, Ohio.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Lincoln, Nebr.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lockport, N. Y.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Logansport, Ind.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Long Beach, Calif.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lorain, Ohio.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Los Angeles, Calif.</td>
<td>24</td>
<td>14</td>
</tr>
<tr>
<td>Louisville, Ky.</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Lynn, Mass.</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Manchester, N. H.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Mankato, Minn.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Marion, Ind.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Marquette, Mich.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Methuen, Mass.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Memphis, Tenn.</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Methuen, Mass.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Milwaukee, Wis.</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>Minneapolis, Minn.</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Mishawaka, Ind.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Montgomery, Ala.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Morristown, N. J.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mount Vernon, N. Y.</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Nashua, N. H.</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Nashville, Tenn.</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>New Britain, Conn.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>New Haven, Conn.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>New London, Conn.</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Public Health Reports: 1920 June
State morbidity reports: 1928-current

MMWR: 1995 January
Data entry by “Digital Divide Data” in Phnom Penh

6,300 weekly reports
35,000 worksheets
200 million keystrokes
Outline

1. Background and vision
2. Development of the Tycho database
3. Applications of the Tycho database
4. Website
5. Future direction
History of disease reporting in the US
Data available per location

- All weekly US NNDSS reports from 1888 to 2009
- All states, > 1500 cities
- 55 diseases
- Total of 100 million reported cases
- Total of 4 million reported deaths
Types of locations

Number of counts per reporting location

- City
- County
- State
- Region
History of disease reporting in the US

Number of conditions included or discontinued per year

-15 -10 -5 0 5 10 15


- New conditions included
- Conditions discontinued
Infectious disease transition over time: average yearly cases in 4 periods

**1900-1924**
- Dengue
- Pellagra
- Encephalitis
- Meningitis
- Poliomyelitis
- Whooping cough
- Typhoid fever
- Pneumonia
- Mumps
- Scarlet fever
- Diphtheria
- Tuberculosis
- Chickenpox
- Influenza
- Measles

**1925-1949**
- Tularemia
- Typhus fever
- Brucellosis
- Meningitis
- Poliomyelitis
- Typhoid fever
- Dysentery
- Diphtheria
- Mumps
- Pneumonia
- Chickenpox
- Whooping cough
- Scarlet fever
- Influenza
- Measles

**1950-1974**
- Meningitis
- Syphilis
- Tuberculosis
- Rubella
- Hepatitis
- Whooping cough
- Pneumonia
- Mumps
- Chickenpox
- Influenza
- Scarlet fever
- Streptococcal sore throat
- Measles
- Gonorrhea

**1975-2009**
- Coccidioidomycosis
- Mumps
- Meningitis
- Lyme disease
- Measles
- Shigellosis
- Syphilis
- Tuberculosis
- Hepatitis
- Aids
- Salmonellosis
- Giardiasis
- Chickenpox
- Gonorrhea
- Chlamydia
The history of measles control in the US

Weekly incidence rates (per 100,000) for US states: 1928-1980
The history of diphtheria control in the US

Z-scores of weekly number of cases in 120 US cities: 1916-1942
- Start of mass vaccination programs in 1928 (red line)
The history of polio eradication in the US

Weekly incidence rates (/100,000) in US states: 1928-1964
- First mass vaccination campaign in 1955-56 (red line)
Pertussis control and resurgence in US

- Weekly pertussis incidence rates (/100,000) in US states for 2 time periods (note scale differences in color legend)
- Approximate time of mass vaccination with whole cell pertussis vaccine in 1948 (red line)
Outline

1. Background and vision
2. Development of the Tycho database
3. Applications of the Tycho database

4. Website

5. Future direction
Roll-out plan for Tycho-USA Website

15 June 2011: Limited “Beta Test” release of version 1.0 at CSTE today

Website access provided to State Epidemiologists and CDC

June –Sept 2011: Testing of Tycho by public health surveillance experts

Feedback to Tycho development team

Changes incorporated

Oct 2011: Public open access to version 2.0
Live demonstration

Tycho website
Welcome

Project Tycho was started in 2009 at the University of Pittsburgh Graduate School of Public Health to address the lack of access to detailed public health data for analysis and policy making. This project was funded by the Bill & Melinda Gates Foundation as part of the Vaccine Modeling Initiative. The goal of project Tycho is to provide a central global public health data access point. Historical as well as current public health data are of great value if accessible for research and analysis. Open access will enable the use of analytical capacity from around the globe for new discoveries of disease patterns and control policies.

The Tycho database currently contains the digital version of the entire US weekly notifiably notifiable disease surveillance system. This system was initiated in 1888 and data continues today. All weekly reports between 1888 and 2009 (4,400 reports) have been entered from PDF files or hard copies using double data entry (200 million keystrokes). Currently, the Tycho database contains 10 million records that each represent a weekly report from a location for a specific disease. All 50 States, 9 territories, and over 4,000 cities have been included. The records include a total of 95 million reported cases and 4 million reported deaths due to notifiable diseases in the United States for the last 112 years.

Currently, the Tycho database is being tested. Invited users have been provided with access to all features of the website and database. A general open access release is planned for later in 2011.

Access the Database

Please log in to access the database.
Home page: www.tycho.pitt.edu

Features accessible after login
Standard search

Project Tycho
Open Access to Public Health Data

Search Tycho

For more search options, use the Advanced Search page.

Disclaimer: Data availability is dependent on historical reporting practice of the US weekly Nationally Notifiable Disease Surveillance System as well as database filters that are currently in place to ensure data consistency (see documentation).

This page shows you the number of records in the Tycho Database for any disease or any location.

- You can search by either Disease or Location.
- You can obtain the number of records reporting Cases or Deaths.
- You can choose to see the number of records per city or per state.
- Use the pull-down menu(s) to select the Disease or location of interest. The items in the pull-down menu show the range of years for which records are available.
- Click the Search button to submit the search.

Search by: Disease Location

Disease outcome: Cases Deaths

Location type: City State

Select disease: MEASLES (1906 - 1956)

Submit
Advanced search

- Comparisons of multiple cities/states
- Aggregation across locations or across time
Outline

1. Background and vision
2. Development of the Tycho database
3. Applications of the Tycho database
4. Website
5. Future directions
Future directions for Tycho dataset

1. Add more USA data (eg complete city level time series)

2. Extend Tycho to other countries:
   
   Partnerships with WHO and country MOH’s

3. Analyze spatio-temporal disease transmission patterns

   “Epidemic ontological units”

4. Link disease patterns to other large scale datasets
   
   (eg Demographics, Environment, Transportation, Evolution)
   
   to discover “laws” of transmission = Kepler activities

5. Improved disease forecasting
Acknowledgements

**Pittsburgh Tycho Development Team:** Wilbert van Panhuis, John Grefenstette, Shawn Brown, Su Yon Jung, Nian Shong Chok, Anne Cross, David Galloway

**University of Pittsburgh:** Steve Wisniewski, Heather Eng, Jia Xu, Raaka Kumbhakar, Vladimir Zadorozhny, Ying-Feng Hsu, Divyasheel Sharma

**Digital Divide Data:** Linda Thomas, Eric Eric Gold, Hetel Patel, Villa Kem and Heak Hok

**Bill and Melinda Gates Foundation:** Steve Buchsbaum

**Discussions and Advice**

**CDC Public Health Surveillance Program Office:** Jim Buehler, Pam Meyer, Kathleen Gallagher, Ruth Jajosky, Pam Diaz

**NCBI:** David Lipman, James Ostell

**Open Government:** Beth Noveck, Aman Bhandari, Todd Park

**CSTE:** Steve Ostroff
Special thanks to the dedicated disease surveillance epidemiologists at local and state health departments and the CDC who have collected surveillance data for the past 122 years!
END