



University of Pittsburgh

Public Health Dynamics Laboratory

Vaccine Decision Information System

SAGE – 25 October 2018

Wilbert van Panhuis, MD PhD

wilbert.van.panhuis@pitt.edu

Content

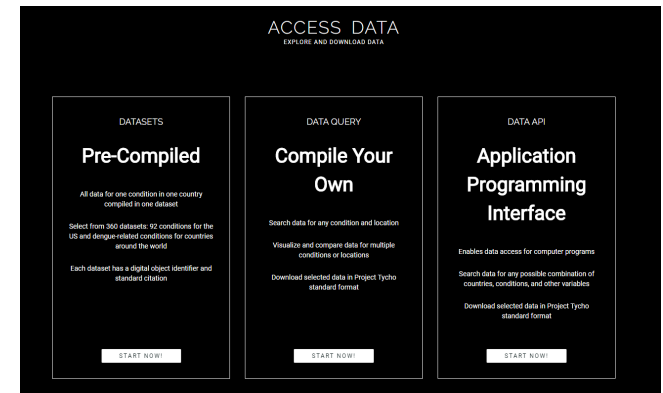
- Unlocking data for research and advocacy: example of the Project Tycho repository
- Using multiple datasets to assess subnational priority areas for vaccination
- Need for interoperability of data systems to enable “using all available data”

Project Tycho: Access to standardized public health data

Difficult-to-use



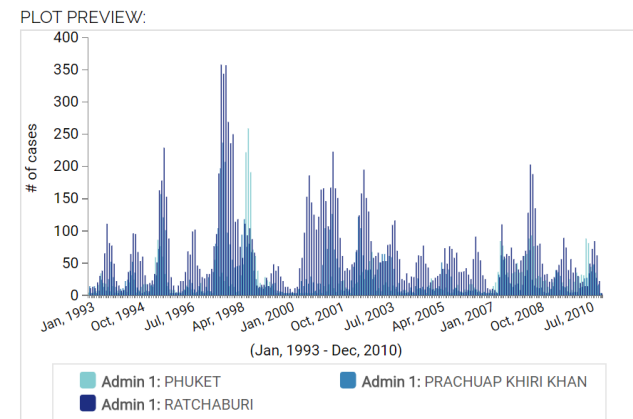
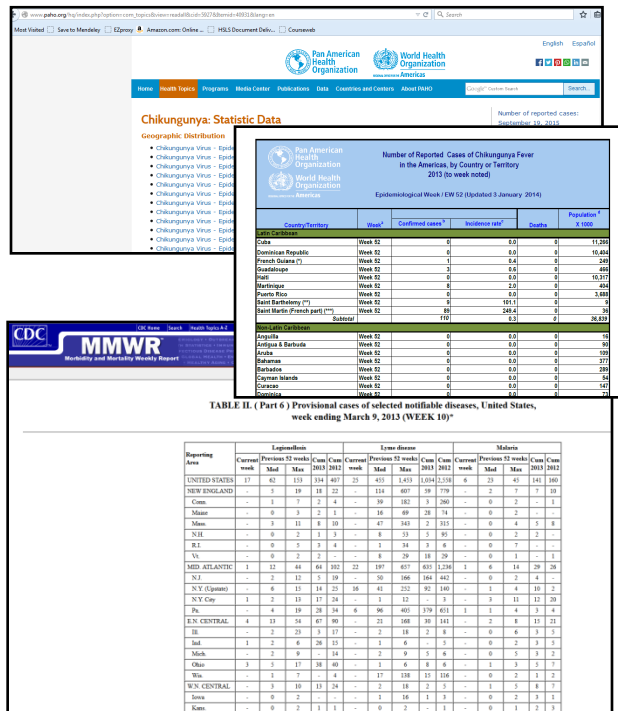
Ready-to-use



Standardization
& Integration

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

~4000 users

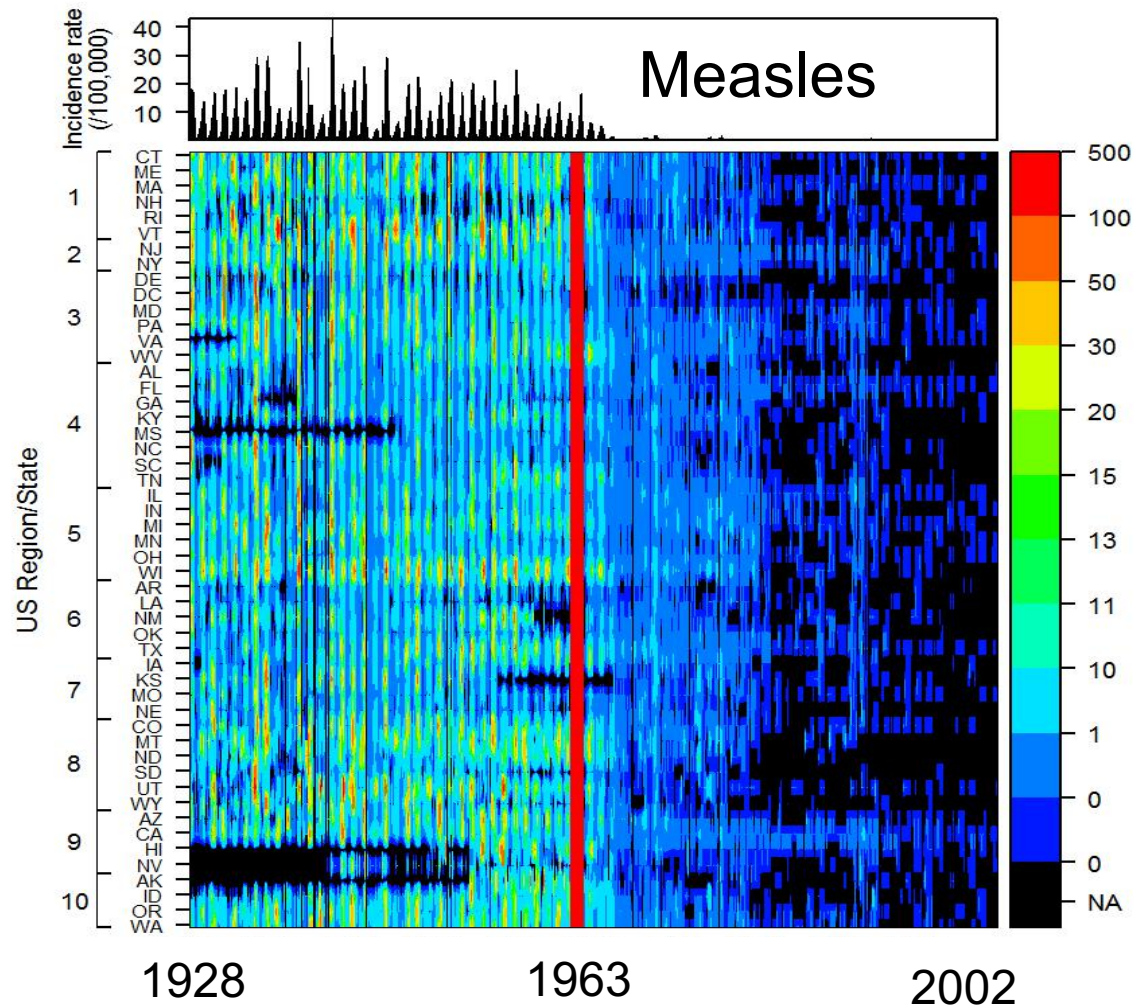


www.tycho.pitt.edu

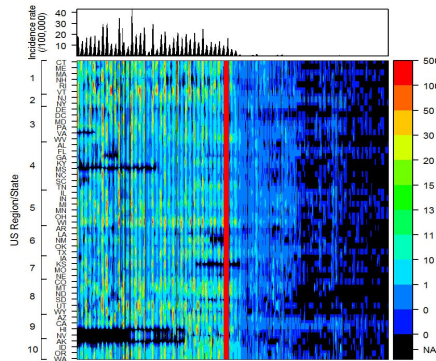
Unlocking data for research



Data digitization = 200
Million keystrokes



Unlocking data for advocacy



The New York Times

The Vaccination Effect: 100 Million Cases of Contagious Disease Prevented

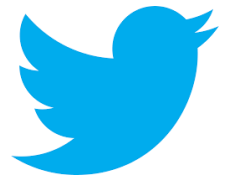
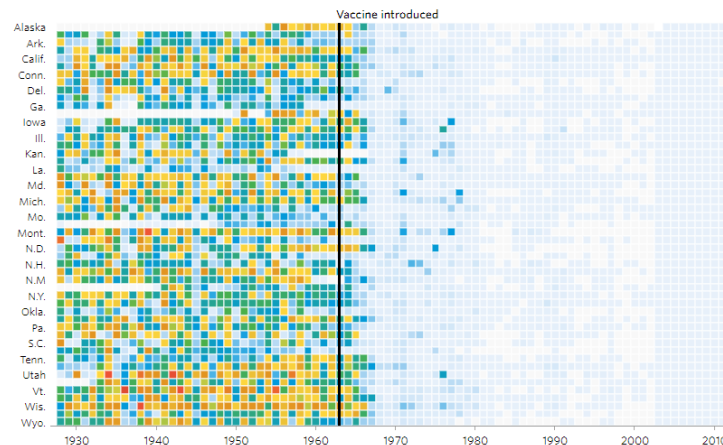
By STEVE LOHR NOVEMBER 27, 2013 5:00 PM 128 Comments

THE WALL STREET JOURNAL.

Battling Infectious Diseases in the 20th Century: The Impact of Vaccines

By Tynan DeBold and Dov Friedman

Published Feb. 11, 2015 at 3:45 p.m. ET



Data waiting to be unlocked....



US, May 2010



Niger, Oct 2009



Laos, Oct 2010



Vietnam, May 2009



Netherlands, April 2014



Nigeria, 2010



Nigeria, 2010



Laos, 2012



Pittsburgh, 2014



India, 2012

Data waiting to be unlocked at WHO



How can we make better use of all available data?

Example 1: Africa 2008-13

RESEARCH ARTICLE

Open Access

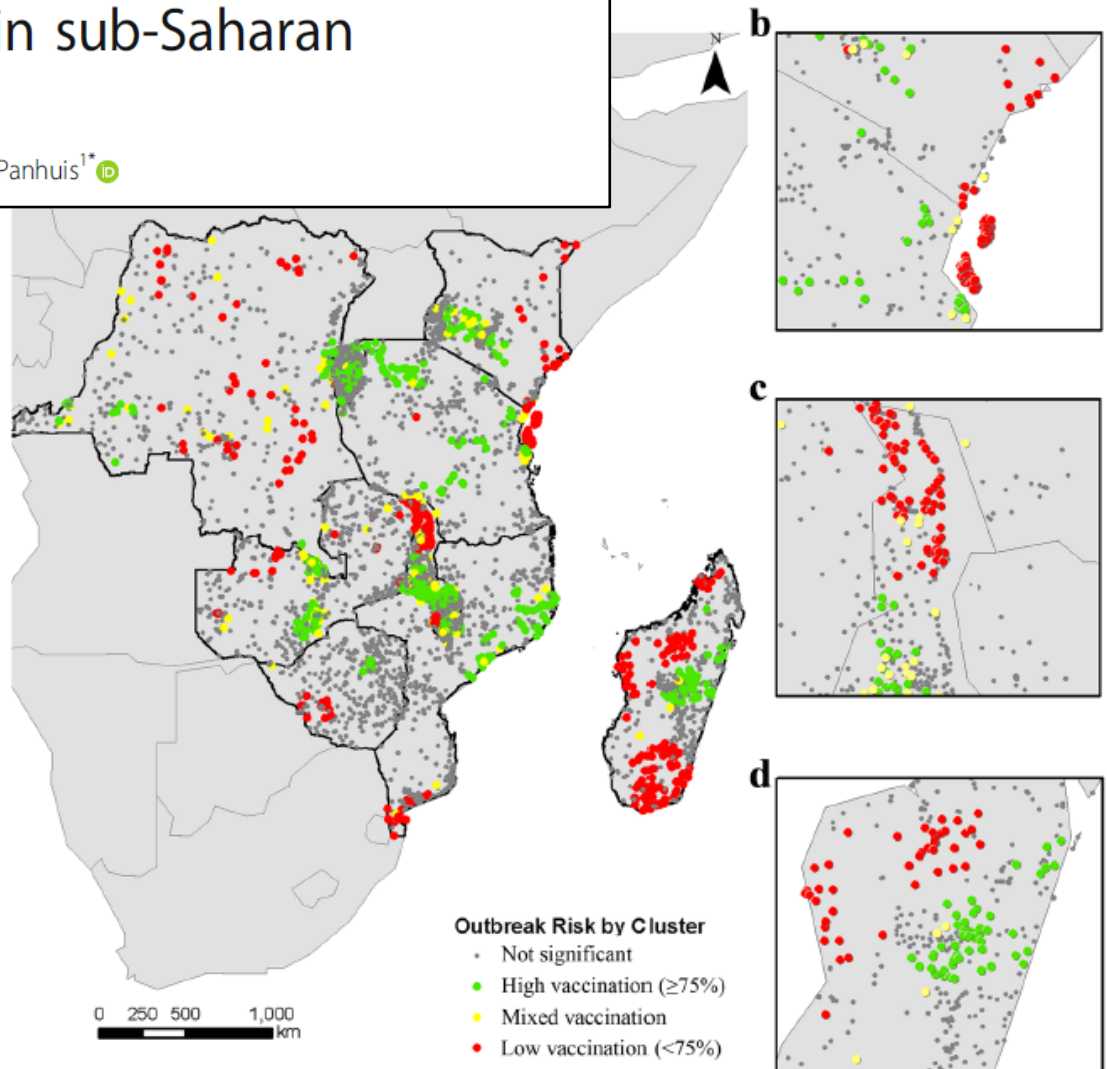
Spatial clustering of measles vaccination coverage among children in sub-Saharan Africa



Tenley K. Brownwright¹, Zan M. Dodson² and Willem G. van Panhuis^{1*} 

Demographic and
Health Surveys:
2008-2013

*Nationwide coverage
rates can mask local
heterogeneity*



Example 2: Malawi 2010

RESEARCH ARTICLE

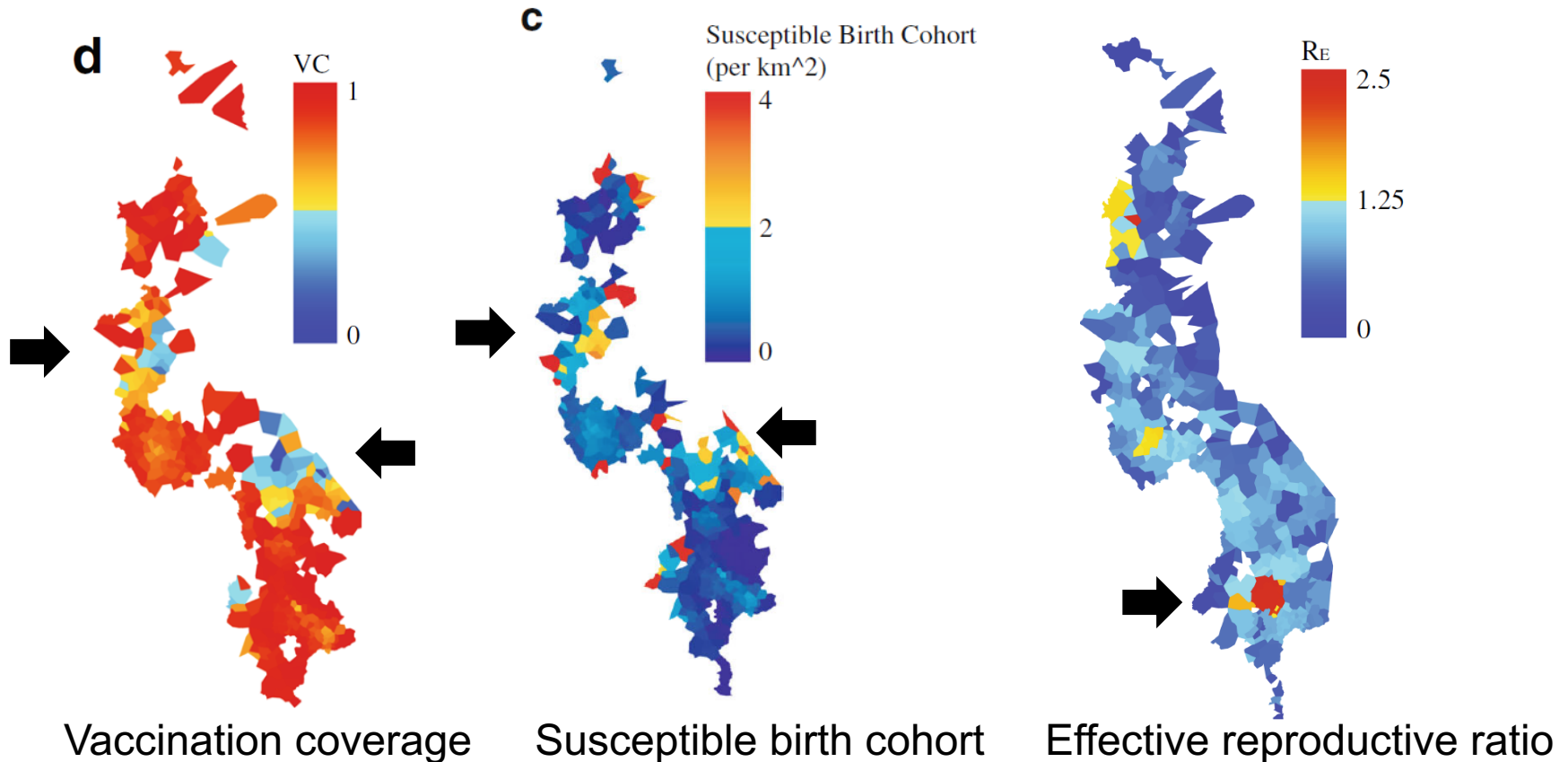
Open Access

Sub-national variation in measles vaccine coverage and outbreak risk: a case study from a 2010 outbreak in Malawi

Avery Kundrick¹, Zhuojie Huang², Spencer Carran², Matthew Kagoli³, Rebecca Freeman Grais⁴, Northan Hurtado⁵ and Matthew Ferrari^{2*}

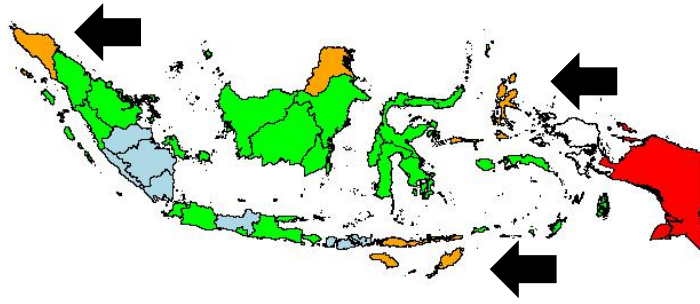


Case-based surveillance and fertility rates -> priority areas for vaccination

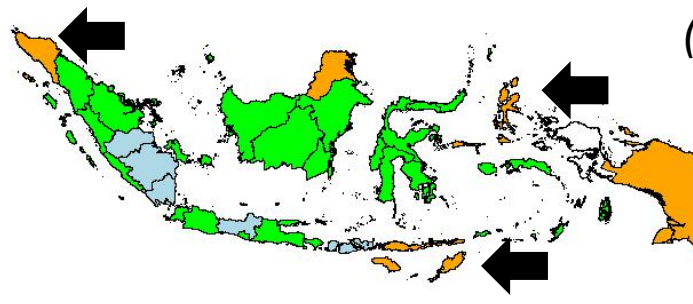


Example 3: Indonesia Diphtheria

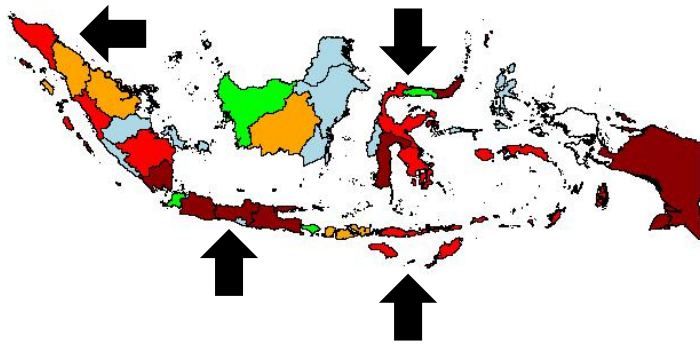
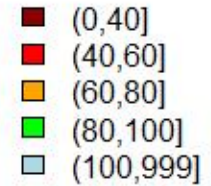
2016



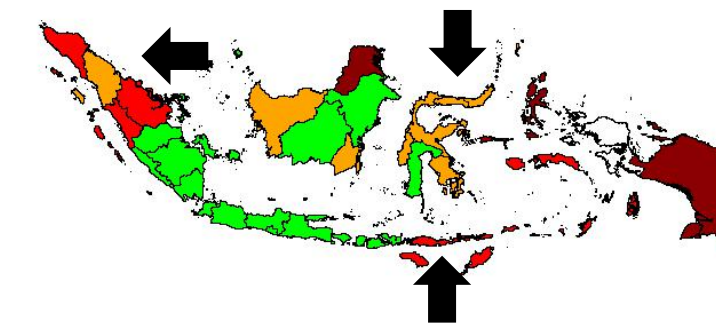
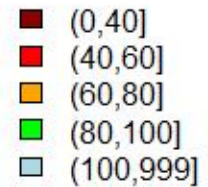
2017



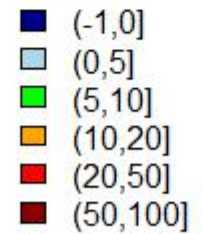
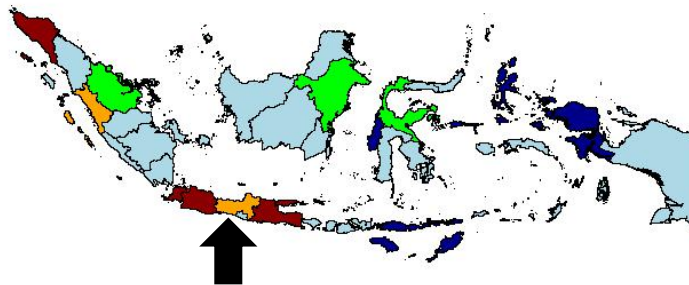
DTP3 coverage
(administrative)



% districts with
DTP3 > 80%



Diphtheria cases



District-level data is
more informative than
province-level data

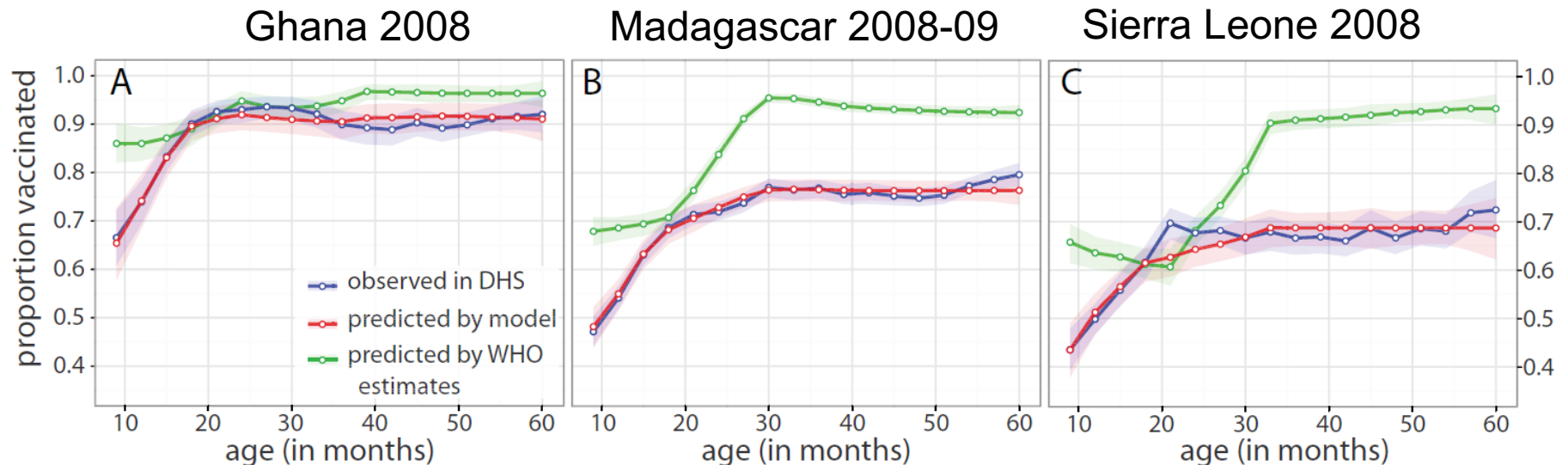
Using multiple data sources together

OPEN ACCESS Freely available online

PLOS MEDICINE

Measuring the Performance of Vaccination Programs Using Cross-Sectional Surveys: A Likelihood Framework and Retrospective Analysis

Justin Lessler^{1*}, C. Jessica E. Metcalf², Rebecca F. Grais^{3,4}, Francisco J. Luquero³, Derek A. T. Cummings¹, Bryan T. Grenfell^{5,6}



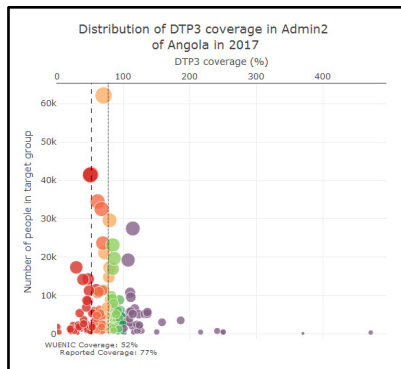
"Combining administrative data with survey data substantially improves estimates of [measles] vaccination coverage"

So far:

- Data are waiting to be unlocked
- Using multiple datasets together leads to better estimates of vaccination coverage or risk areas
- Methods have been developed to use datasets together

Data needs preparation before they can be used for action -> “data readiness”

WHO HQ Subnational coverage data

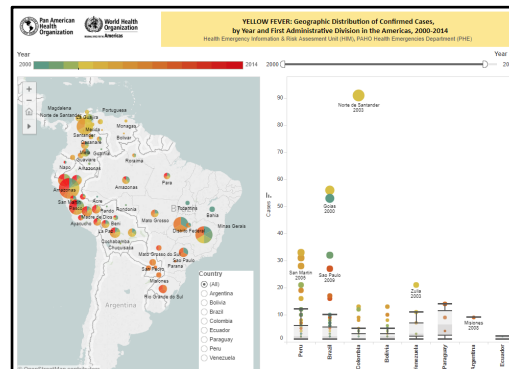


Individual case surveillance data

The **DHS** Program
Demographic and Health Surveys

The Humanitarian
Data Exchange

WHO Regional and country data



 **Cochrane**

 **PubMed**

Census data

Migration data

Model data

Currently in different systems that are not interoperable

Need:
Interoperable
Vaccine Decision
Information
System

What is interoperable?

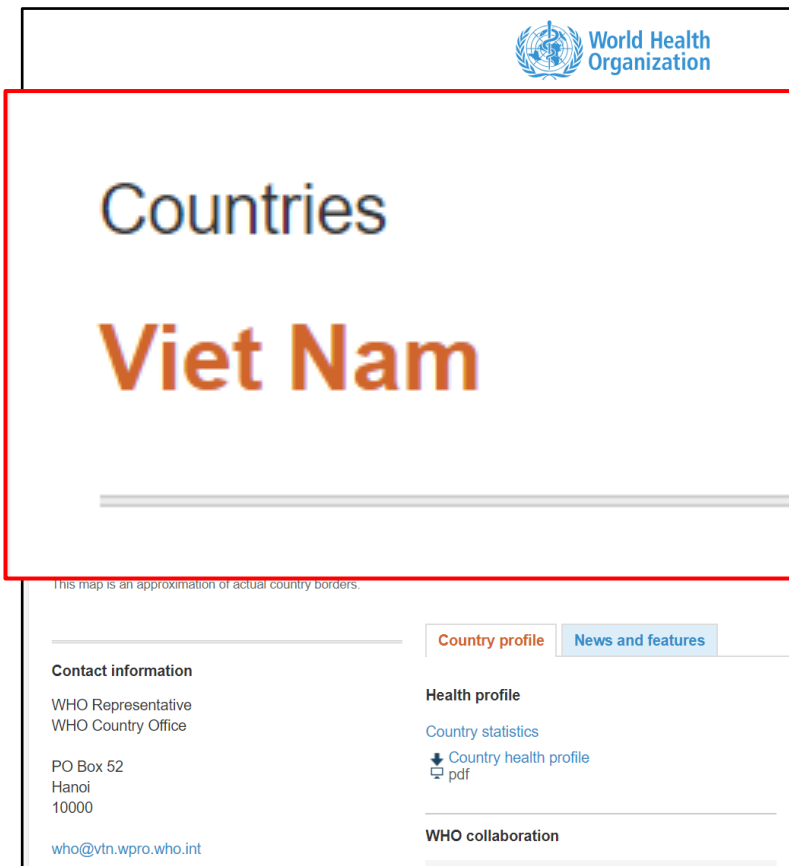
Syntactic interoperability = datasets in the same format
-> minimal *transformation* needed to use together

How: standard data format or data transformation algorithms

Semantic interoperability = datasets contain comparable information / comparable meaning
-> minimal *translation* needed to use together

How: structured vocabularies/ontologies to represent the meaning of information and translation algorithms

No semantic interoperability



World Health Organization

Countries

Viet Nam

This map is an approximation of actual country borders.

[Country profile](#) [News and features](#)

Contact information

WHO Representative
WHO Country Office

PO Box 52
Hanoi
10000


who@vtn.wpro.who.int

Health profile

[Country statistics](#)

[Country health profile](#)
pdf

WHO collaboration



IHME | GHDx | GBD Compare

Vietnam

Life expectancy, 1990-2016

Expected

	1990	2016
Females	67.5	75.0
Males	63.4	69.1

Observed

	1990	2016
Females	71.1	78.1
Males	65.3	70.9

Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says



Gil

I write

Opinion

TWEET THIS



data science
mining



76% of

TECHNOLOGY

For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

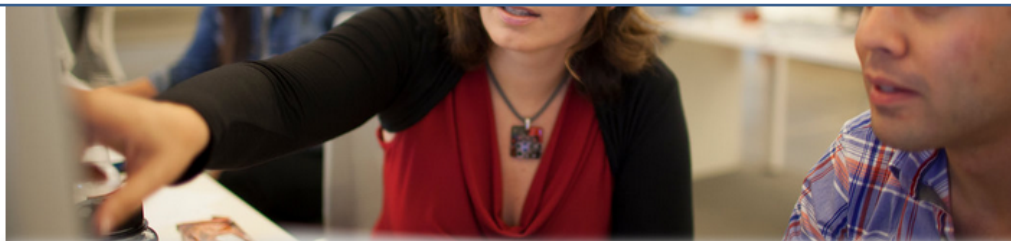
By STEVE LOHR AUG. 17, 2014



Data preparation accounts for about 80% of the work of data scientists

CrowdFlow
scientists. I

Data prep
scientists



Discover it with us. Get The Times at the Academic Rate. Just \$1 a week.

ALREADY A SUB



Discover the truth with us.

Discover it with us. Get The Times at the Academic Rate. Just \$1 a week.

SEE MY OPTIONS



Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says



Gil Weinberg
I write
Opinion

TECHNOLOGY

For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

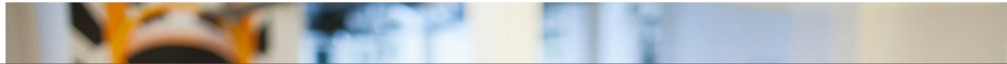
By STEVE LOHR AUG. 17, 2014



TWEET THIS

data science
mining

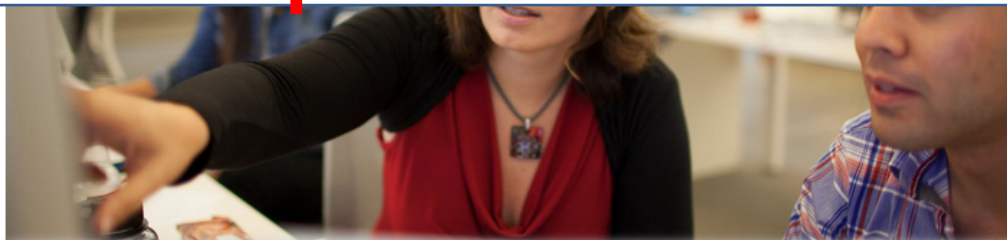
76% of



Data preparation accounts for about 80% of the work of data scientists **and public health officers**

CrowdFlow
scientists.

Data preparation
scientists



Discover it with us. Get The Times at the Academic Rate. Just \$1 a week.

ALREADY A SUBSCRIBER



Discover the truth with us.

Discover it with us. Get The Times at the Academic Rate. Just \$1 a week.

SEE MY OPTIONS



Semantic interoperability

Standard vocabulary:
ISO 3166

Viet Nam = VNM
Vietnam = VNM
VNM =



Countries

Viet Nam

Vietnam

This map is an approximation of actual country borders.

Contact information

WHO Representative
WHO Country Office

PO Box 52
Hanoi
10000

who@vtn.wpro.who.int

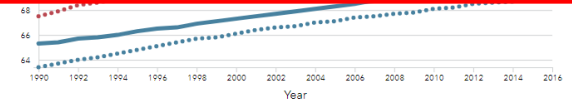
Country profile

Health profile

Country status

Country status
pdf

WHO collaboration



	Expected		Observed	
	1990	2016	1990	2016
Females	67.5	75.0	71.1	78.1
Males	63.4	69.1	65.3	70.9

Life expectancy, 1990-2016

Benefits of an **interoperable** Vaccine Decision Information System

- All available information can be used together to inform vaccination strategy decisions
- Reduction in duplicative/repetitive data cleaning -> cost and time savings
- Information can be accessed by all public health stakeholders and shared with research collaborators
- More efficient quality assessment and targeted data quality improvement strategies

No overhaul of all data systems, but make them interoperable with each other

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Research data

» Publication
characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

Roadmap that defines how to make
datasets **FAIR**:

Findable
Accessible
Interoperable
Reusable

Conclusions

There is value in subnational coverage, case, and other data and information systems should be developed to better use these data on a routine basis.

Routine administrative and WHO data should be augmented with other data sources to identify priority areas for vaccination strategies.

Different siloed vaccine-related data systems should be made interoperable, using FAIR data principles, to prevent redundancies and save time and cost

Acknowledgements



Donald Burke, MD
Dean, Graduate
School of Public
Health



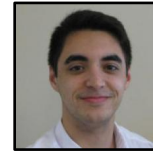
Mike Wagner, MD PHD
Professor of Biomedical
Informatics and Intelligent
Systems



Anne Cross, MLIS
Database manager



John Levander
Software design,
MIDAS ISG



Max Sibilla
Software design,
MIDAS ISG



Tejaswi Anataraju, MS
Data programmer



Sharon Crow, MEd
Project Coordinator

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

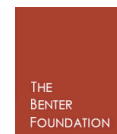
Laos Ministry of Health; Taiwan Ministry of Health; Thailand Ministry of Public Health; Cambodia Ministry of Health, Vietnam National Institute for Hygiene and Epidemiology; Philippines Ministry of Health; Singapore National Environment Agency; Malaysia Ministry of Health; Brazil Ministry of Health;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

Funders

BILL & MELINDA
GATES *foundation*



National Institute of
General Medical Sciences



Extra slides

SCIENTIFIC DATA



OPEN

SUBJECT CATEGORIES

» Research data

» Publication

characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

Principle	Key attributes of FAIR research products
Findability	Unique identifier, rich metadata, indexed in catalogues
Accessibility	Retrievable through identifier, universal access protocol
Interoperability	Formal, standardized format and content
Reusability	Metadata has many relevant attributes, usage license

FAIR principles for scientific data management and stewardship

<http://www.nature.com/articles/sdata201618>

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards