



Institut-Hôpital  
neurologique de Montréal  
Montreal Neurological  
Institute-Hospital

November 18, 2019 | #OpenScienceinAction | [www.theneuro.ca](http://www.theneuro.ca)



# A Framework for Sharing High-Risk Data

**Open Science in Action**

Isabella Chu, MPH

Associate Director, Data Core, Stanford Center for Population Health Sciences

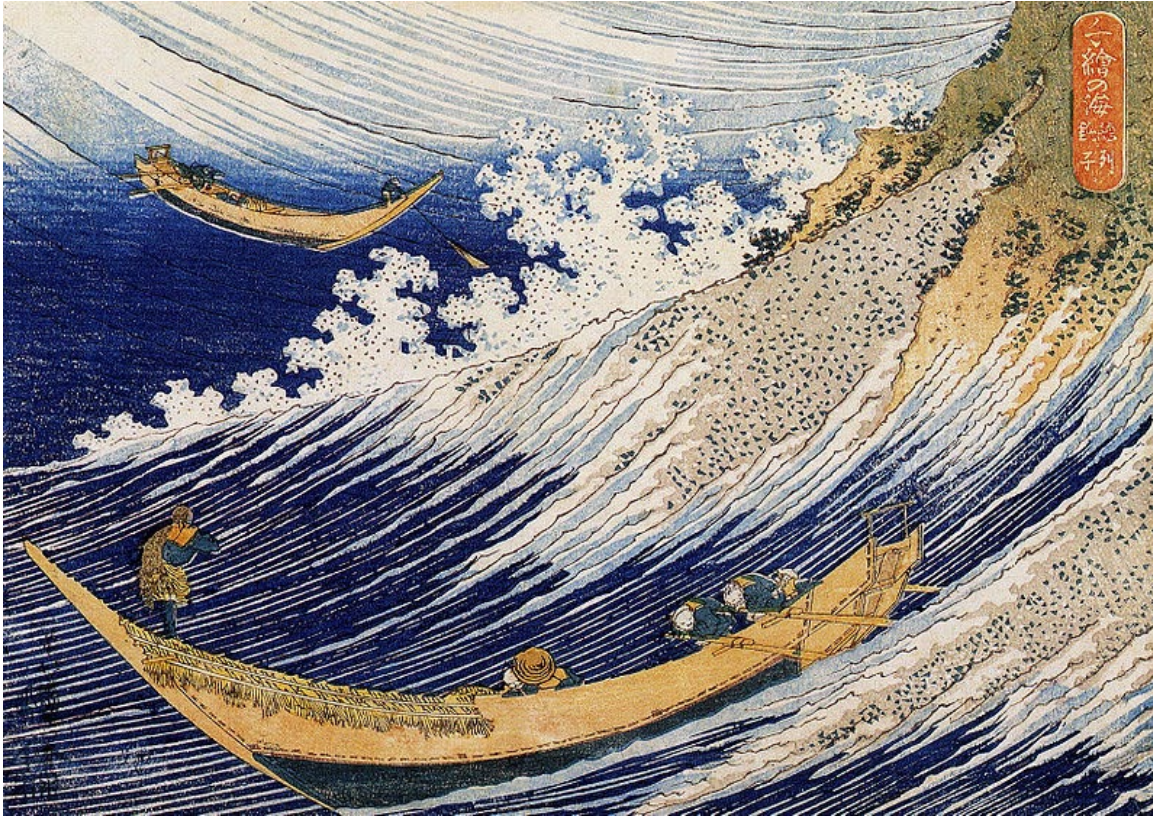
November 18, 2019





**Open Science,  
to accelerate discovery  
and deliver cures**

# The challenge: Large, high-risk datasets



- Commercial Claims
  - 58 million, 8 years
  - 149 million, 7 years
- Medicare
  - 11 million, 8 years
- EMRs
  - 80 million

## Traditional Model

## PHS Data Center

Discoverability

Personal relationship or literature review



Searchable using standard ontologies such as Medical Subject Headings ([MeSH](#))

Access

Available across disciplines

Administration

Streamlined processes

Data Quality

Common model for interoperability

Hosting

Standard ontologies such as Medical Subject Headings ([MeSH](#))

Computing Capacity

Limited bandwidth and ad hoc security for large-scale analysis



Secure, cloud computing for performance and scalable use



**Goal:** Research innovation—*not disparate data access*—defines the competitive advantage.

## Traditional Model

## Optimal Model

Discoverability

Personal relationship or literature review



Searchable using standard ontologies such as Medical Subject Headings (MeSH)

Access

Limited to single teams and/or discipline



Data and agreements available across teams and/or disciplines

Administration

Dependent on research team knowledge and management



Centralized, automated processes

Data Quality

Fragmented, un-harmonized data sets requiring individual cleaning



Transformed data into common model for ease of replication and interoperability

Hosting

Individual computers or lab-specific servers



Cloud storage for ease of access and customization

Computing Capacity

Limited bandwidth and ad hoc security for large-scale analysis



Secure, cloud computing for performance and scalable use

# A Data Ecosystem for High-Risk Data

## Model Data Ecosystem

Goal : Combine disparate datasets together to answer population health questions

Strategic Priorities: ✓ *Data Governance* ✓ *Data Management* ✓ *Infrastructure* ✓ *Incentives* ✓ *Capacity Building*

### Secure Cloud Storage

Capable of storage for data of many sizes and types



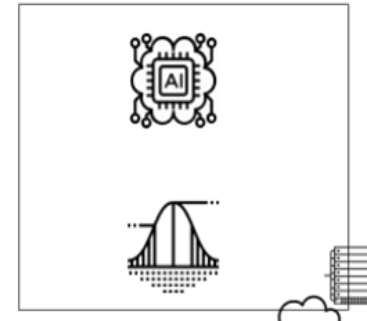
### Data Discovery & Access

Host metadata, at a minimum  
Manage access



### Compute platforms

Flexible depending on risk-profile,  
data type, analysis and budget



### Reproducibility and Knowledge Sharing

Academic websites,  
Github, etc.



Stanford  
affiliated user

1. Explore metadata;  
Frequency counts
2. Complete access  
requirements, as needed

1. Generate analytical  
datasets (link data,  
apply research criteria  
and filters)
2. Perform analyses
3. Generate statistical and  
machine learning  
models


1. Share findings
2. Share code and  
methods
3. Share data

🔒 High risk



# PHS Data Portal

Log in Sign up

 **Stanford**  
MEDICINE  
Center for  
Population Health Sciences

## Stanford Center for Population Health Sciences

★ Star 12

Overview Dataset groups Datasets About

☰ StanfordPHS Entities

- ☰ patient per calendar ...
- ☰ pharmaceutical claim
- ☰ households
- ☰ persons
- ☰ single hospitalization...
- ☰ laboratory test result...
- ☰ claim per patient per ...
- ☰ eligibility period spa...

### At a Glance

Founded in 2015, the Center's mission is to improve individual and population health by bringing together diverse disciplines and data to understand and address social, environmental and biological factors on both a domestic and global scale.

[See more >](#)

<b>13</b> Dataset Groups >	<b>42 billion</b> Records >	<b>87</b> Datasets >
-------------------------------	--------------------------------	-------------------------

[phsdata.stanford.edu](https://phsdata.stanford.edu)

## Looking forward

- Research on incentives
- New tools
- Grants to share what we've learned.
- New ontologies
- Data management and annotation
- Data landscaping



**Join the conversation online!**



**@TheNeuro\_MNI**  
**#OpenScienceinAction**