上海交通大学医学院附属 新華醫院
Xinhua Hospital Affiliated To Shanghai Jiaotong University School Of Medicine

# Initiatives for Data Harmonization & Sharing Across Biobanks and Cohorts

**Charles W. Wang, MD, PhD**

**MOE Key Lab of Environmental and Children's Health**
**Xinhua Hospital, School of Medicine**
**Shanghai Jiao Tong University**
**Shanghai, China**

# Outline

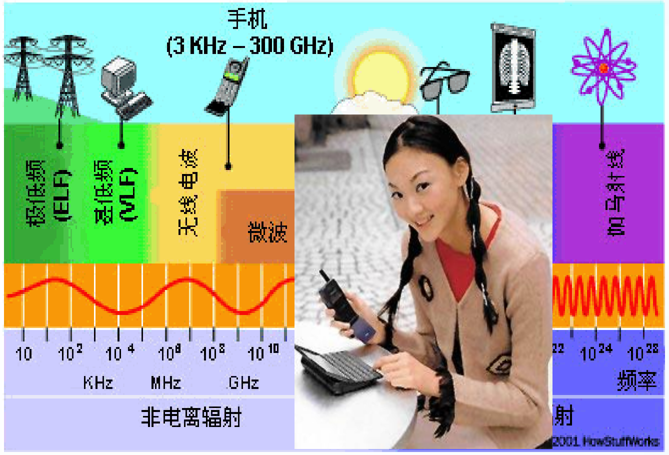


- **Shanghai Birth Cohort (SBC) – Introduction**
- **Analysis across Cohorts/Biobanks – Data Harmonization**

# Environmental Pollution

# Emerging Exposures

Flame retardants




Electronic waves

Plastic additives
PFOS(A)
Triclosan

Formaldehyde, flame retardants

# China Accounts for ¼ of Global Mercury Emission



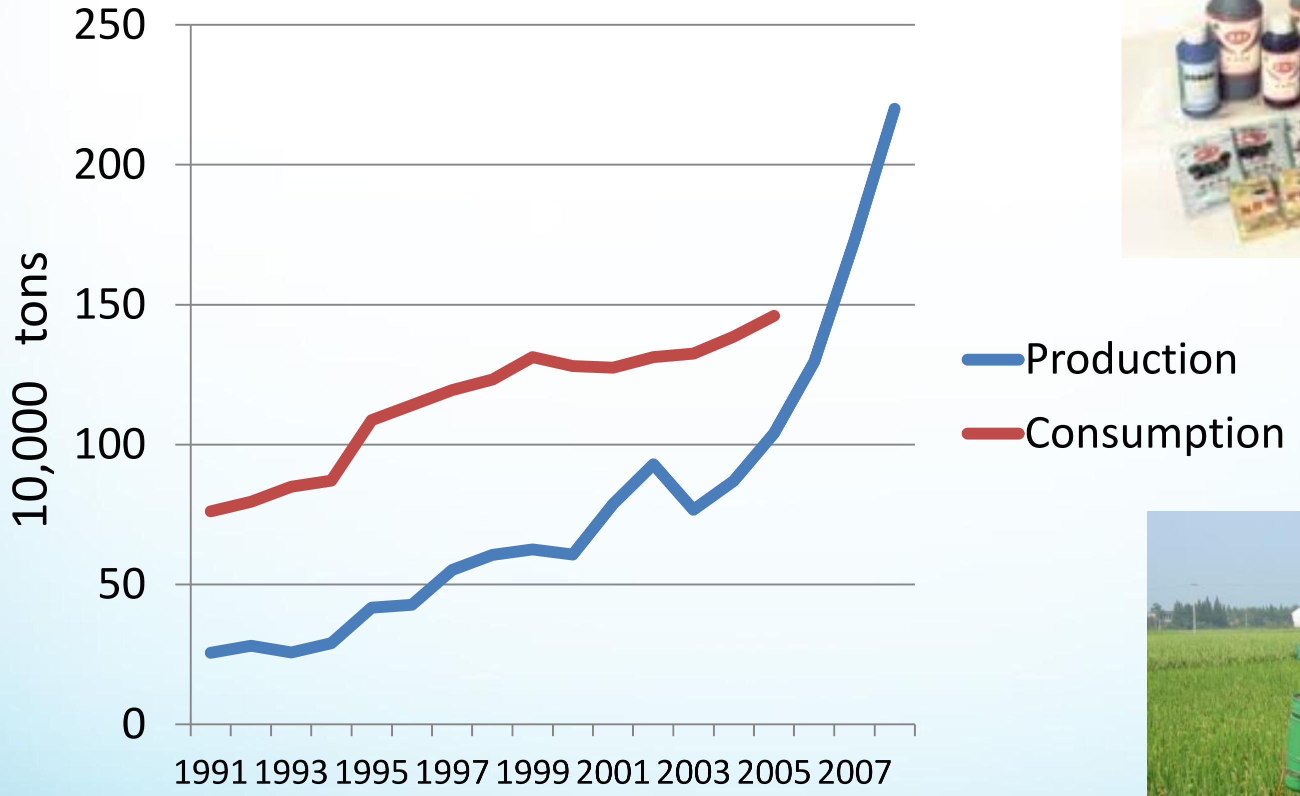
**Table 5.** Global emissions of total mercury from major anthropogenic sources ($Mg\ yr^{-1}$).

| | SC[a] | NFMP | PISP | CP | CSP | MP | GP | WD | O | T | Reference year | Reference[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S. Africa | 32.6 | 0.3 | 1.3 | 3.8 | 0.0 | 0.0 | 0.3 | 0.6 | 1.3 | 40.2 | 2004 | (1) |
| China | 268.0 | 203.3 | 8.9 | 35.0 | 0.0 | 27.5 | 44.7 | 14.1 | 7.6 | 609.1 | 2003 | (2) |
| India | 124.6 | 15.5 | 4.6 | 4.7 | 6.2 | 0.0 | 0.5 | 77.4 | 7.5 | 240.9 | 2004 | (3) |
| Australia | 2.2 | 11.6 | 0.8 | 0.9 | 0.0 | 0.0 | 0.3 | 0.2 | 0.6 | 16.6 | 2005 | (4) |
| Europe | 76.6 | 18.7 | 0.0 | 18.8 | 6.3 | 0.0 | 0.0 | 10.1 | 14.7 | 145.2 | 2005 | (5) |
| Russia | 46.0 | 5.2 | 2.6 | 3.9 | 2.8 | 0.0 | 4.3 | 3.5 | 1.5 | 69.8 | 2005 | (5) |
| N. America | 65.2 | 34.7 | 12.8 | 15.1 | 10.3 | 0.0 | 0.0 | 13.0 | 1.7 | 152.8 | 2005 | (6) |
| S. America | 8.0 | 13.6 | 1.8 | 6.4 | 2.2 | 0.0 | 16.2 | 0.0 | 1.5 | 49.7 | 2005 | (5) |
| Total | 623.2 | 302.9 | 32.8 | 88.6 | 27.8 | 27.5 | 66.3 | 118.9 | 36.4 | 1324.3 | | |
| Rest of the world | 186.8 | 7.1 | 10.4 | 147.1 | 135.1 | 22.5 | 333.7 | 68.5 | 28.2 | 939.4 | 2006 | (7) |
| Total | 810.0 | 310.0 | 43.2 | 235.7 | 162.9 | 50.0 | 400.4 | 187.4 | 64.6 | 2319.7[c] | | |

Pirrone et al. Atmos Chem Phys 2010;10:5951-64.

# Pesticides in China

上海交通大学医学院附属新华醫院
Xinhua Hospital Affiliated To Shanghai Jiaotong University School Of Medicine




**Ministry of Agriculture of China (stats.gov.cn)**
**Proc Intl Acad Ecol Environ Sci 2011;1:125-44.**

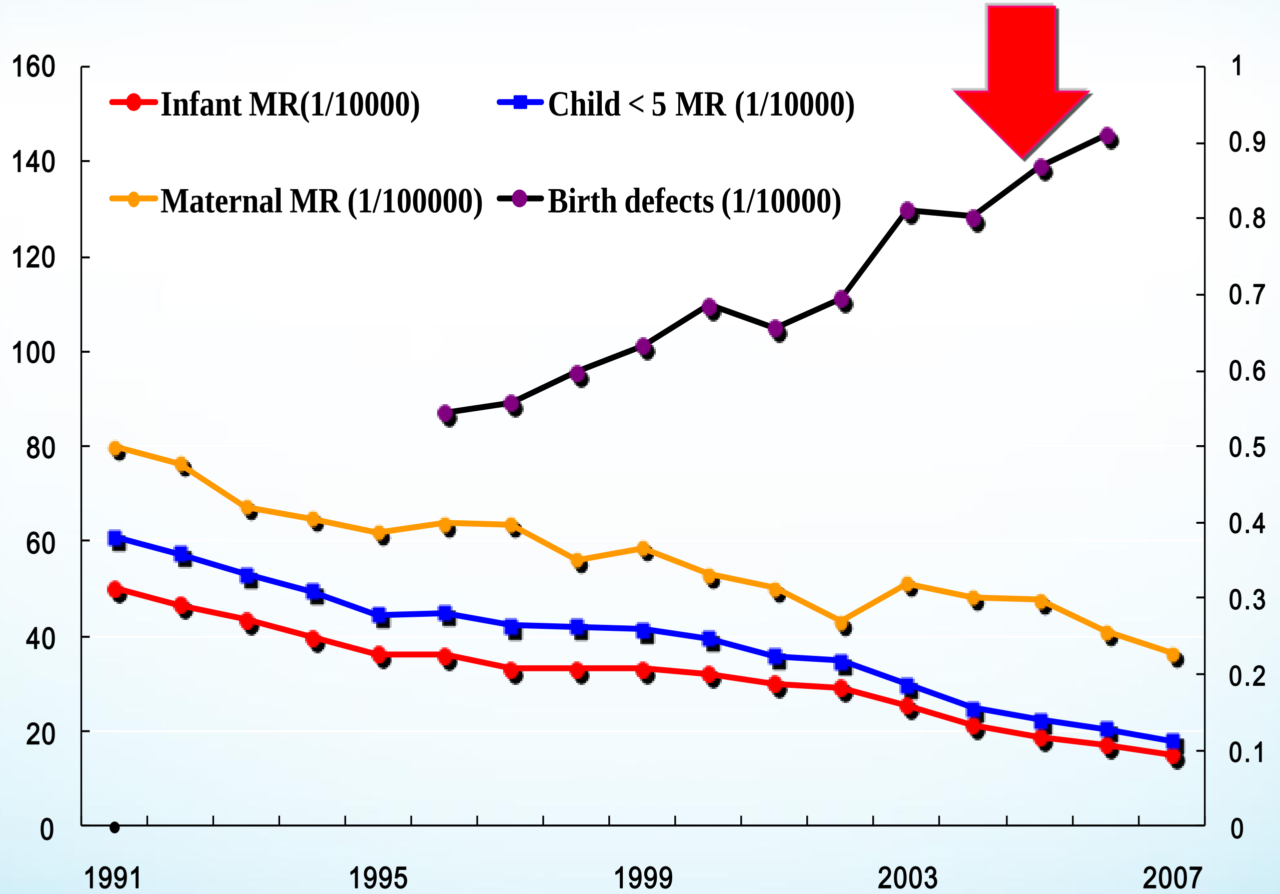
# Developmental Origins of Health and Diseases (DOHaD)

Link

**Child Diseases：Congenital Anomalies, ADHD, Autism, Asthma, Mental Retardation**
**Adult Diseases: Cardiovascular Diseases, Diabetes, PCOS, Cancer, Psychiatric Disorders, Osteoporosis,**

# Incidence of Birth Defects

上海交通大学医学院附属 新華醫院
Xinhua Hospital Affiliated To Shanghai Jiaotong University School Of Medicine




Source: 中国卫生部和中国国家统计局 2008

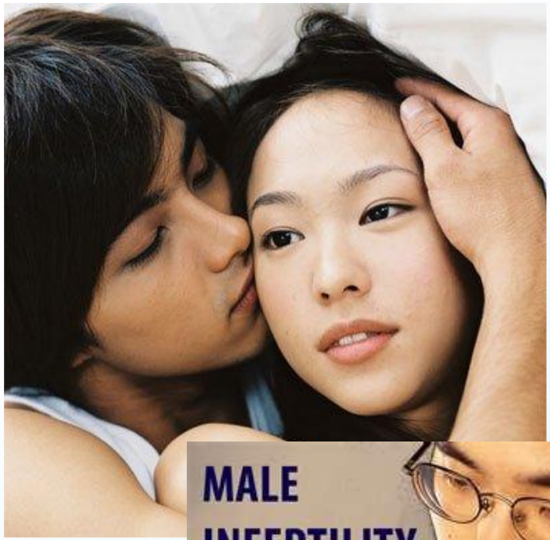
# Possible Impact on Reproduction

- In 1988, infertility rate in a national survey was 6.9%
- In 2010, primary infertility rate was 10-12%

《中华计划生育杂志》2011

- 40 million infertile people

2010《中国不孕不育现状调研报告》



# Childhood Diseases in China



- **Asthma survey**
  - Chongqing 3.34% in 2000; 7.45% in 2010

- **Between 1996 and 2006, prevalence of overweight and obesity in children aged 0 – 6 years increased 4-5 times**
  - In 2006, overweight = 19.8% ; obese = 7.2%

中华儿科杂志 2008;46:179-84.

# Mission – Translated Research

Identify questions from clinical practice → Conduct scientific research → Translate results into health policy → Improve child health

**Provide Evidence for Environment and Health-Related Policy Making and Translational Medicine**

# SBC at A Glance



To study the effects of genetic, environmental and behavioral factors on reproductive health, pregnancy outcomes, child growth, development and risks of diseases.

| preconception | pregnancy | infancy | childhood | adolescence |
| --- | --- | --- | --- | --- |
| *infertility* | *miscarriage prematurity fetal growth restriction, stillbirth* | *birth defect metal retardation* | *asthma, ADHD autism, obesity precocious puberty* | *mental , behavioral & endocrine disorders* |





# Visit Schedule: Pregnancy



Hospital

**Preconception :**
consent
Interview, sample
Partner
Telephone follow-up

**Early：**
(≤16 weeks)
(consent)
Interview
Sample

**Mid, late：**
(22-28，32-36 weeks)
Interview
Sample

**Birth：**
Physical measures chart abstraction
Samples

Samples

Blood, urine

Blood, urine

Blood, urine, hair, nail

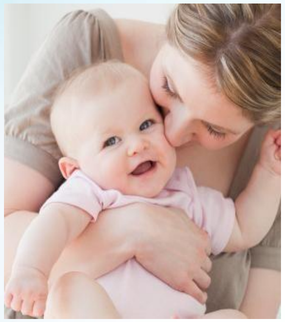
Cord blood, placenta, blood spot, maconium, father buccal swab

Home visit

Environmental sampling
Diet, nutrition, environment questionnaire

# Visit Schedule: Child

上海交通大学医学院附属 新華醫院
Xinhua Hospital Affiliated To Shanghai Jiaotong University School Of Medicine




Hospital

**42 day：**

Postpartum health
Feeding, habit
Physical measure
Neonatal diseases

**6-month：**

Feeding, habit
ASQ
Physical measure
Disease history

**12-month:**

Feeding, habit, environment,
ASQ
Physical measure
Disease history

**24-month：**

Feeding, habit, environment
ASQ, M-CHAT
Intelligence test
Physical measure
disease history

Sample

milk

Urine?

Blood, urine, hair, nail

Tier II

Psychology & behavior
Family environment

Psychology & behavior

# Data and Sample Collection

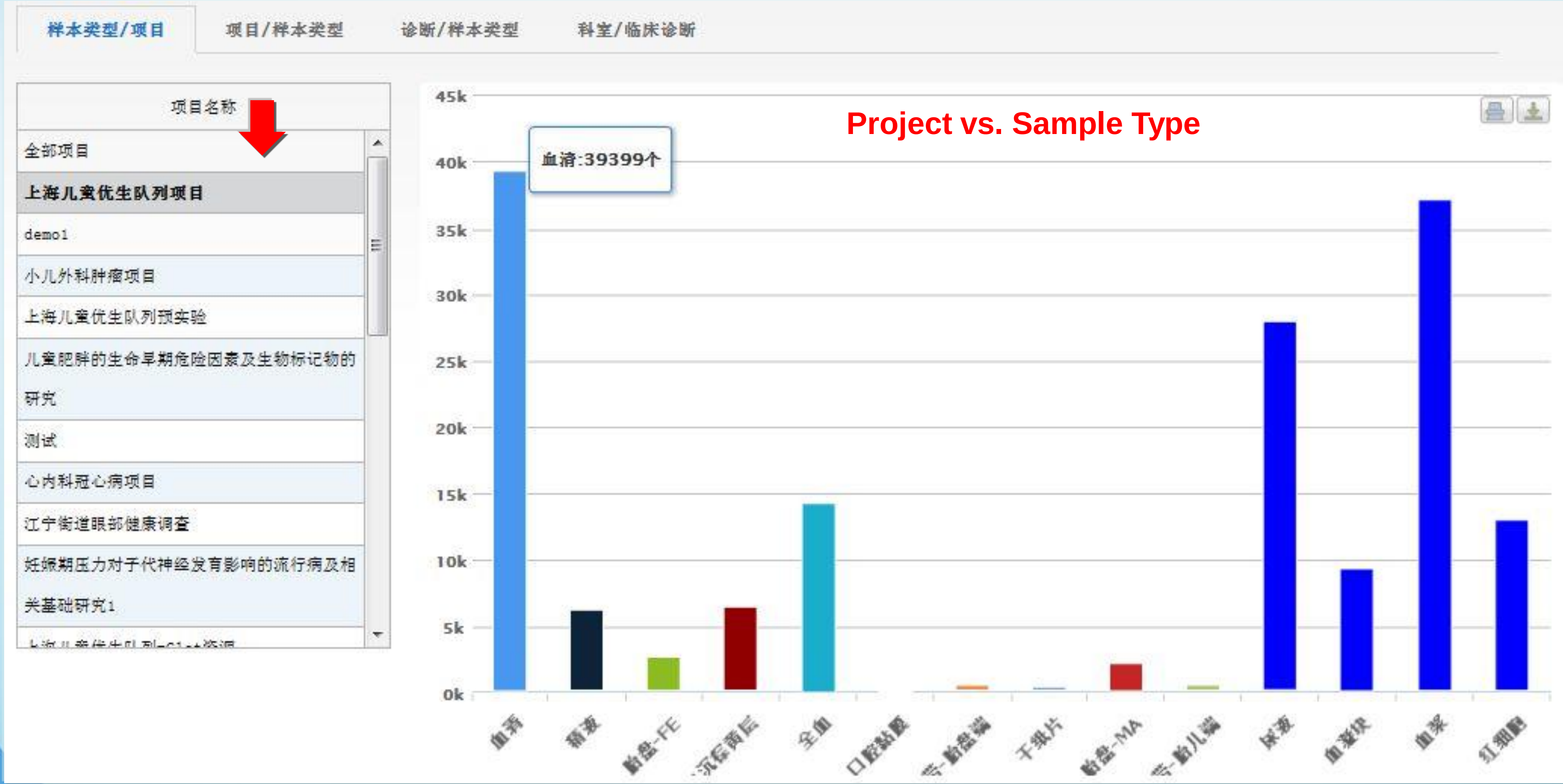
**Interoperability**

# Sample Type vs. Temperature

| 样本类型 | 分装容积 | 储存温度 |
| --- | --- | --- |
| 全血 | **0.5 ml** | **-80C** |
| 血浆 | **0.5 ml** | **-80C** |
| 血清 | **0.5 ml** | **-80C** |
| **PBMC** 白细胞层 | **1 ml** | **-80C** |
| **RBC** | **1 ml** | **-80C** |
| 血凝块 | **N/A** | **-80C** |
| 尿液 | **15 ml** | **-20C** |
| 头发 | **>20** | **Ambient** |
| 干血纸片 **(DBS)** | **1** | **-20C** |
| 指甲 | **>10** | **Ambient** |
| 胎粪**(meconium)** | **2** | **-80C** |
| **Breast Milk** | **1 ml** | **-80C** |
| 脐带血 | **1 ml** | **-80C** |
| 胎盘和脐带 | **N/A** | **-80C** |
| | | |

# Project-based Samples

上海交通大学医学院附属 新华醫院
Xinhua Hospital Affiliated To Shanghai Jiaotong University School Of Medicine



# Key Scientific Questions of SBC

1. Environmental endocrine disrupters on infertility, abortion and adverse pregnancy outcomes.
2. Environment-gene interaction on birth defects
3. Pregnancy Stress and Micronutrients on child development and diseases
4. Early life Exposure to Environmental Pollutants on Children's Neurological and Mental Development and Allergies
5. Environmental endocrine disrupters on Child Obesity and Child Precocious Puberty
6. Early Life Familial and Social Environment on Adolescent Psychological and Behavioral Development

# Questionnaire

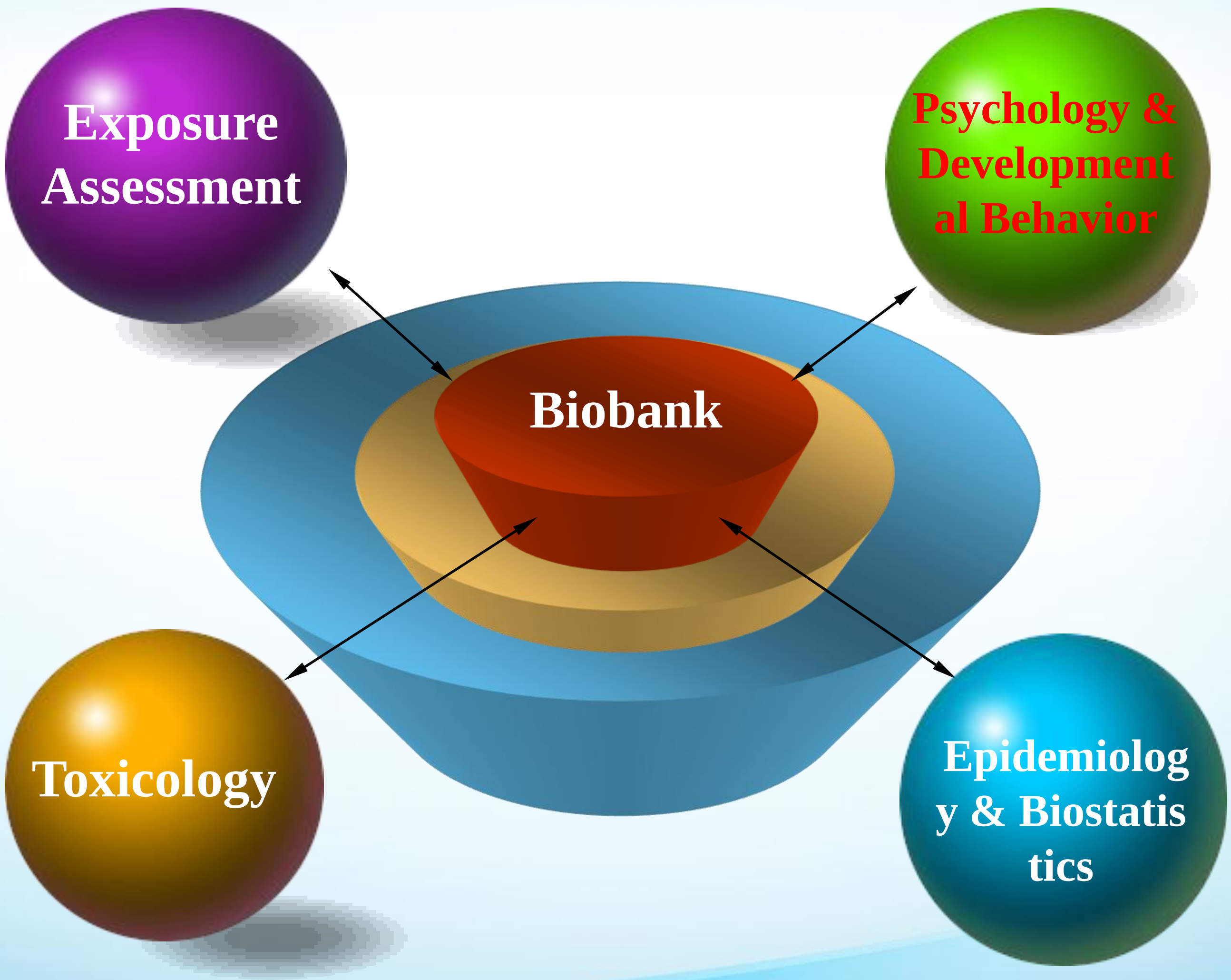


- Questionnaire
  - Socio-economic status
  - Social support
  - Health behavior： physical activity, sleeping, smoking, alcohol, tea, drugs
  - Reproductive history
  - Medical history
  - Medication and supplements*
  - Family history

- Environment, occupation
- Psychology: stress, anxiety and depression
- Diet and nutrition
- Infant feeding and habit
- Family and community environment
- Child developmental tests
- Child ASQ，M-CHAT
- Child psychological behavior
- Child diseases

# Research Platforms





# Heterogeneity vs. Inoperability



- There's confusion when we talk about it because we are not always talking about the same thing；
- Inoperability is critical to minimize heterogeneity but maximize the value of cohorts/specimens for sharing
- We need to better understand similarity and difference across studies and resources.

# Goal

上海交通大学医学院附属新华医院
Xinhua Hospital Affiliated To Shanghai Jiaotong University School Of Medicine

- Etiology study of diseases, especially the rare diseases, requires large number of cases and biological samples.
- The birth cohorts by Canadian and Shanghai share much in common.
- Incompatibility of datasets across cohorts and ethical and legal issues challenge sharing and collaboration.
- Thus harmonization of cohort data, and an infrastructure to boost statistic power of cohort study analysis.

# Example：Data Collection



**Study 1:** In the last month, were you exposed to secondhand smoke at home (Y/N)?

**Study 2:** Does your husband smoke when at home?

**Need to generate compatible data:**
exposure to secondhand smoke at home

**Study 3:** How many people smoke at home (excluding yourself)?

The way to question, data collection and format, for example: smoke, you smoke, other smoke, site, degree of exposure, etc.

**Study 4:** Does anyone of your family living together with you smoke? (Y/N)?

# Harmonization and Federation



1. Document study
2. Define variables targeted for harmonization
3. Assess harmonization potential
4. Develop data processing algorithms for Harmonized Datasets
5. Interconnect harmonized databases for federated data analysis

# Study Documentation



| Name | Description | Start | End |
|---|---|---|---|
| MIREC Prenatal Visit 1 | MIREC first trimester (between 6 to 13 weeks gestation) visit with the mother. | 2008 (February) | 2011 (March) |
| MIREC Prenatal Visit 2 | MIREC second trimester (between 16 to 21 weeks gestation) visit with the mother. | 2008 (May) | 2011 (June) |
| MIREC Prenatal Visit 3 | MIREC third trimester (between 32 to 34 weeks gestation) visit with the mother. | 2008 (July) | 2011 (September) |
| MIREC Post-Partum | MIREC post-partum (day 1 or 2) follow-up. | 2008 (September) | 2011 (October) |
| MIREC Delivery | MIREC at delivery follow-up. | 2008 (September) | 2011 (October) |
| MIREC Milk Collection | MIREC post delivery (2 to 10 weeks) follow-up. | 2008 (October) | 2012 (April) |
| MIREC-ID Birth | MIREC-ID (MIREC: infant development to 6 months) follow-up at birth. | 2010 (January) | 2011 (December) |
| MIREC-ID 6 months | MIREC-ID (MIREC: infant development to 6 months) follow-up at 6 months of age. | 2010 (August) | 2012 (May) |
| MIREC-CD3 | MIREC-CD3 (MIREC: child development at age 3) survey. (Ongoing) | 2012 (August) | 2014 (March) |
| MIREC-CD Plus Biomonitoring | MIREC-CD PLUS Biomonitoring visit (Children aged 15 to 35 months). (Ongoing) | 2013 (March) | 2015 (March) |
| MIREC-CD Plus Neurodevelopment | MIREC-CD PLUS Neurodevelopment visit (Children aged 36 to 47 months). (Ongoing) | 2013 (April) | 2015 (March) |

# Comparison：Data Dictionaries

CN

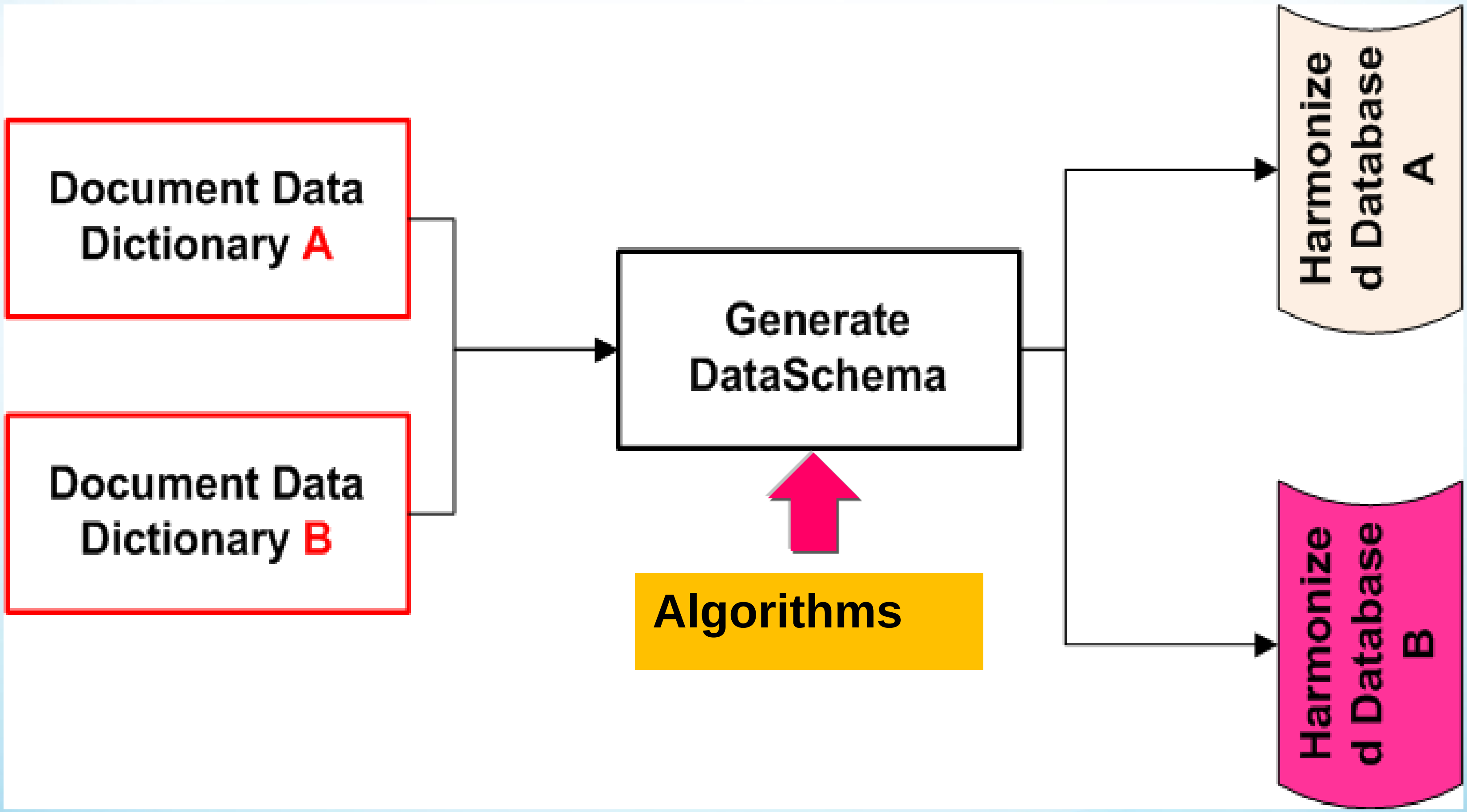
| table | name | valueType | unit | label:en | description:en |
|---|---|---|---|---|---|
| HRB_S | HRB_S1 | Integer | | mother's smoking status | 您现在吸纸烟吗?(吸烟:一生中至少吸过100支香烟 (约5包)) |
| HRB_S | HRB_S1_2 | Integer | | amount of cigarettes per day | 您目前每天平均吸多少支烟? |
| HRB_S | HRB_S4 | Integer | | family members' smoking status | 与您同住的其他家庭成员是否有人吸烟? |
| HRB_S | HRB_S5 | Integer | | colleagues' smoking status | 与您同一办公室的同事在上班时抽烟吗? |
| HRB_S | HRB_S5_1 | integer | | amount of colleagues smoking | 共有几个同事吸烟? |

CA

| table | name | valueType | unit | label:en | description:en |
|---|---|---|---|---|---|
| SMOKINGHISTORY | SH2 | Integer | | Current smoker | At the present time, do you smoke cigarettes daily, occasionally or not at all? |
| SECHANDSMOKE | SHS3 | Integer | | number of cigarettes are smoked indoorly in your home | On a typical day, how many cigarettes are smoked inside your home? |
| SECHANDSMOKE | SHS1 | Integer | | presence of smokers inside home | Including both household members and regular visitors, does anyone smoke inside your home, every day or almost every day? Note : Include cigarettes, cigars and pipes. |
| SECHANDSMOKE | SHS5 | Integer | | exposition to second-hand smoke at workplace | During your pregnancy, has anyone in your workplace smoked in your presence? (including breaks, lunch) |

# Harmonize Variables and Dataset





# Harmonization Potentials



**VARIABLE: Current quantity of cigarettes consumed**
***Definition:*** **Average number of cigarettes consumed by the participant per day; *Unit:* cigarettes per day; *Format:* open; *Type:* integer**

**Completely Match:** In a typical week, how many cigarettes do you smoke per day? (integer)

**Possible:** In a typical week, how many cigarettes do you smoke per day? (1–3, 4–6, 7–9, 10 or more)

**Impossible:** In a typical week over the past 3 years, how often have you been exposed to secondhand smoke inside your home? (little, few, some, many)

# Harmonize Variable: Algorithms



**Dataschema variable:** BMI in kg/m$^2$
**Description:** Body Mass Index calculated using measured weight and height (Mass in Kg / (Height in M)$^2$) . **Value type:** Integer

**Study 1:** BMI collected

**JavaScript algorithms:**
$('BMI').whenNull(99);

**Study 2:** BMI not collected, only height and weight collected

**JavaScript algorithms:**

```
var height = $('MOH');
  var weight = $('MOW');
  if ((height.isNull().or(weight.isNull())).value()) {
    return newValue(99, 'integer');
  } else {
    return weight.div(height.unit('cm').toUnit('m').
pow(2));
  }
}
```

**Variable names legend:**
'MOH'=participant's height at visit
'MOW'=participant's weight at visit

**Harmonized variable:** BMI in kg/m$^2$

# Study and Variable Catalogues



From Maelstrom

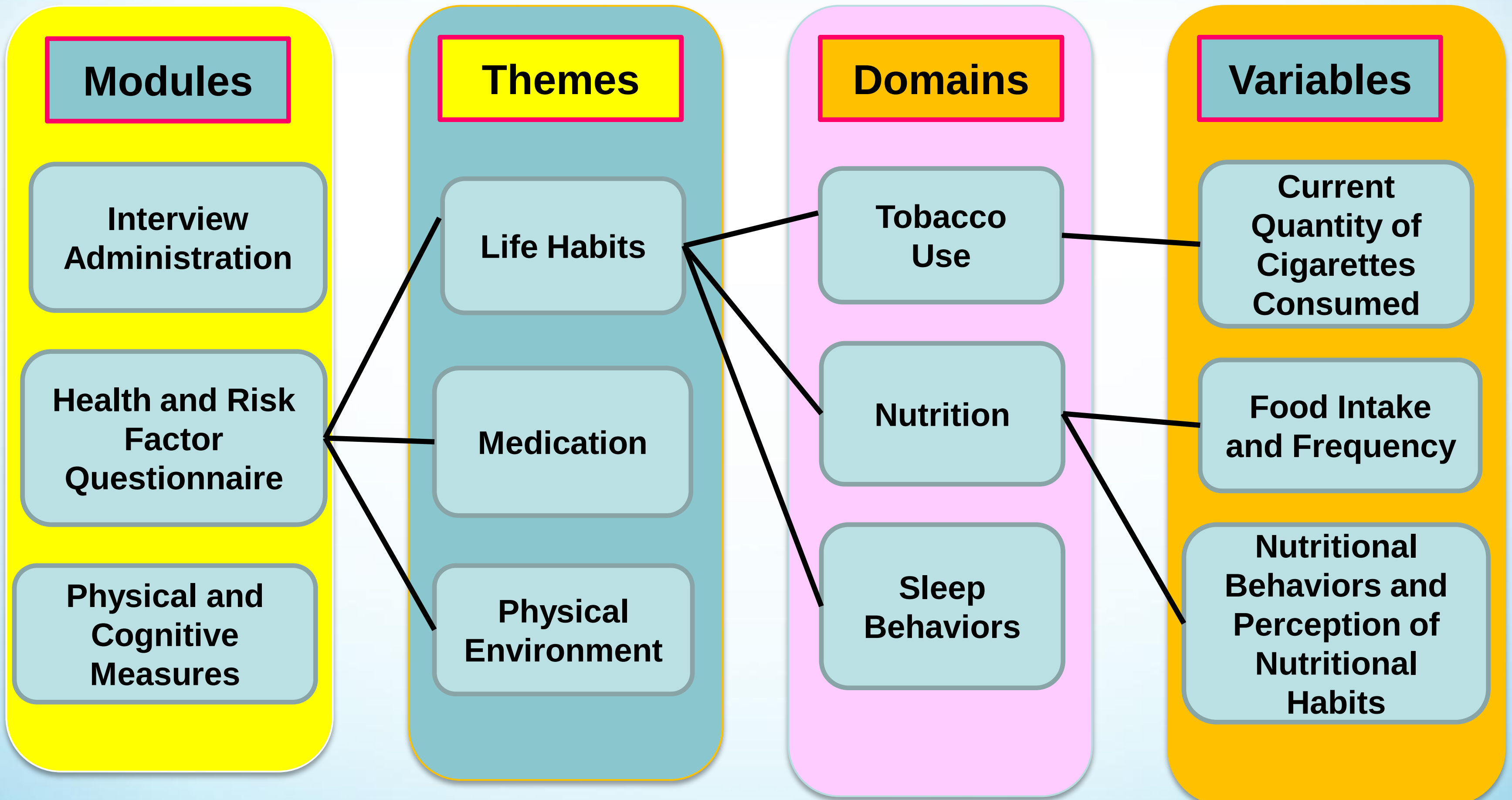
- DATA COLLECTION EVENTS
- DATA SOURCE
- DISEASES HISTORY AND RELATED HEALTH PROBLEMS
  - Diseases of the circulatory system (33)
  - Diseases of the musculoskeletal system and connective tissue (12)
  - Endocrine, nutritional and metabolic diseases (11)
  - Neoplasms (6)
  - Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (6)
  - Show more
- MEDICAL HEALTH INTERVENTIONS/HEALTH SERVICES UTILIZATION
  - Physical/Physiological therapeutic interventions (9)
  - Health services utilization (5)
  - Cognitive, psychosocial and sensory therapeutic interventions (2)
  - Other healthcare interventions (1)
- MEDICATION
  - Medication Intake (30)
  - Adherence to medication (2)
- REPRODUCTIVE HEALTH AND HISTORY
  - Menstruation (9)
  - Menopause (7)
  - Pregnancy (6)
  - Fertility and reproductive problems (2)
  - Number of children (1)
- PARTICIPANT'S EARLY LIFE/CHILDHOOD
  - Newborn/Childhood health (1)
- LIFE HABITS/BEHAVIOURS
  - Nutrition (32)
  - Tobacco use (19)

---

- DISEASES HISTORY AND RELATED HEALTH PROBLEMS
- MEDICAL HEALTH INTERVENTION/HEALTH SERVICES UTILIZATION
- MEDICATION
- REPRODUCTIVE HEALTH AND HISTORY
- PARTICIPANT'S EARLY LIFE/CHILDHOOD
- LIFE HABITS/BEHAVIOURS
- SOCIODEMOGRAPHIC/SOCIOECONOMIC CHARACTERISTICS
- PHYSICAL ENVIRONMENT
- SOCIAL ENVIRONMENT
- PERCEPTION OF HEALTH/QUALITY OF LIFE
- ANTHROPOMETRIC STRUCTURES
- BODY STRUCTURES
- BODY FUNCTIONS
- LABORATORY MEASURES
- ADMINISTRATIVE INFORMATION

# Hierarchy of Data Dictionary





# Variable Classification Taxonomy



**Diseases History And Related Health Problems**
**Medical Health Interventions/Health Services Utilization**
**Medication**
**Reproductive Health And History**
**Participant's Early Life/Childhood**
**Life Habits/Behaviours**
**Socio-Demographic/Socio-Economic**

**Physical Environment**
**Social Environment**
**Perception Of Health/Quality Of Life**
**Anthropometric Structures**
**Body Structures**
**Body Functions**
**Laboratory Measures**
**Administrative Information**

# Basic Harmonization Steps



Define Scientific Questions

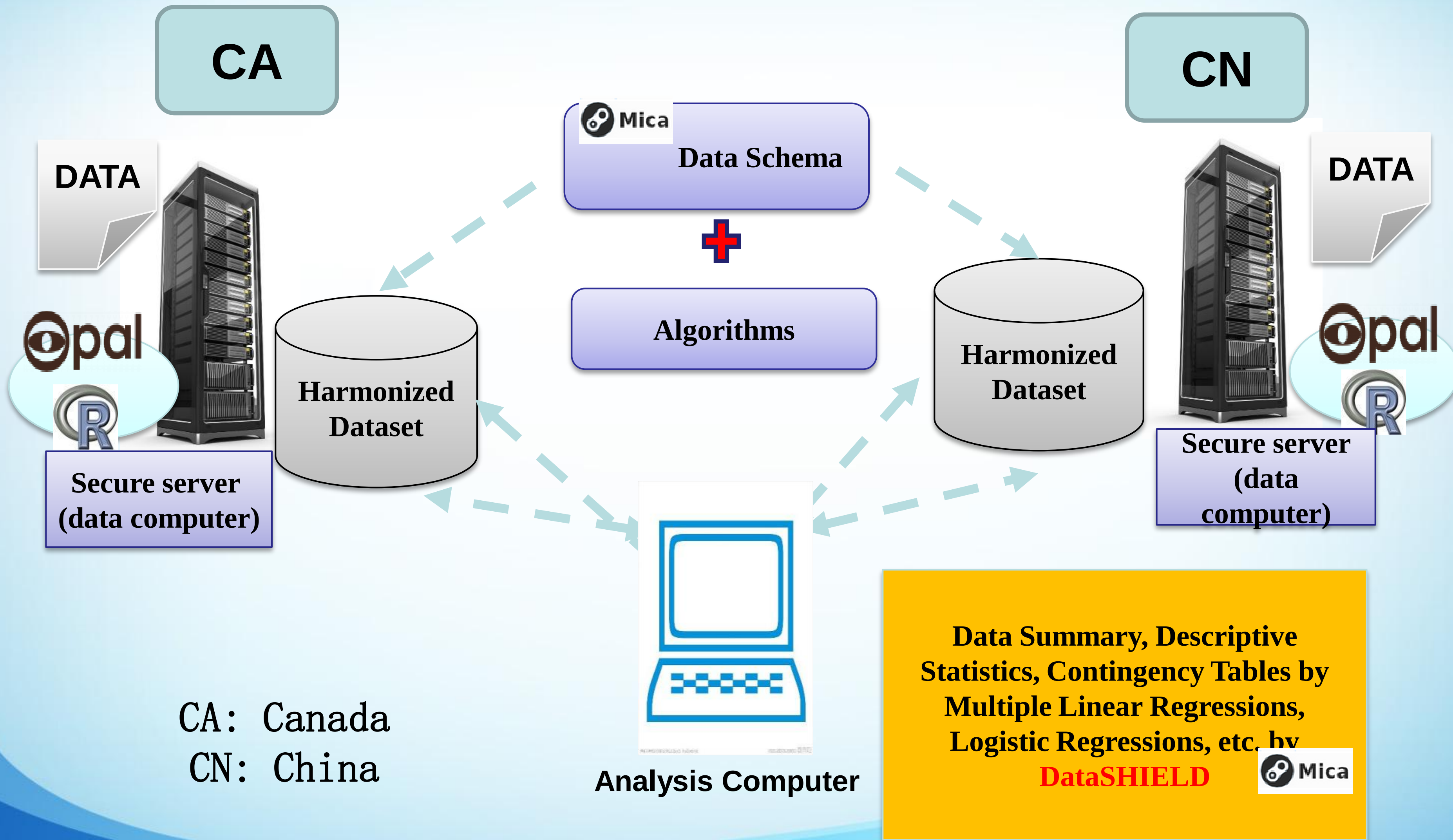
Evaluate Potential for Harmonization

Process Data in a Common Format

Analyze to use the harmonized data

# Federated Analysis Model





# Summary



1. To investigate gene-environment interactions, other less common events, requires larger number and statistical power;
2. Understand similarities and differences across Studies, which direct discovery-driven research;
3. Ultimately, to prompt harmonization and sharing for collaborative endeavor is the key to maximizing the value of limited resources, which could be of value beyond measure.

上海交通大学医学院附属新華醫院
Xinhua Hospital Affiliated To Shanghai Jiaotong University School Of Medicine

# THANK YOU！


www.xinhuamed.com.cn