

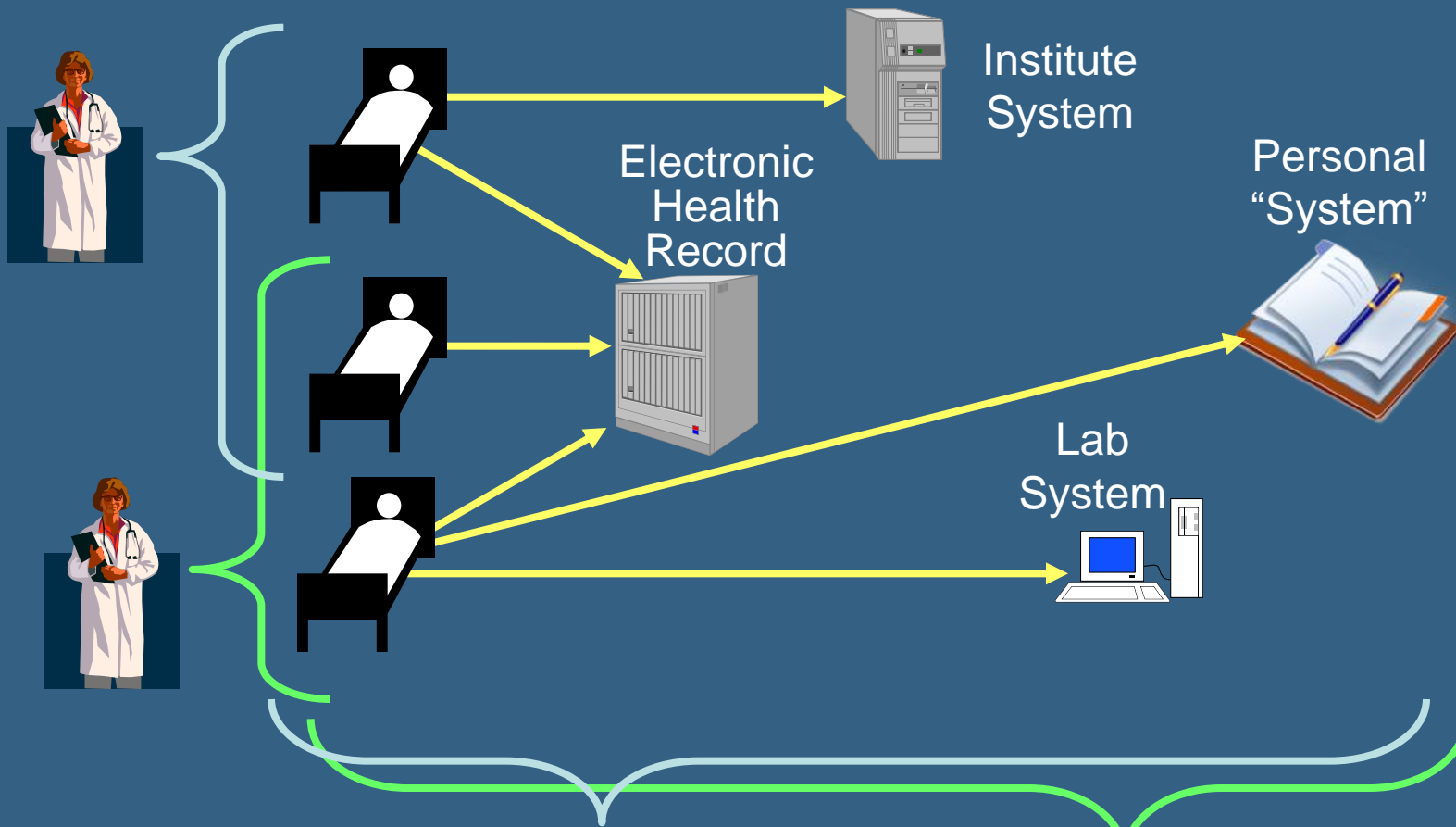
Experiences with Cohort Lists for Re-Identification and Recruitment Across Networks

James J. Cimino, MD, FACMI, FACP
*Laboratory for Informatics Development
NIH Clinical Center
Bethesda, Maryland, USA*

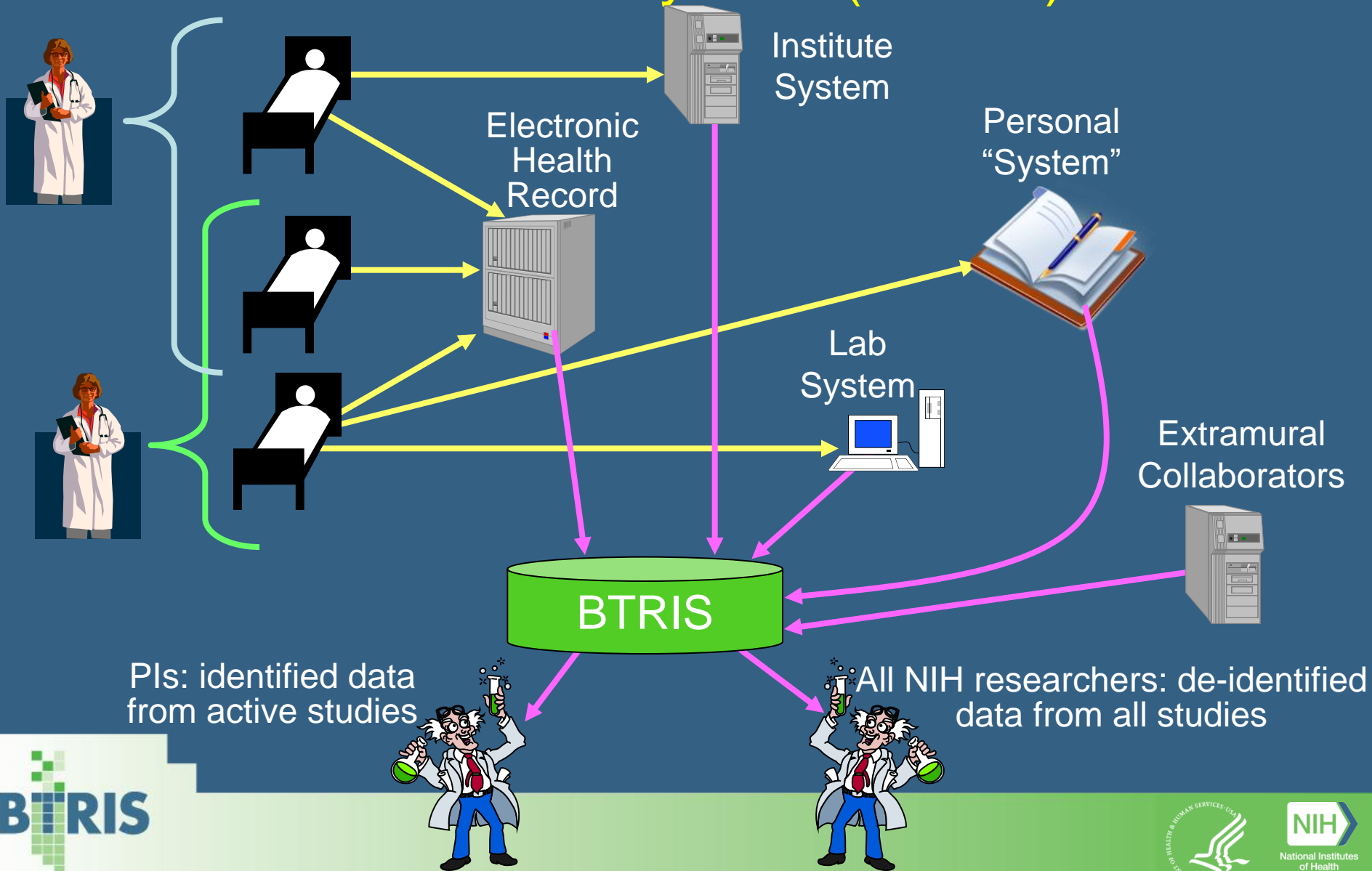
Experience with Cohort Identification

- Background: BTRIS at NIH
- Challenges to access
- Challenges to use

Clinical Research Data at NIH



The Biomedical Translational Research Information System (BTRIS)





BIOMEDICAL TRANSLATIONAL RESEARCH INFORMATION SYSTEM

National Institutes of Health

Home > Subject Attribution & Designee Assignment > De-Identified Data

Select Data Type

- Diagnosis
- Demographics
- Medications
- Lab Tests

- [Training & User Guides](#)
- [Data Source](#)
- [BTRIS Policies](#)
- [Searching in BTRIS](#)

BTRIS Customer Support
[Submit Comment](#)
 M-F 8:30am to 5pm
 301-827-8270
BTRISsupport@nih.gov

System Hours
 M-F 8:30am to 5 pm
 Maintenance and upgrades are
 scheduled for after hours and
 may impact system access and
 performance

Updated: Feb 21, 2013

Result Totals

of Diagnosis1: 14
 # of Diagnosis2: 9 # of Distinct Subjects: 3
 # of Labs4: 5372 # of Medication Orders3: 196

[Download Detailed Result](#)

Query

[Close Query](#)
[How to Query](#)

Review & Run Query

Diagnosis1

[Diagnosis Help](#)

Search for Diagnosis: **Include**

Contains all of these words

Remove All	Selections
<input checked="" type="checkbox"/>	Chronic Granulomatous Disease (CGD)

Enter Description: **Include**

Diagnosis Date: **Between** and

Age at Diagnosis: **Between** and

AND

Diagnosis2

[Diagnosis Help](#)

Search for Diagnosis: **Include**

Contains all of these words

Remove All	Selections
<input checked="" type="checkbox"/>	Unspecified Pulmonary Tuberculosis, Unspecified Examination

Enter Description: **Include**

Diagnosis Date: **Between** and

Challenges to Access

- Honest broker services still needed (may need clinical expertise)
- Original researchers have intellectual interest in data
- Disjoint policies on data re-use versus new collection (specimen reanalysis, natural history, intervention)

Challenges to Use

- Lack of patient identifiers across institutions (or across data sets within institutions)
- EHR data fields not used consistently (PII, ICD-9-CM)

Looking for Chronic Granulomatous Disease

ICD	ICD Name	Text Description
288.1	Functional Disorders of Polymorphonuclear Neutrophils	CHRONIC GRANULOMATOUS DISEASE
288.1	Functional Disorders of Polymorphonuclear Neutrophils	AUTOSOMAL RECESSIVE CHRONIC GRANULOMATOUS DISEASE
288.1	Functional Disorders of Polymorphonuclear Neutrophils	RULE OUT CHRONIC GRANULOMATOUS DISEASE
288.1	Functional Disorders of Polymorphonuclear Neutrophils	CARRIER CHRONIC GRANULOMATOUS DISEASE
288.1	Functional Disorders of Polymorphonuclear Neutrophils	MOTHER OF AFFECTED CHRONIC GRANULOMATOUS DISEASE PATIENT, RULE OUT CARRIER STATE
288.1	Functional Disorders of Polymorphonuclear Neutrophils	NOT A CARRIER OF X-LINKED CHRONIC GRANULOMATOUS DISEASE

Challenges to Use

- Lack of patient identifiers across institutions (or across data sets within institutions)
- EHR data fields not used consistently (PII, ICD-9-CM)
- Terminologies not used consistently across networks
 - Laboratory tests

Sharing Laboratory Data

Baorto DM, Cimino JJ, Parvin CA, Kahn MG.
Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC). *Int J Med Inform.* 1998 Jul;51(1):29-37.

[Examined LOINC coding of local lab terms]

...In the present study, 75% of failures to match the same tests between different institutions using LOINC codes were due to differences in local coding choices...

Challenges to Use

- Lack of patient identifiers across institutions (or across data sets within institutions)
- EHR data fields not used consistently (PII, ICD-9-CM)
- Terminologies not used consistently across networks
 - Laboratory tests
 - Medication terms

Sharing Medication Data

Cimino JJ, McNamara TJ, Meredith T, Broverman CA, Eckert KC, Moore M, Tyree DJ. Evaluation of a proposed method for representing drug terminology. Proc AMIA Symp. 1999:47-51.

[Examined the drug descriptions matched on NDC codes across five commercial pharmacy knowledge bases]

...We found that the components of the term descriptions matched 68-87% of the time and that the overall descriptions matched 53% of the time....

Challenges to Use

- Lack of patient identifiers across institutions (or across data sets within institutions)
- EHR data fields not used consistently (PII, ICD-9-CM)
- Terminologies not used consistently across networks
 - Laboratory tests
 - Medication terms
- Users need to understand data model:
 - Medication lists versus orders versus administration
 - ICD-CM-9 codes versus narrative text field
- Data standards may not meet all needs
 - CDISC Study Data Tabulation Model (SDTM)
 - demographics: marital status, name, death date, address
 - lab data: lab comment, data source