In order to mine data...

P. Pearl O'Rourke, MD
Partners HealthCare
Boston, MA

In order to mine data...

You need a Mine

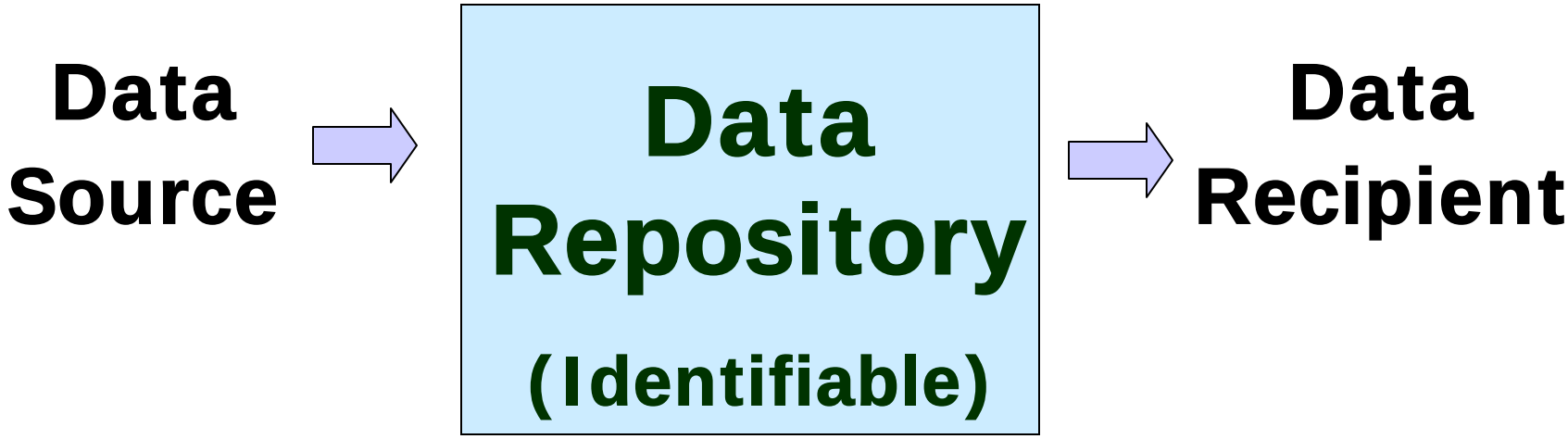
P. Pearl O'Rourke, MD

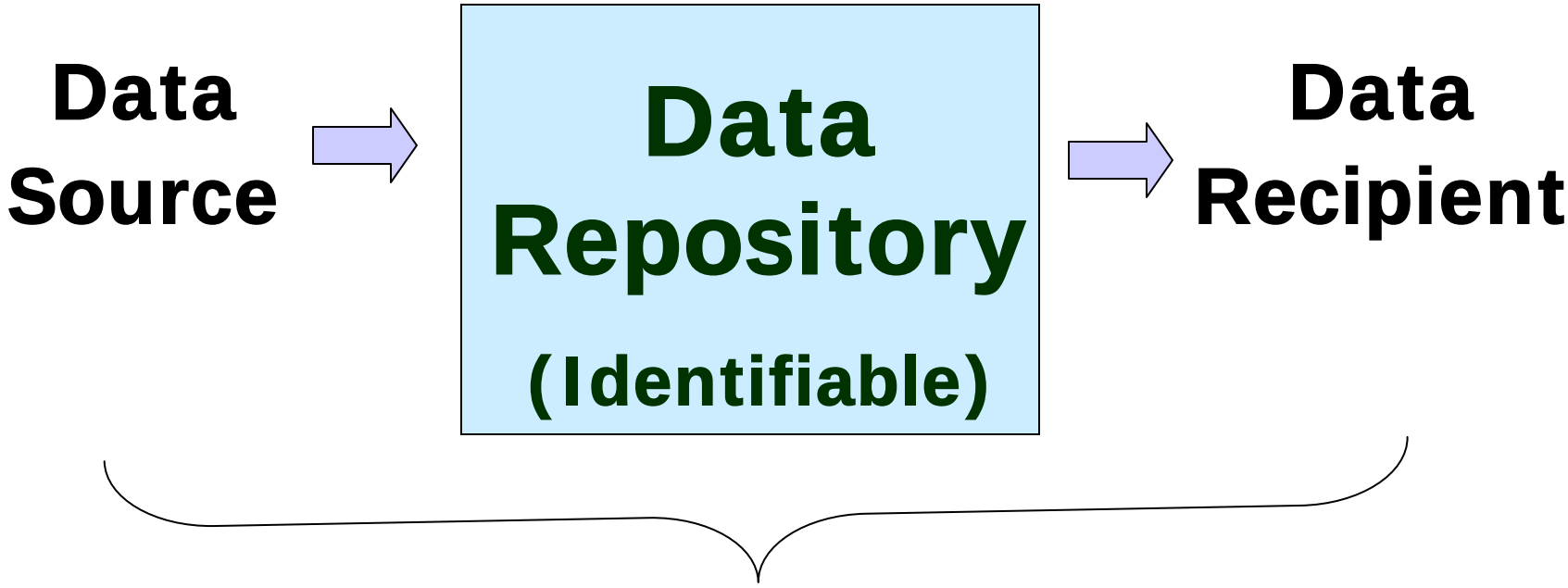
Partners HealthCare

Boston, MA

Assumptions

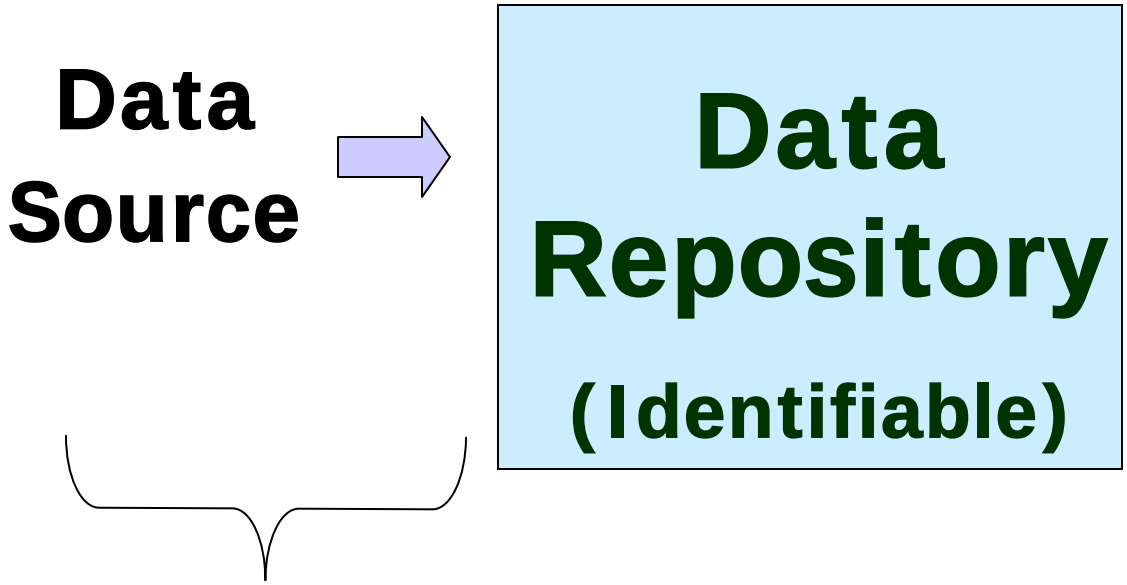
- Current science requires access to large data bases (mines)
- Most valuable databases contain identifiable longitudinal data
- Databases are only as good as the quality of their query tools
- Databases require rules of operation for:
  - Data in
  - Maintenance of data in the repository
  - Data out


Data Source
Data Repository
(Identifiable)
Data Recipient



IRB review of and approval for

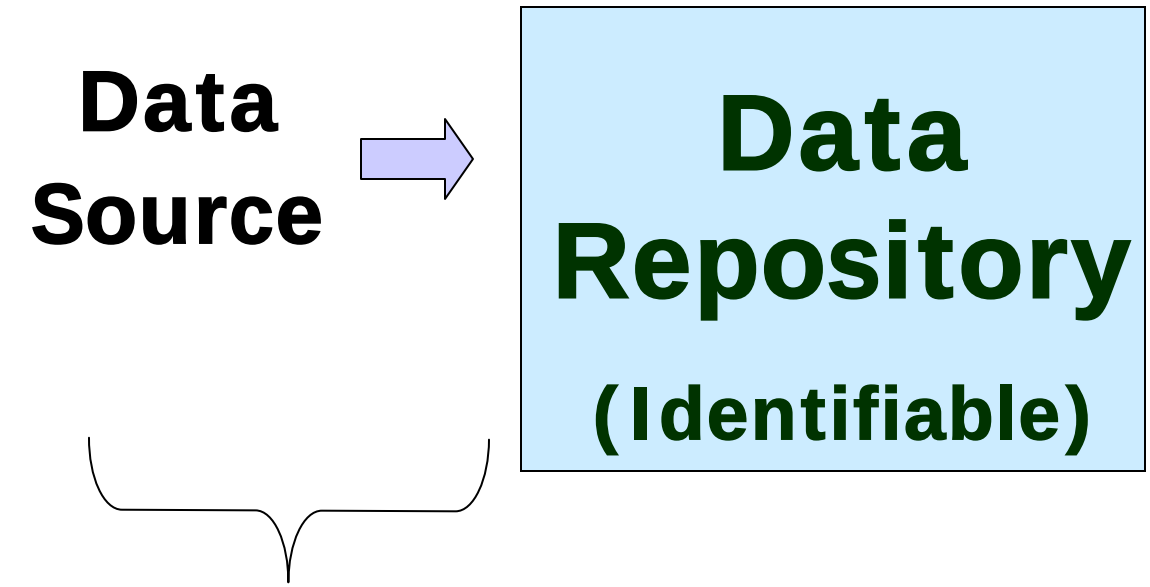
- The repository itself
- Operating rules for inputting data
- Operating rules for providing data



Inputting data generated during

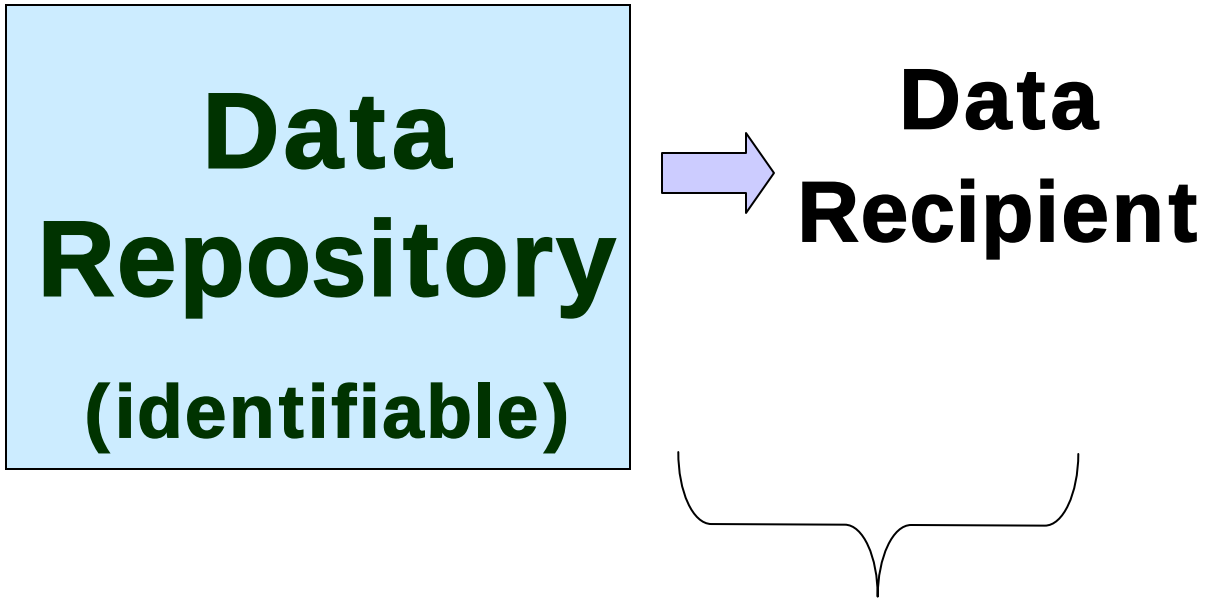
ROUTINE CLINICAL CARE:

- Informed consent/authorization or
- Waiver
  - *makes the most sense for large clinical databases*



Inputting RESEARCH data requires:

- IRB approval for placement of data into the Repository
  - As part of initial protocol submission
    - Include in any informed consent/authorization
  - As an amendment
    - New consent /authorization may be required



Providing data from the Data Repository:

- Cumulative numbers and de-identified data:
  - Covered by Repository rules alone
- Any identifiable data
  - Requires IRB review and approval
    - IRB to consider need for informed consent/authorization

Waivers

Informed Consent and Authorization

- Benefits
  - More practical logistics for large data bases
  - Avoid sample bias
- Challenges
  - Disclosures must be tracked (HIPAA)
  - May limit secondary uses: I.e., placement into databases that require consent
  - Limited ability to re-contact individuals

Summary

- Robust clinical data Repositories can be built
- The rules of operation must be carefully developed
- The addition of research data to any Repository adds complexity
  - Research data and Clinical data have distinct characteristics and regulations
- The Repository ‘owners’ must have a working relationship with Institutional Oversight

Research Patient Data Registry exists at Partners Healthcare to allow for exploration of clinical data

1) Queries for aggregate patient numbers

- Warehouse of in & outpatient clinical data
- 3.6 million Partners Healthcare patients
- 1.1 billion diagnoses, medications, procedures, laboratories, physical findings & genomics coupled to demographic/visit data
- Authorized use by faculty status
- Clinicians can construct complex queries
- Queries cannot identify individuals, internally can produce identifiers for (2)

Query construction in web tool

De-identified Data Warehouse

Z731984X
Z74902XX
...
..

Encrypted identifiers

2) Returns identified patient data

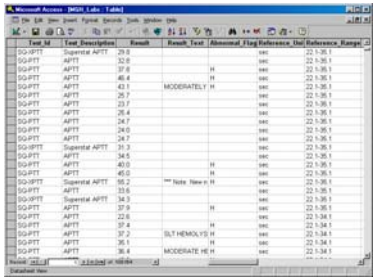
- Start with list of specific patients, usually from (1)
- Authorized use by IRB Protocol
- Returns contact and PCP information, demographics, providers, visits, diagnoses, medications, procedures, laboratories, microbiology, reports (discharge, LMR, operative, radiology, pathology, cardiology, pulmonary, endoscopy), and images into a Microsoft Access database and text files.

0000004
2185793
...
...

OR

0000004
2185793
...
...

Real identifiers



Security and Patient Confidentiality of Step 1

- All patients at Partners are added
  - HIPAA notification that their data may be used for research upon registration.
- RPDR data is anonymized at the Query Tool.
  - Aggregated numbers are obfuscated to prevent identification of individuals; automatic lock out occurs if pattern suggests identification of an individual is being attempted.
- Queries done in Query Tool available for review by RPDR team, a user lock out will specifically direct a review.
- Medical record numbers are encrypted and obvious identifiers are removed from the data warehouse.
- Concept of “established medical investigator” is promoted by classification as a faculty sponsor.

Security and Patient Confidentiality of Step 2

- Only studies approved by the Institutional Review Board (IRB) are allowed to receive identified data.
- Queries may be set up by workgroup member, but faculty sponsor on IRB protocol must directly approve all queries that return identified data.
- Special controls exist when distributing data regarding HIV antibody and antigen test results, substance abuse rehab programs, and genetic data, due to specific state and federal laws.
- Queries that return identified data are reviewed (retrospectively) by the IRB.