Big Data at Penn Medicine

Patient Care and Research

Brian P. Wells
Associate VP Health Technology and Academic Computing
brian.wells@uphs.upenn.edu
wellsb@mail.med.upenn.edu

Penn Medicine

Penn Medicine at a glance

PENN Medicine 3.8 billion operating budget

UNIVERSITY OF PENNSYLVANIA HEALTH SYSTEM

University of Pennsylvania School of Medicine
Founded 1765
720 students, 1006 housestaff

Hospital of the University of Pennsylvania (HUP)
695 beds, 1595 physicians
38,213 adult admissions

Penn Presbyterian Medical Center (PPMC)
317 beds, 832 physicians
15,419 adult admissions

Pennsylvania Hospital (PAH)
515 beds, 888 physicians
23,903 adult admissions

Clinical Practices of the University of Pennsylvania (CPUP)
1200 faculty
200+ practice locations
1.4M office visits/yr

Clinical Care Associates (CCA)
250 providers
55+ practice locations
800K office visits/yr

Sources of Big Data in Penn Medicine

Tissue/liquid samples, genetics, proteomics, clinical trials, PROs…

Patient Research Data
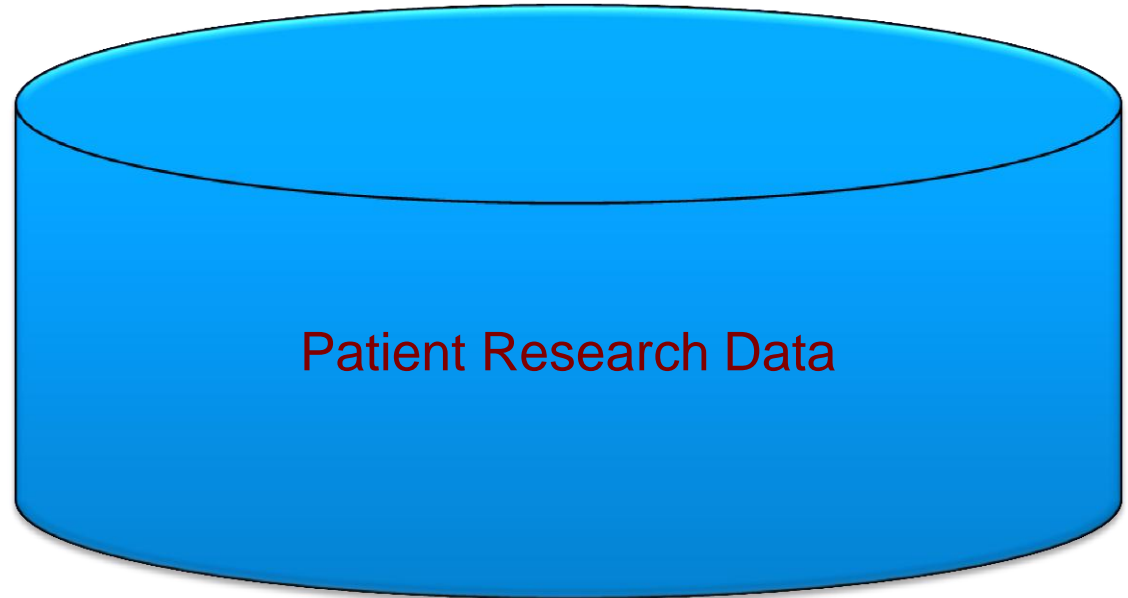
Exam notes, reports (Rad, Path, GI, Derm, Neurology, Cardiology, …), discharge summary

Patient Clinical Unstructured Data
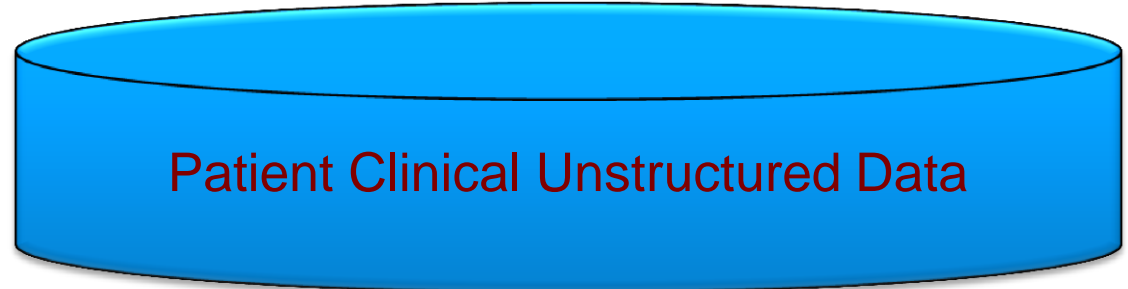
Diagnostic tests, results, medications, infections

Patient Clinical Structured Data

Demographics, billing

Patient Administrative Data

The Opportunity

Combine all the discrete/structured, unstructured and research data into one integrated warehouse of patient information and then utilize it to:

Patient Care

- Improve patient quality and safety through tracking metrics and benchmarks
  - Infections
  - Falls and other incidents
  - Adverse drug events
  - Evidence based medicine
- Develop impactful clinical decision support rules that evaluate at the point of care within the EMR(s)
- Simplify medication reconciliation
- Improve financial performance

Research

- Identify cohorts of patients for clinical trials
- Complete Genome Wide Association Studies (GWAS)
- Identify unique predictive biomarkers for specific diseases
- Discover, develop and test molecules that modify genetic and proteomic molecular processes
- Complete observational studies to identify ways to improve care

The Challenges

- ◆ **Source systems and data integration**
  - UPHS utilizes over 50 systems to run the clinical and billing aspects of the health system alone
  - HL7 is only the ***structural*** messaging standard
  - PSOM has many islands of research data
- ◆ **Standards**
  - Healthcare in general does not utilize ***semantic*** data standards beyond those required for billing
    - ICD-9
    - CPT-4
  - Clinical and research data standards are poorly adopted
  - Therefore clinical and research data exchange between disparate systems is nearly impossible
  - The USA do not have a national patient identifier!

"Standards are like toothbrushes – everyone has one but nobody wants to use yours"
Doug Fridsma, Director Office of Standards and Interoperability, ONCHIT

Penn Medicine

Translational Research Data Map

Allscripts: Sunrise Clinical Manager: Inpatient electronic medical record

AHRQ:

Aria: Radiation oncology

Cerner: Cerner Millennium Laboratory and pathology diagnostics system – remotely hosted by Cerner Corp in Kansas City.

CV Surg STS:

UHC:

Clintrac: Workflow management for registration and financial planning and billing.

CBIG

SBIA

GEPACS/ Imagecast: Radiology information system.

Press Ganey:

Radiology Research: placeholder for Radiology, including servers moved from Simbfire 3495

HDM: Coding (diagnosis/procedures) all inpatients and certain Outpatients at HUP and PPMC. Coding data is stored in HDM and is also interfaced out to UMS for billing.

Theradoc: infection prevention and management system real time surveillance tools monitor changes in patient condition, adverse events, and threats to patient safety.

EMTRAC: Emergency Department

Cardio

Derm

GI

HemOnc

Pulmonary

Lawson

OBGYN

Knowledge Link

MedView: Physicians portal

Report

Siemens : Inpatient ADT, billing, scheduling and census.

Epic:
Beaker: Epic's medical oncology product.
Cadence: Epic's Enterprise Scheduling product. Used to schedule and track patient appointments.
Canto: Enables UPHS physicians to access patient data on an iPad.
Care Everywhere: Provides access at the point of care to the patient's medical records from other organizations.
Clarity: Epic's data warehouse
EpicCare Ambulatory: Outpatient medical record application.
EpicLink: An application that allows affiliate organization providers to view a patient's clinical data in our Epic system.
Haiku: Enables UPHS physicians to access patient data on a iPhone.
Identity: EMPI for the application and the health system.
MyChart: MyPennMedicine - allows patient to view their medical records and interact with their physicians.
Prelude: Epic's registration application.
Resolute: Epic's Patient Accounting product for clinics.

OTTR:

FPDS: Financial Penn Data Store HPMI: Data warehouse
Used by financial staff, administration, and researchers for detailed operational and financial and analyses including cost, revenue and margin across UPHS entities

CPDS: Clinical Penn Data Store Penn Medicine data warehouse

UPDS: Unstructured Penn Data Store

I2B2: data warehouse Used by HRI with select data from Penn Data Store.

IDS Clinical (ITMAT Protocol)

IDS

WebIDS

Med Dispense: dispense emergency meds and studies using controlled substances

Cell Center Stockroom: Order, inventory and billing for products sold.

Translational Core Lab: Provides specialized lab services on human and animal samples to researchers

TRC Lab Services

Microscopy Core: Service providing microscopes and application supporting scheduling and billing for use

Proteomics: application supporting ordering and billing of services provided by proteomics core

SAM: enables Principal Investigator, designated Lab Manager or Business Administrator to authorize an individual to place orders to service centers/cores against a specific grant fund.

SOMERA: SOM Electronic Research Administration, pre-dates PennERA, still used to manage school's grant proposal approval process.

DNA Sequencing: Ordering, data delivery, billing for DNA services to research projects

Transgenic Mouse: Ordering, inventory management and billing of services provided by Transgenic and Chimeric Mouse Facility

Pathcores: Applications supporting the ordering of services and billing for multiple Pathology-based cores.

PennERA: Data warehouse collection for querying electronic grant submission data to federal funding agencies.

NIH Grants.gov

Clinical Trial Auditing: Application supporting internal audit of clinical trials

IRB Data Feed (existing): Interfaced to SOMS and used in multiple applications requiring IRB data.

Clinical Research Registry

Research Subject Interest: Biographical information for participation research at Penn, limited to a specific protocol.

DCI: Departments, Centers and Institutes; assist in map an organization in DCI to zero in the financial org code

RBA: Research Billing Application - generates a Research Billing Number (RBN)

Research Billing Number: Account Number

Research Billing System

Tumor Registry

Peers

Pharma

Research Subject Registry: data about subject enrolled in research studies at Penn

DSMC CTSRMC: Data Safety Monitoring Committee, Cancer Center Clinical Trials Scientific Review Committee

Velos ACC

Oracle Clinical

CHOP

Brovo

FEDS: Faculty Expertise Database System, stores expertise data used for CV, faculty profiles.

DMU: Psychiatry Data Management Unit; Treatment Research Center

Hyperbaric Medicine: Institute for Environmental Medicine

IMACS: Siperian Contract management Calculates contracted payer amount

SAE: Serious Adverse Events

Study Protocol Viewer

Velos Training

PRA Viewer: Prospective Reimbursement Analysis

ACC Membership

Way to Health

Velos SOM

Progeny

User core log

Path BioResources Billing

ca tissue

ACCARD

VIVO: Collaboration and data visualization of researcher networks based on publication, grant and other researcher profile data. Key deliverable of CTSA grant

Stem Cell & Cadaver Inventory: application to manage requests for and inventory cadavers, human body parts and stem cells.

PGFI: Penn Genomics Frontiers Institute, a multi-school research institute (Type 1, accountable to Provost)

FADS: Faculty Affairs Database System, stores data on faculty appointments and positions.

POLARIS: (ULAR) Application for ordering, management and billing of animals for research.

PCBI: a unit within PGFI

DB Freezer: Store and track simple sample data Freezer/Shelf/Rack storage systems

Freezer Works: Freezer Inventory and Sample Tracking

Cost Finder: Delivers research cost for procedures at Penn Medicine's three hospitals.

LIMS: Laboratory Information Management System

WSL

IMPACT

Grants List

PFS: Patient Family Services

ICU Registry

SPS

RedCAP

Trusted Broker Service

RPDS: Research Penn Data Store

Clinical Trial X: Trial listing service

Clinicaltrials. gov

NSF

NIH CTSA

HHS

Clinical Trial Listing ITMAT

External Collaboration Sites

NIH Grants.gov

CRCU: Clinical Research Computing Unit; developed and supports custom applications to support large clinical trials, other functions.

CMP HSR: Center for Mental Health Policy and Services Research, receives HHS data.

Version 29

Legend:
- Existing SOM
- To be created
- External sources
- Existing Penn Medicine
- To be included in system
- Data Warehouse
- System
- Feeds Data Warehouse

Brian P. Wells
Associate VP Health Technology and Academic Computing
brian.wells@uphs.upenn.edu

Created by Vince Frangiosa
vince.frangiosa@uphs.upenn.ed

CONFIDENTIAL

The Challenges

- **Volume increasing**
  - UPHS
    - 4.5 million patients
    - 42 million encounters
      - 2+ million added each year
    - 400 million orders and results
    - 40 million system-to-system messages a month across 350 unique interfaces
  - PSOM
    - 2 million samples
    - Genetic data exploding
      - Half a terabyte of data per full patient genome sequence
      - Rapidly increasing sequencing speed, accuracy and fidelity with decreasing duration and cost
- **Network speeds not following Moore's law**
  - 10 gigabit per second max between buildings
  - 1 gigabit per second at the wall plate
  - 3 hours per terabyte on a *dedicated* connection

The Challenges

- **Security tightening**
  - HIPAA
    - Breach notification
  - HITECH
    - Personal liability
    - Fines, imprisonment
  - GINA
    - Needs strengthening
  - Full de-identification of unstructured data requires manual review
- **Science**
  - Microbiome sequencing
  - New genetic biomarkers constantly being discovered
- **Liability / Ethics**
  - If we "know" your entire genetic profile must we notify you when a new marker is discovered that you already possess?
  - Must we notify your offspring? Parents? Can you sue us if we fail to?

What is Penn Medicine Actually Doing?

- **Penn Data Store**
  - Financial
  - Clinical
  - 2 to 3 billion rows of structured information
  - Dashboards and reports
    - Financial
    - Clinical quality
    - Patient satisfaction
    - Research requests
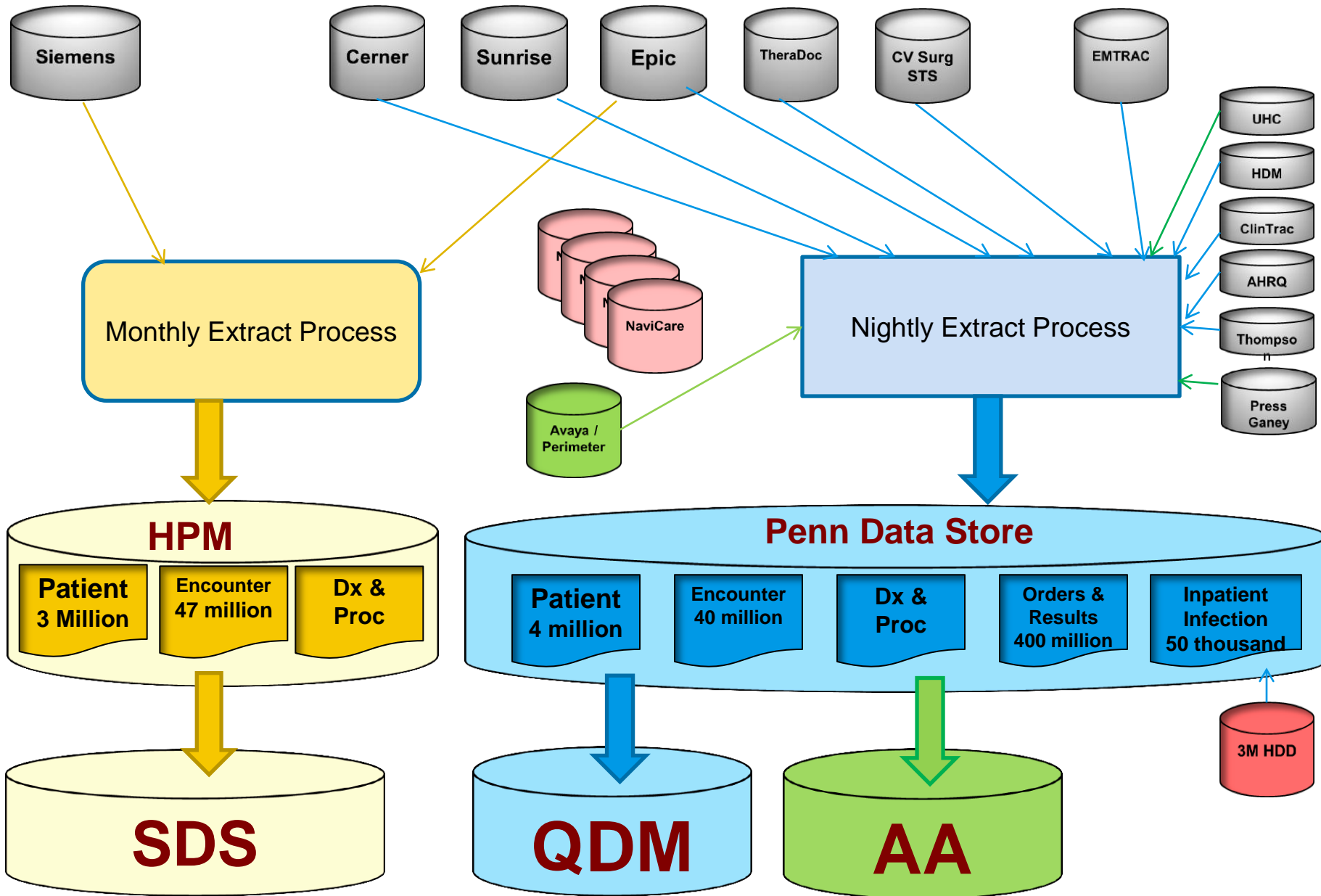
What is Penn Medicine Actually Doing?

- **Penn Data Store**
  - Financial
  - Clinical
  - 2 to 3 billion rows of structured information
- **Work in progress**
  - High Performance Computing
    - New local, $3 million cluster
    - 1024 high memory cores
    - 1 petabyte of spinning disk
    - 3 petabytes of tape archive
  - Best Practice Advisories
    - Clinical trial recruiting
    - Clinical care alerts
  - Predictive analytics
    - Sepsis "sniffer"
  - Unstructured text mining
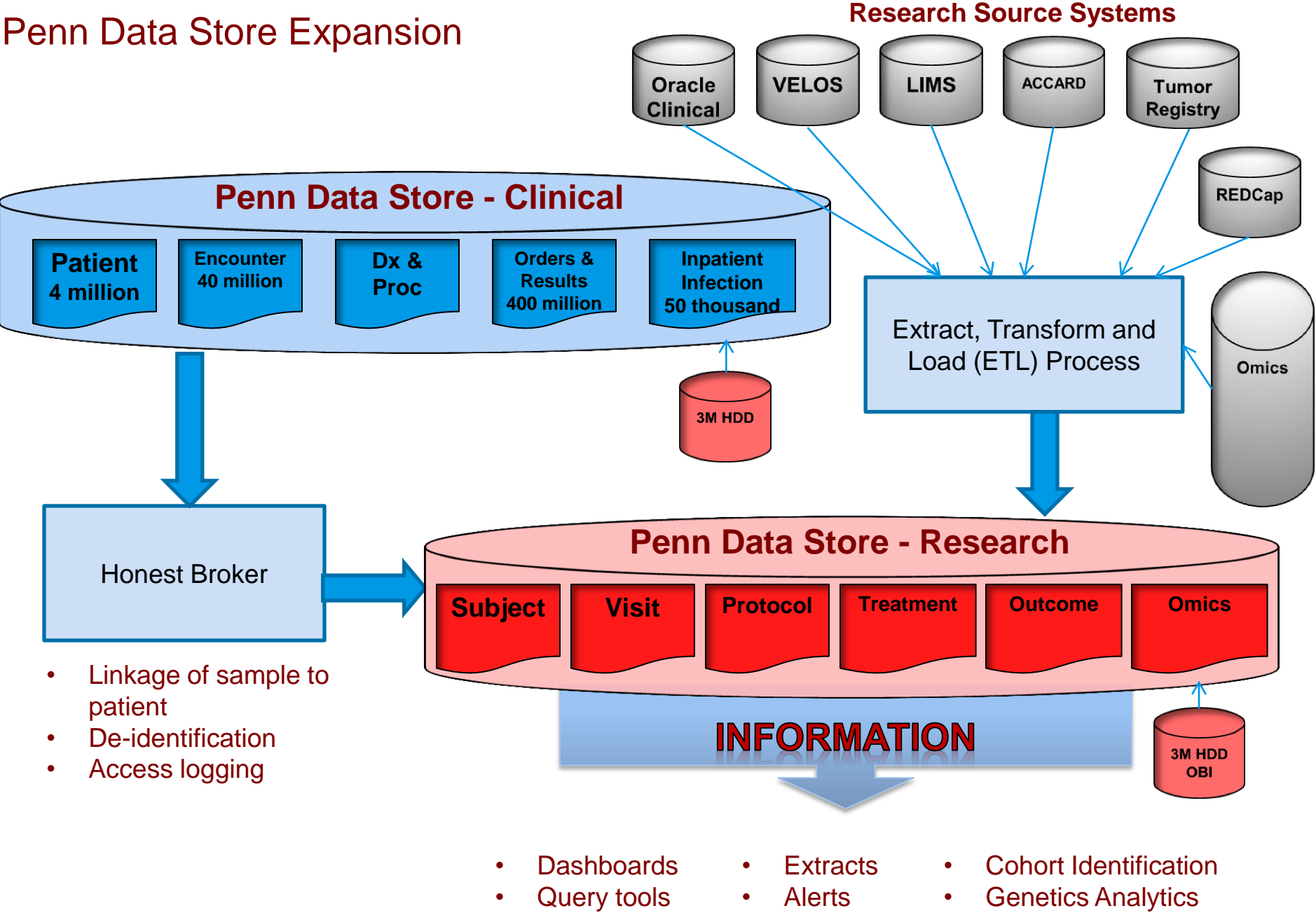   - 15 million documents and counting
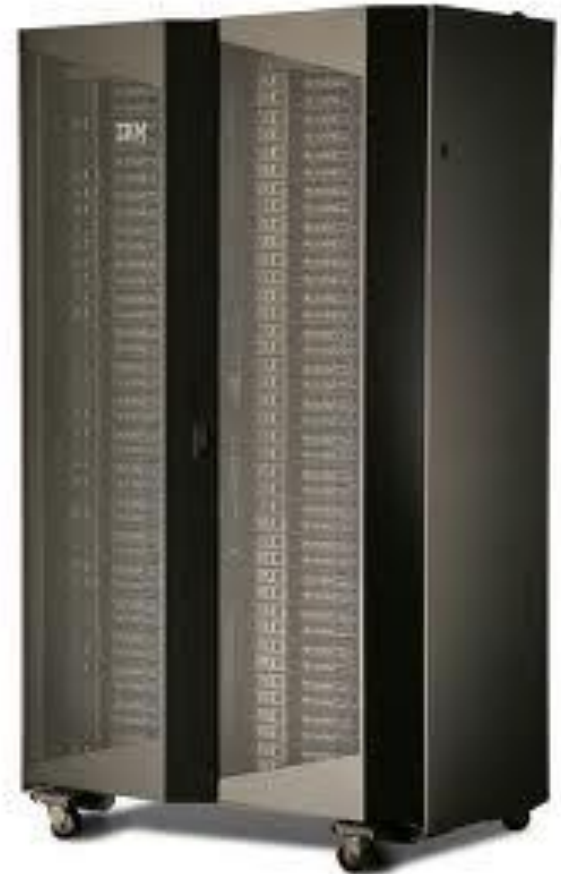
What is Penn Medicine Actually Doing?

◆ **Future projects**

- Bio-bank operational system (LIMS)
- Simulation
  - What happens if …
- Research section of Penn Data Store
  - Genetic data
  - Bio-bank
  - Tumor Registry
  - Outcomes
  - Use cases
    - ○ Find patients that are poor responders for drug Y and have a mutation in the promoter region of Gene X
    - ○ What is the expression level of TP53 mutants by cancer tissue?
    - ○ How many patients have disease Z, responded to treatment, have a chromosome 18 deletion and have blood samples in the bio-bank?
    - ○ Mine the breast and ovarian TCGA data for the somatic mutation data associated with tumors with germline BRCA1 and BRCA2 mutation

Penn Medicine

Penn Data Store Expansion

Research Source Systems

Oracle Clinical | VELOS | LIMS | ACCARD | Tumor Registry

REDCap

Omics

Penn Data Store - Clinical

| Patient 4 million | Encounter 40 million | Dx & Proc | Orders & Results 400 million | Inpatient Infection 50 thousand |

3M HDD

Extract, Transform and Load (ETL) Process

Honest Broker

- Linkage of sample to patient
- De-identification
- Access logging

Penn Data Store - Research

| Subject | Visit | Protocol | Treatment | Outcome | Omics |

3M HDD OBI

INFORMATION

- Dashboards
- Query tools
- Extracts
- Alerts
- Cohort Identification
- Genetics Analytics

Questions?