Enhancing Community Health Center PCORI Engagement (EnCoRE)

Funded by:
This work was partially supported through a Patient-Centered Outcomes Research Institute (PCORI) Program Award (NCHR 1000-30-10-10 EA-0001).

With support from:
N² PBRN – Building a Network of Safety Net PBRNs
funded by:
Grant # 1 P30 HS 021667
Project Partners

Clinical Directors Network (CDN)  
New York, NY  
Jonathan N. Tobin, PhD  
JNTobin@CDNetwork.org

National Association of Community Health Centers (NACHC)  
Washington D.C.  
Michelle Proser, MPP  
MProser@NACHC.org  
Michelle Jester, MA  
MJester@NACHC.org

The Association of Asian Pacific Community Health Organizations (AAPCHO)  
Oakland, CA  
Rosy Chang Weir, PhD  
rceir@aapcho.org

Access Community Health Network  
Chicago, IL  
Danielle Lazar, Danielle.Lazar@accesscommunityhealth.net

Institute for Community Health (ICH) a Harvard Affiliated Institute  
Cambridge, MA  
Shalini, A. Tendulkar, ScM, ScD  
stendulkar@challiance.org  
Leah Zallman  
lzallman@challiance.org

The South Carolina Primary Health Care Association (SCPHCA)  
Columbia, South Carolina  
Vicki Young, PhD  
vickiy@scphca.org
EnCoRE Partners’ Geography
2014-2015
EnCoRE: Enhancing Community Health Center PCORI Engagement

AIM: To build health center capacity to engage in patient-centered outcomes research through an interactive 12-month long training curriculum, walking health centers through the steps and skills needed to develop a patient-centered research proposal
Goal:
To adapt, enhance, and implement an existing year long training curriculum designed to educate and engage Health Center teams including patients, clinical and administrative staff in Patient Centered Outcomes Research (PCOR).

Objectives:
• Build infrastructure to strengthen the patient-centered comparative effectiveness research (CER) capacity of Health Centers as they develop or expand their own research infrastructure
• Develop, implement, and disseminate an innovative online training, which will be targeted to and accessible at no cost to all Health Centers and other primary care practices.
• Content will prepare Health Center patients, staff, and researchers in the conduct of community-led PCOR
During this live training, you may ask questions at any time in the Chat Window. This area is located in the lower left hand corner of your screen. These questions will be answered at the end of the presentation.
Audio Setup

Configure your PC for Audio

Configure Your PC

- Click the Microphone/Gears icon or
- Go to: Tools > Audio > Audio Setup Wizard

1. DEMO ROOM – Blackboard Collaborate
   - Tools > Audio > Audio Setup Wizard

2.Roger Noia
   - Tools > Audio > Audio Setup Wizard...
This program has been reviewed and approved for up to 1.5 Prescribed CME credits by the American Academy of Family Physicians (AAFP).

Please complete the CE Evaluation launched at the end of the presentation and email eLearning@CDNetwork.org with a request for credits.
Presenters

Fred D. Rachman, MD
Chief Executive Officer
Alliance of Chicago Community Health Services

Heidi Crane, MD, MPH
Associate Professor of Medicine
University of Washington;
Associate Director, CFAR Clinical Epidemiology and
Health Services Research
University of Washington and Fred Hutchinson Cancer
Research Center CFAR

Fe Seligman, MA Economics
Director of Grants, Operation Samahan Health Centers
EnCoRE Session 8: Bioinformatics
May 19, 2015

Fred D. Rachman, MD
Chief Executive Officer
Alliance of Chicago Community Health Services

Heidi Crane, MD, MPH
Associate Professor of Medicine
University of Washington;
Associate Director, CFAR Clinical Epidemiology and
Health Services Research
University of Washington and Fred Hutchinson Cancer
Research Center CFAR

Fe Seligman, MA Economics
Director of Grants, Operation Samahan Health Centers
Patient reported outcomes, PROMIS®, and implementation in clinical care for patient centered outcomes research:

Our experience with clinical sites in CNICS

H Crane MD, MPH
University of Washington
hcrane@uw.edu
Outline

- Why collect PROs
- How: assessment platform
- Clinic buy-in
- PRO implementation in care: ever evolving
- Patient and provider views on implementation
- All PRO collection in clinical care is local
- Key stakeholders view domains differently
- Impact on clinical care
- Impact on clinical research
- PROMIS
- Summary
Why do it?

- Patient-reported outcomes (PROs) such as mental health, substance abuse, symptom burden, and medication adherence can:
  - Enhance patient-provider communication
  - Improve care
  - Facilitate clinical research
Figure 1A. Common situation in routine clinical care

- Structural barriers
  - Provider barriers
  - Patient barriers to assessment

- System not aware

- Adherence not assessed

- No intervention

- Poor adherence

- Continued poor adherence

- Poor HIV outcomes

- Patient factors
  - *substance abuse
  - *mental illness
  - *other

Figure 1B. Situation with valid adherence measurement incorporated into clinical care

- System aware

- Intervention

- Adherence assessed

- Adherence assessed

- Poor adherence

- Good adherence

- Better HIV outcomes

- Patient factors
  - *substance abuse
  - *mental illness
  - *other
Extension of Chronic Care Model

- Clinical information system, delivery system design, decision support
  - “Listening to the patient’s voice in a systematic standardized way”
  - Delivering data to providers using 21st century informatics tools
Assessment

- We use an open-source, non-proprietary web-based survey software application
- Surveys are completed on touch-screens
  - Facilitate data collection, decrease staff burden eliminating scoring and data-entry time compared with the use of paper forms, and also allows immediate access to results
  - Highly acceptable and feasible among HIV-infected patients in routine clinical care
    Skip patterns, CAT
- Encrypted SSL/TLS
- English and Spanish and most recently Amharic
- Tracks patient eligibility and time since last assessment, time to complete each assessment as well as time to complete each item and instrument for each patient

Fredericksen R et al. Journal of AIDS and HIV Research
## Domains

<table>
<thead>
<tr>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARV adherence</td>
</tr>
<tr>
<td>Depression</td>
</tr>
<tr>
<td>Anxiety</td>
</tr>
<tr>
<td>Alcohol use</td>
</tr>
<tr>
<td>Substance use</td>
</tr>
<tr>
<td>Health related quality of life</td>
</tr>
<tr>
<td>Symptom burden</td>
</tr>
<tr>
<td>Body morphology</td>
</tr>
<tr>
<td>HIV Risk Behavior</td>
</tr>
</tbody>
</table>

Assessments on tablet PCs with touch screens every 6 months, contains between 69 and 127 items depending on responses.
Clinic buy-in: provider assessment of adherence

- 62 of initial 500 patients self-reported very poor adherence
- Providers documented (same day):
  - Inadequate adherence for only 17 (27%)
  - No mention of adherence for 25 (40%)
  - Good adherence for 20 (32%)

Provider documented adherence assessments among 62 with poor adherence:

<table>
<thead>
<tr>
<th>17 inadequate</th>
<th>25 no mention</th>
<th>20 good adherence</th>
</tr>
</thead>
<tbody>
<tr>
<td>green</td>
<td>yellow</td>
<td>red</td>
</tr>
</tbody>
</table>
Ever evolving, do not “set it and forget it”

- 4 mos after implementation: survey questions were re-ordered
  - Items most relevant to clinical care presented first
  - Response data reports are printed & delivered to clinicians prior to PRO completion if necessary
- 5 mos after implementation: social workers receive PRO results
  - …when criteria for high-risk behavior are met in areas of adherence, HIV transmission risk, substance abuse, and depressive symptoms
  - PRO print-out automatically generated for social workers when these criteria are met
  - Approximately 70% of our patients meet these high risk criteria
- 7 mos after implementation: we added a Spanish option
- Still evolving languages, clinic flow, EHRs
Findings from observational work flow analyses and usability testing

- System is promoting awareness of previously unrecognized/under-recognized issues: e.g. unrecognized alcohol use
- Reports serve as “conversational icebreakers” for providers to engage patients
- System implementation has been minimally disruptive to clinic workflow
PROs in CNICS

- Extension to Fenway (Boston), UCSD, (large Latino population; Spanish essential), others (UCSF, JHU, UNC)
  - Different EHR systems, different leadership / clinic cultures, different patient groups, different patient flow
- UAB as example: patient flow circuit, “ticket numbers”, ROS at every visit
- All PRO collection is local!
## Patient vs. provider priorities

<table>
<thead>
<tr>
<th>Domain</th>
<th>Provider ranking</th>
<th>Patient ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication adherence</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>HIV symptoms</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Alcohol abuse</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Cognition</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Sexual risk behavior</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>HIV stigma</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Positive affect</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Sexual function</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Social roles</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Spirituality</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROMIS I Domain</th>
<th>Provider ranking</th>
<th>Patient ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Physical function</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Pain</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Anxiety</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Fatigue</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Anger</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

Fredericksen *et al.*, AIDS Care, 2015, *In press*
Impact on care

Provider documentation in the 8 months before and after initiation of provider feedback for patients with at-risk symptoms and behaviors:
- Before
- After

- Depression
- Inaccurate adherence
- At-risk alcohol
- Substance use
Selected CNICS findings after 45,078 completed assessments from >10,000 individuals: basis for clinical research relevant to improving care for persons living with HIV, 20+ papers, etc.

- Moderate to severe depression 22%
- Anxiety 20%
- Moderate-severe lipoatrophy 4%
- Moderate-severe lipohypertrophy 7%
- Any illicit drug use 68%
- Current illicit drug use inc. marij. 49%
- Current illicit drug use excl marij 28%
- At-risk alcohol use 34%
- Current ART 88%
- Missed any doses in prior 4 days 23%
PROMIS

- NIH’s largest investment in PROs and psychometrics
- Designed by people with expertise in instruments
  - Standardization across domains
  - Readability
  - Translatability
- Allows short forms or even better CAT
- Population based norms
- Captures longitudinal change
- Many strengths but not always the domains or content most relevant for clinical care
Lessons learned

- It is feasible to collect PROs in busy, multi-provider HIV clinics with feedback results to the providers in real time to impact care.
- Stand alone platform allowed us to implement across sites with many different EHRs (4 of whom have changed EHR since implementation).
- All clinical care is local and changes over time.
- Once integrated into a clinic, we have worked hard to maintain support.
- Additional features such as real-time, automated pager notification when patients indicate suicidality are especially valuable to providers.
- Automated notification of qualification for particular studies such as depression treatment studies based on elevated depression scores particularly valued by researchers.
Acknowledgements

- Patients, providers, and staff members
- Colleagues: too many to name
- University of Washington Madison HIV clinic
- CNICS sites: UW, UAB, UCSD, UCSF, Fenway, JHU, UNC
- CFAR
- PROMIS
- CHARN, particularly the Fenway node
- NIH NIMH R01
- PROMIS PCORI
HIT in the Context of Research

Fred D. Rachman, MD
Reasons for Health Centers to participate in research

• Mechanism to evaluate existing an new interventions to benefit patients
• Recruitment and retention strategy for practitioners
• Opportunity to highlight care provided and issues faced by patient populations
• Access to new resources
Benefits of CHC Sector

Representation of disparate populations
Complete spectrum of health services
Stability of service population
Quality orientation – desire to implement evidence based practice and contribute to improvement
Network infrastructure to support multiple site studies
Experience with HIT
Challenges for Community Health Centers

• Negative experiences of communities with research
• Lack of infrastructure at Health Center level
• Discontinuity between academic and clinical settings
• Relative lack of resources directed to community setting.
Challenges in the Healthcare Setting

- Increasing complexity of Health Care
- Challenges in coordination/communication among medical providers and between supportive services and disciplines
- Pressures to focus on providing care: limited time for interaction
- Difficulty accessing information in timely/organized fashion for use in decision making at the point of service
- Labor intensity limiting population based data to inform system change
Potential of EHRS for Research

• Decision Support
  ➢ Client recruitment
  ➢ Study protocols

• Performance Measurement
  ➢ Data collection
  ➢ Study protocols

• Remote hosting
  ➢ Communication between study sites
  ➢ Off site intervention

• Data warehouse/reporting
  ➢ De-identification of data
  ➢ Aggregation and analysis of data
Types of Research Studies

• Descriptive
• Correlational
• Experimental
• Meta Analysis
Cohort Studies

Residents without breast CA identified

Followed till they develop breast CA

Patients with colon cancer identified

Stage of colon cancer classified

Patients followed till they die
Case Control Studies

Ask them to recall their diet, medication use, treatment, etc.

Patients with breast CA who died

Patients with breast CA who are alive
Approaches to Use of EMR

• Direct reporting from EMR or other source system
• Extract data and place in another analytic environment
• Create a data registry
• Create data mart or data warehouse
Data Mart

• Often holds only one subject area
• May hold more summarized data (although many hold full detail)
• Concentrates on integrating information from a given subject area or set of source systems
• Uses a dimensional model to facilitate queries and analysis
Data Warehouse

- Holds multiple subject areas
- Holds very detailed information
- Works to integrate all data sources
- Does not necessarily use a dimensional model but feeds dimensional models.
20 states
>45 health center organizations using Alliance-developed CDS for their Electronic Health Record (EHR)
200+ service delivery sites
500+ FTE medical providers
200+ FTE behavioral health providers
>800,000 patients served
1,200,000 patient encounters annually
Chicago Health Center Average Body Mass by Community Area

Legend
Average BMI
- 22.65 - 23.43
- 23.43 - 24.31
- 24.31 - 25.51
- 25.51 - 26.73
- 26.73 - 42.13
Algorithm 3: All patients that had any 3 encounters with an SBP >= 140 OR DBP >=90 12 months previous to their most recent encounter (The most recent encounter may be the third qualifying encounter).

Definition 3 Results

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAGNOSED</td>
<td>2172</td>
<td>86%</td>
</tr>
<tr>
<td>UNDIAGNOSED</td>
<td>353</td>
<td>14%</td>
</tr>
</tbody>
</table>
Alliance Research Priority Areas

- Areas of Interest:
  - Health Disparities
  - Cost Effectiveness
  - Testing New Devices
  - Organizational Process Research
  - Clinical Effectiveness
  - Secondary Analysis of Existing Data

- Eager to collaborate with research partners with shared vision
Internal Research Activities

• Prep to research queries
• Clinical quality measure testing
• Clinical Decision Support design and testing
• Simple EMR driven intervention trials
• Observational HIT studies
• Public Health Surveillance
Microsoft’s Big Data Solution

- Microsoft SQL Server 2012
- Microsoft Analytics Platform System (APS)
- Microsoft HDInsight (Hadoop): Unstructured, Free Text Data

Data Ingestion Layer

- SAP Data Services
- SSIS
- Sqoop (Scoop into Hadoop)

Data Sources

- GE EMR
- Clinical Data
- Research Data
- Other EMR Data
- Future Data
Community Health Applied Research Network (CHARN II)

Funder: HRSA; Contract with Kaiser Permanente
Timeline: 4/10/14 – 4/9/17
Sites: Erie, Near North, North Country, PCC, Heartland Health Outreach, Howard Brown Health Center

Deliverables:

<table>
<thead>
<tr>
<th>Data Use Agreements and IRB</th>
<th>Website Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approve DUAs and Conduct IRB modifications/approvals</td>
<td>Update internal project management website</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Steering Committee Participation</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation on weekly telephone/bi-annual in-person meetings; Consultation on Data Access and Sustainability plans and Safety-Net Research Agenda</td>
<td>Provide input on Data Plan, assist with identification of new data elements for CHARN Data Warehouse, and update Warehouse on regular basis.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subcommittee Membership</th>
<th>Dissemination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation on Executive Committee and Research Planning, Data Sharing, and IRB Subcommittees</td>
<td>Lead the coordination of at least one manuscript each project period (one year).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Work Group Participation</th>
<th>Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead and administratively staff one project-related work group. Examples include Insurance, Cardiovascular Disease, and Smoking Cessation work groups.</td>
<td>Lead the development of one proposal each project period (one year), identifying external funding sources to support community-driven topics.</td>
</tr>
</tbody>
</table>
Aims of CHARN

• Create infrastructure for pooling EMR data across different sites
• Develop improved approaches for transferring research findings into practice
• Foster practice-based collaboration among CHCs, practitioners and academic medical center researchers
• Develop and conduct study protocols
• Train CHC personnel in research methods and protocols
• Develop research proposals for additional funding through other federal agencies
Governance Structure

Steering Committee

Executive Committee

Research Planning & Publication Subcommittee

IRB & Regulatory Subcommittee

Data Sharing & Registry Design Subcommittee

Communication & Community Relations
CHARN Registry Tables

- **Encounter Data**
  - Date, type, patient ID number,

- **Patient Data**
  - Age, sex, race, ethnicity, primary language

- **Diagnosis/Comorbidities**
  - Conditions: HIV, HEP B and C, CVD, Hyperlipidemia, Asthma, COPD, Mental Illness

- Lab data
- Medications
- Insurance
- Vital signs
<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Ethnicity</th>
<th>Encounter Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients</strong></td>
<td>519,636</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>302,311</td>
<td>Hispanic or Latino</td>
<td></td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>217,169</td>
<td>Not Hispanic or Latino</td>
<td></td>
</tr>
<tr>
<td><strong>&lt; 18 years old (male and female)</strong></td>
<td>155,531</td>
<td>Missing (reported unknown)</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td>Missing (left blank)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>314,487</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black/African American</td>
<td>94,849</td>
<td>Primary Care/Outpatient NOS</td>
<td>217,616</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>4,500</td>
<td>ER/Urgent Care</td>
<td>15,407</td>
</tr>
<tr>
<td>Asian/NHOPI</td>
<td>91,092</td>
<td>Specialty/Referral</td>
<td>27,615</td>
</tr>
<tr>
<td>Multi-racial</td>
<td>4,964</td>
<td>Lab</td>
<td>59,705</td>
</tr>
<tr>
<td>Other</td>
<td>26,848</td>
<td>Pharmacy</td>
<td>20,622</td>
</tr>
<tr>
<td>No race indicated (missing)</td>
<td>57,328</td>
<td>Other</td>
<td>32,978</td>
</tr>
</tbody>
</table>
Current CHARN Research Activities

• Cardiovascular Disease
• Substance Abuse
• Risk and Resiliency
• Infectious Disease
• Insurance
Chicago area Patient-Centered Outcomes Research Network (CAPriCORN)

- To capture longitudinal clinical information on more than 1 million patients (~50% nonwhite)
- Develop the capacity to efficiently conduct comparative effectiveness research trials and observational studies
- Establish procedures for clinical data standardization and inter-operability across the national patient-centered research network of clinical data research networks (CDRNs) and patient-powered research networks (PPRNs)
- Engage patients, clinicians, and health system leaders in governance and use of CAPriCORN resources.
Who is CAPriCORN?
• The PopMedNet™ software application enables simple creation, operation, and governance of distributed health data networks.

• It facilitates distributed analyses of electronic health data to support medical product safety, comparative effectiveness, quality, medical resource use, cost-effectiveness, and related studies.
Capabilities of PopMedNet

• Allows users to send questions to the data
• Provides secure, customized, private portals, and file transfer capabilities that allow users to query data held by disparate partners
• Allows participating network data partners to maintain physical and operational control over their data
• Supports both menu-driven analyses and distribution of complex analytic programs
• Accommodates any network size, from single datasets held by a single study to multi-year projects encompassing dozens of organizations and multiple projects
• Accommodates any data model
WHY HEALTHCARE DATA IS DIFFICULT

- **Data Format**: It’s not all digitized in healthcare
- **Data Definitions**: Subjective based on source
- **Data Complexity**: From humans to data warehouses
- **Data Structure**: Structured vs non-structured
- **Location**: Where is your data?
- **Regulations & Requirements**: Keeping up with the government

![Late Binding Data Warehouse](image)
Data processes

Concept
  Definition
    Specification (data element and measure)
    Capture
    Analysis
    Reporting
    Validation
    Visualization
What does the data truly represent?

Capture of data element from data source outside the EHRS – no formal arrangement (e.g. colonoscopy)

Capture of data element from data source outside the EHRS – formal arrangement for resulting (e.g. eye exam from formal referral resource)

Capture of data element requiring entry of observation in standardized way by practitioner (e.g. foot exam)

Capture of data element as easily objective defined observation captured by EHRS (e.g. blood pressure)

Direct electronic of data element and/or result through order entry or interface (e.g. Hgb A1C measure and result)
De-identified Data

- De-identified health information is individually identifiable health information from which all potentially identifying information has been removed.
- Examples of identifiable information: Name, address, social security number, birthdate, zip code, full date of clinical date.
Correlation between EHRS elements and research plan for clinical trials

Evidence based practice guideline

- Research protocol

Data elements defined

- Subject criteria, pre and post data elements

End user form designed to provide decision support at point of patient care

- Study protocols

Measures defined and Data elements mapped to reports

- Baseline and study data collection plan
Challenges in use of de-identified data

- Places limits on some variables that might be of interest – eg location/community area
- Makes it difficult to track patients across multiple locations/databases
- May result in duplicate counts

Solutions: Matching algorithms and limited data sets.
Limited Data Sets

• a limited set of identifiable patient information as defined by “HIPAA” that may be disclosed to an outside party without a patient’s authorization if certain conditions are met:
  • The purpose of the disclosure may only be for research, public health or health care operations.
  • The person/entity receiving the information must sign a data use agreement

• Examples of limited data set information include:
  • dates such as admission, discharge, service, DOB, DOD;
  • city, state, five digit or more zip code; and
  • ages in years, months or days or hours.
Reliability and Validity

- Reliability: Is the data collected consistently
- Validity: Is the data accurately representing the concept
Data Transformation

• Conversion of the data format of a source data system into the data format of a destination data system such as a data warehouse

• Data **transformation** can be divided into two steps:
  • **data mapping** maps **data elements** from the source data system to the destination data system and captures any transformation that must occur
  • **code generation** that creates the actual transformation program
Big Data

A collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools. The challenges include:

• Capture
• Storage
• search
• Sharing
• Analysis
• Visualization.