Analytics, Big Data, and Partnerships

Connie White Delaney, PhD, RN, FAAN, FACMI
Dean & Professor, School of Nursing*

Tom Clancy, PhD, MBA, RN, FAAN
Clinical Professor & Associate Dean for Faculty Practice, Partnerships, & Professional Development*

Karen Monsen, PhD, RN, FAAN
Associate Professor & NI Specialty Coordinator*

Judy Pechacek, DNP, RN
Clinical Assistant Professor & Director of DNP Programs*

* University of Minnesota

November 7, 2015
8:30-4:00
Orlando, FL USA

Big Data: It’s a Big Deal!

Thomas R. Clancy, PhD, MBA, RN, FAAN
Clinical Professor and Associate Dean for Faculty Practices, Partnerships and Professional Development, University of Minnesota School of Nursing
Disclaimer

- I have no relevant financial interest to disclose nor am I endorsing any of the commercial products identified in this presentation.

Objectives

- Review emerging trends that are driving the creation of “Big Data”.
- Describe the technology infrastructure supporting “Big Data”.
- Provide exemplars on how “Big Data” will change healthcare.
Emphasis on EHR & Meaningful Use

Utilize technology
- Basic EHR functionality, structured data
- Privacy & security protections
- Care coordination
- Patient informed
- Structured data utilized
- Privacy & security protections

Transform health care

Access to information

Stage 1 MU
Stage 2 MU
PCMH 3-Part Aim
ACO’s “Stage 3 MU”

The Digitization of Healthcare Meaningful Use

- Projections are for 90 percent of providers to have access to a fully operational electronic health records by 2019, up from 34-35 percent in 2011.

United States EMR Adoption Model™

<table>
<thead>
<tr>
<th>Stage</th>
<th>Cumulative Capabilities</th>
<th>2012 Q2</th>
<th>2013 Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 7</td>
<td>Complete EMR, CDA transactions to other EMR, Data warehousing, Data continuity with EMR/other systems</td>
<td>1.7%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Stage 5</td>
<td>Physician documentation (structured responses), EMR/CDI, variance &amp; compliance, IT (IPACS)</td>
<td>6.0%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Stage 5</td>
<td>Closed loop medication administration</td>
<td>10.5%</td>
<td>12.3%</td>
</tr>
<tr>
<td>Stage 4</td>
<td>CPOE, Critical Decision Support (clinical pathways)</td>
<td>13.3%</td>
<td>14.2%</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Nonclinical documentation (flow sheets), CDA, lab results (LOINC)</td>
<td>42.3%</td>
<td>41.3%</td>
</tr>
<tr>
<td>Stage 2</td>
<td>CPR, Controlled Medical Vocabularies, CDS, may have inconsistent imaging, MI: capable</td>
<td>11.7%</td>
<td>11.5%</td>
</tr>
<tr>
<td>Stage 1</td>
<td>Ancillaries - Lab, Radiology, Pharmacy - All Installed</td>
<td>5.1%</td>
<td>4.8%</td>
</tr>
<tr>
<td>Stage 0</td>
<td>All Ancillaries Not Installed</td>
<td>7.5%</td>
<td>7.4%</td>
</tr>
</tbody>
</table>

Data from HIMSS Analytics Database 6/30/12

N = 15350
N = 2116
**Big Data: Healthcare**

- **CMS** – Center for Medicare & Medicaid
- **NIH** - Clinical Translational Science Award
- **Commercial** – Optum Labs Research Collaborative

- 2 billion data points per year
- 62 medical research institutions in 32 states
- 150 million lives, 3200 data points per life, over 20 years.

---

**Emphasis on EHR & Meaningful Use**

<table>
<thead>
<tr>
<th>Access to information</th>
<th>Transform health care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilize technology</td>
<td></td>
</tr>
<tr>
<td>Basic EHR functionality, structured data</td>
<td>Improved population health</td>
</tr>
<tr>
<td>Privacy &amp; security protections</td>
<td>Enhanced access and continuity</td>
</tr>
<tr>
<td>Care coordination</td>
<td>Data utilized to improve delivery and outcomes</td>
</tr>
<tr>
<td>Patient informed</td>
<td>Patient self management</td>
</tr>
<tr>
<td>Evidenced based medicine</td>
<td>Patient engaged, community resources</td>
</tr>
<tr>
<td>Structured data utilized</td>
<td>Patient centered care coordination</td>
</tr>
<tr>
<td>Privacy &amp; security protections</td>
<td>Team based care, case management</td>
</tr>
<tr>
<td>Registries for disease management</td>
<td>Registries to manage patient populations</td>
</tr>
<tr>
<td>Privacy &amp; security protections</td>
<td></td>
</tr>
</tbody>
</table>

| Stage 1 MU | Stage 2 MU | PCMH 3-Part Aim | ACO’s “Stage 3 MU” |
Moores Law

1. Exponential growth in computer processing speed,
2. The digitization of everything,
3. Build-out of the Intranet,
4. Recombination of existing technologies

---

Exponential Growth in Processing Speed

**ASCII RED: Year 1996**
- Speed: 1 TFLOP
- Size: 1625 Square Feet
- Cost: $55 Million

**Play Station 3: Year 2005**
- Speed: 1.8 TFLOP
- Size: DVD player
- Cost: $500

---

http://www.top500.org/featured/systems/ascii-red-sandia-national-laboratory/
Digitization of Everything & Sensor Technology

- Transformation from analog to digital format has massively accelerated growth in technology.
- Once created digitized data is practically free to copy & reproduce.
- *One trillion sensors today!*

The Quantified Self Movement

**Wearable Computing**

- Activity monitors
- Diet and weight
- Sleep and mood
- Vital signs
- Insulin and blood sugar
- Medication compliance
Geo-Spatial Data

Data which has a geographical or geospatial aspect including the use of geographic information systems and geometrics (GPS).

The geospatial web – blending physical and virtual spaces, Arno Scharl in receiver magazine, Autumn 2008


Hot Spotting

Intranet of Things


Build-Out of the Intranet

- Every day, 2.5 quintillion bytes of data is created.
- 90% of the data in the world today has been created in the last two years alone.

Quintillion - (a number equal to 1 followed by 18 zeros) 1,000,000,000,000,000,000

Social Media Data
Top Five US Websites in 2015

<table>
<thead>
<tr>
<th>Website</th>
<th>Subscribers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>900,000,000</td>
</tr>
<tr>
<td>Twitter</td>
<td>310,000,000</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>255,000,000</td>
</tr>
<tr>
<td>Pinterest</td>
<td>250,000,000</td>
</tr>
<tr>
<td>Google+</td>
<td>120,000,000</td>
</tr>
<tr>
<td></td>
<td>1,835,000,000</td>
</tr>
</tbody>
</table>
**Emobile Explosion**

- 85% of adults own a cellphone.
- 76% own a computer
- 80% have access to the Intranet
- Mobile device growth is estimated at 30% per year.
- 10 billion users by 2020

---

**Mobile Health (mHealth)**

Medical and public health practices supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices.

http://healthinformatics.wikispaces.com/mHealth

mHealth: New horizons for health through mobile technologies, World Health Organization, Global Observatory for eHealth series - Volume 3, 2011
Whole Gene Sequencing Data

- Process that determines the complete DNA sequence of an organism's genome at a single time.
- Data will be an important tool to guide therapeutic intervention in the future.

There are six billion base pairs in each human diploid genome

---

Growing Value of Secondary Use Data

<table>
<thead>
<tr>
<th>Secondary Data</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google search data</td>
<td>Forecasting epidemics</td>
</tr>
<tr>
<td>Insurance claims data</td>
<td>Personalized premiums</td>
</tr>
<tr>
<td>Streaming data</td>
<td>Event prediction</td>
</tr>
<tr>
<td>Social network (twitter)</td>
<td>Moods swings (depression)</td>
</tr>
<tr>
<td>Electronic health records</td>
<td>Descriptive and predictive Health analytics</td>
</tr>
</tbody>
</table>

Human Graphic Information System (GIS)

- Multiple layers of demographic, physiologic, anatomic, biologic and environmental data about a particular individual.
- Possible, but not feasible.

Knowledge Engineering

Data Mash-ups

Knowledge Value

Knowledge Engineering

Knowledge engineering of big data will dramatically change how healthcare providers treat and care for patients through their ability to **describe, predict and prevent** health events. Knowledge Engineering and Personalized Medicine

Knowledge Engineering

- Decision trees
- Association rules
- Artificial neural networks
- Support vector machines
- Clustering
- Bayesian networks
- Genetic algorithms

Machine Learning

A simple neural network

![A simple neural network diagram](image)
Healthcare Data Constraints

Data takes on multiple forms:
- Structured & Unstructured text
- Audio (dictation)
- Images (PAC's)
- Videos
- Waveforms (ECG)
- Streaming (Sensors)

Healthcare Data Constraints

Data is stored in multiple locations and silos.
- Registration/Scheduling
- EHR
- Financial
- Human Resources
- Support (lab, radiology)
- Material Management
- Portals and personal health records
- External (clinics & other)
What Makes Data “Big Data”?

When data processing grows beyond conventional methods of analysis data becomes “Big Data”.

Healthcare Data Solutions

Traditional Data Management
- Centralized data warehouse
- Expensive & time consuming
- Limited to structured data
- Relational databases & tabular structure
- Data is normalized
- Slow processing

New Data Management
- Distributed data warehouse
- Inexpensive and efficient
- Structured and Unstructured
- Parallel processing
- Data is “messy”
- Fast processing
## The New Data Management Platform

### The New Enablers of Big Data
- Hadoop
- MapReduce
- Star
- Spark
- Cassandra
- NoSQL
- Lucene
- Solr
- Hive
- Spark
- Shark

### IBM BlueGene

**Massive Parallel Processing**

---


---

<table>
<thead>
<tr>
<th>Big Data Governance</th>
<th>Operations Management</th>
<th>Security, Monitoring, Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation Layer</td>
<td>Analytics Layer</td>
<td>Data Management Layer</td>
</tr>
<tr>
<td>Data Visualization</td>
<td>Data Mining, Pattern</td>
<td>Data Integration/Management</td>
</tr>
<tr>
<td>Tools, EMR System,</td>
<td>Recognition Engine,</td>
<td>Meta-data repository,</td>
</tr>
<tr>
<td>Analytics-enabled</td>
<td>Predictive Modeling</td>
<td>Distributed data warehouse,</td>
</tr>
<tr>
<td>Workflow, Live</td>
<td>Engine, Classification</td>
<td>Data Integration,</td>
</tr>
<tr>
<td>Dashboards,</td>
<td>Engine, Classification</td>
<td>Management,</td>
</tr>
<tr>
<td>Applications User</td>
<td>Engine, Inference</td>
<td>Data Integration,</td>
</tr>
<tr>
<td>Interface</td>
<td>Engine, Natural Language Processing</td>
<td></td>
</tr>
</tbody>
</table>

### Data Connection Layer
- Data Extraction
  - Data Ingestion Tools, Data Extract, Transfer, Load, ETL (tools), HL7/XDX/CCDA Exchange Pipelines, Data Bus Manager
1. Monitoring Systems: Clinical and Business Intelligence Dashboards

Performance Analytics Dashboard

CAKE: Knowledge base

- CAKE is a very large knowledge base of associations amongst clinical concepts
  - KB is automatically generated by extracting and curating associations from different knowledge sources using proprietary big data algorithms
  - Associations are assigned strengths

With permission from eemeen Karimi, PhD, Aziz Boxwala, MD, PhD, Randolph Barrows, MD, MS
Problem-Oriented View

2. Predictive Systems: Modified Early Warning System (MEWS)

Scoring is based on:
- Respiratory rate
- Heart rate
- Systolic blood pressure
- Conscious level
- Temperature
- Hourly urine output (for previous 2 hours)

- WebMD Mobile
- iTriageHealth
- Medscape Mobile
- Diagnosaurus DDx
- Symptoms TakTools
- iHeadache
- SignsSx Handbook
- Symptom Mate
- Differential Dx i-pocket
- STATworkUP
- eRoentgen Radiology DX
- Symptom Minder

Pocket Symptom Analyzer

https://www.youtube.com/watch?v=CQqBMG578tA

Image: http://www.thinklabs.com/#/thinklink/cbor

Recombination of Technologies

iPhone
- Proximity sensor
- Ambient light sensor
- Accelerometer
- Magnetometer
- Gyroscopic sensor
- Camera/Video
- Voice recognition (Siri)
- Phone
- Email/Text
- Intranet
- NLP
4. Prevention: Asthma Attacks

Sensor Cluster:
- Air quality
- Pollen
- Inhaler use
- Geo-location
- Breath nitric oxide
- Lung function – Smartphone app
- RR, Temp, O2 sat.

Modern Healthcare

Registry-based clinical trial puts heart treatment to the test

A study challenging the value of a device used when unclogging blocked arteries is raising the prospect that inexpensive registry trials, which use observational data to assess outcomes, may soon be widely used to challenge expensive add-ons to medical procedures that may not improve patient care.

Experts say the new model holds promise as a quick, low-cost way to conduct large-scale research from registries and patient records, whose results can be disseminated far more quickly than traditional clinical trials. The growing availability of electronic health records is spurring interest in such trials to determine if high-cost medical devices and drugs perform as well in actual medical settings as they did in the original randomized clinical trials that led to their regulatory approval.
THE WALL STREET JOURNAL

TECHNOLOGY

Big Data's High-Priests of Algorithms

'Data Scientists' Meld Statistics and Software for Find Lucrative High-Tech Jobs

Saba Zuberi, an astrophysicist working as a data scientist at TaskRabbit, said working for a consumer Internet firm can be surprisingly rewarding. Ramin Rahimian for The Wall Street Journal

Hype or Hope?