Big Data Management and Advanced Analytics in Healthcare

Jason Burke
System VP and Chief Analytics Officer
UNC Health Care

Agenda

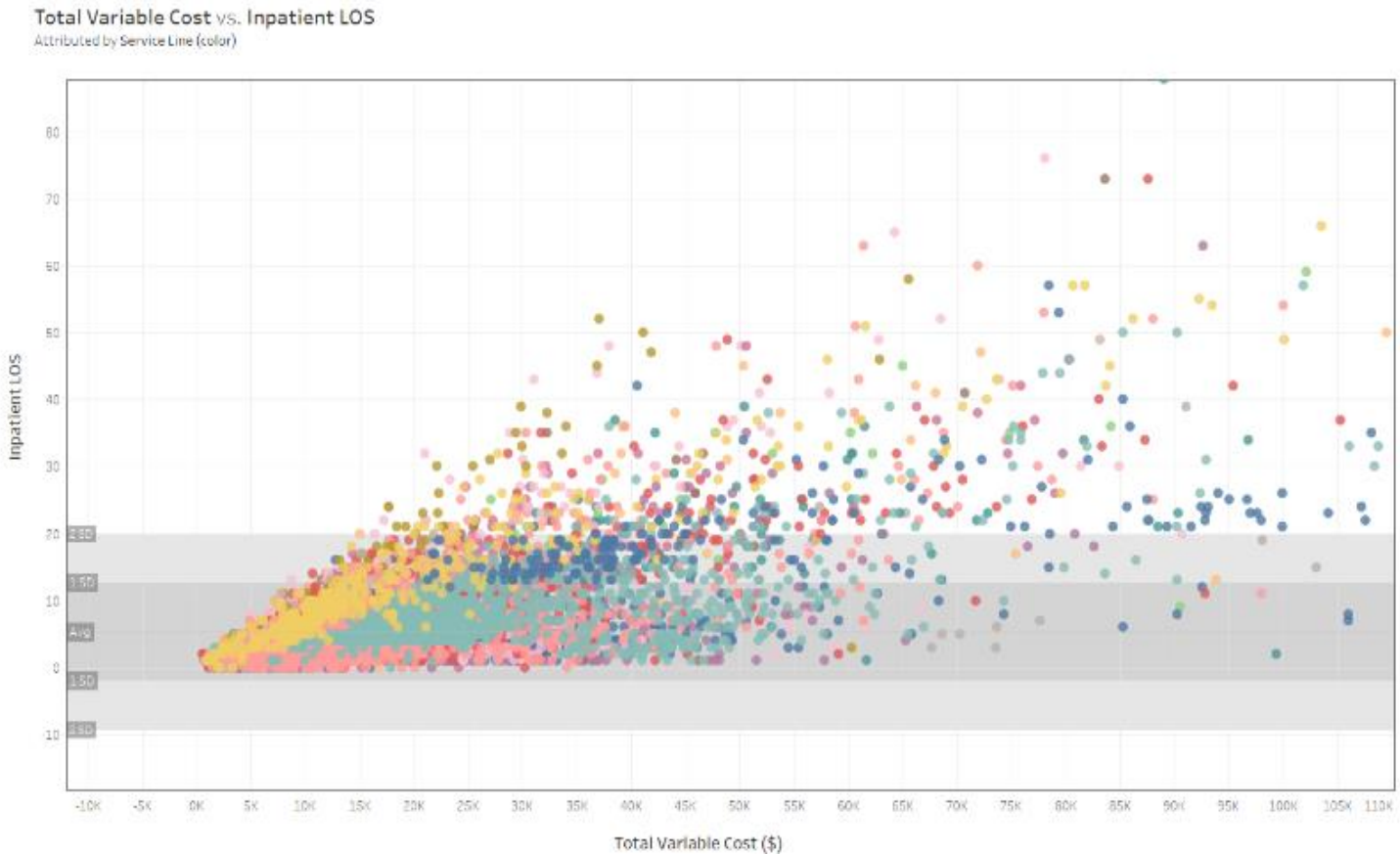
Define the evolving role of data management in the field of health care "big data"

Discuss some of the new competencies for effectively leveraging big data at scale

Share some examples of how this is being applied at a large health care system

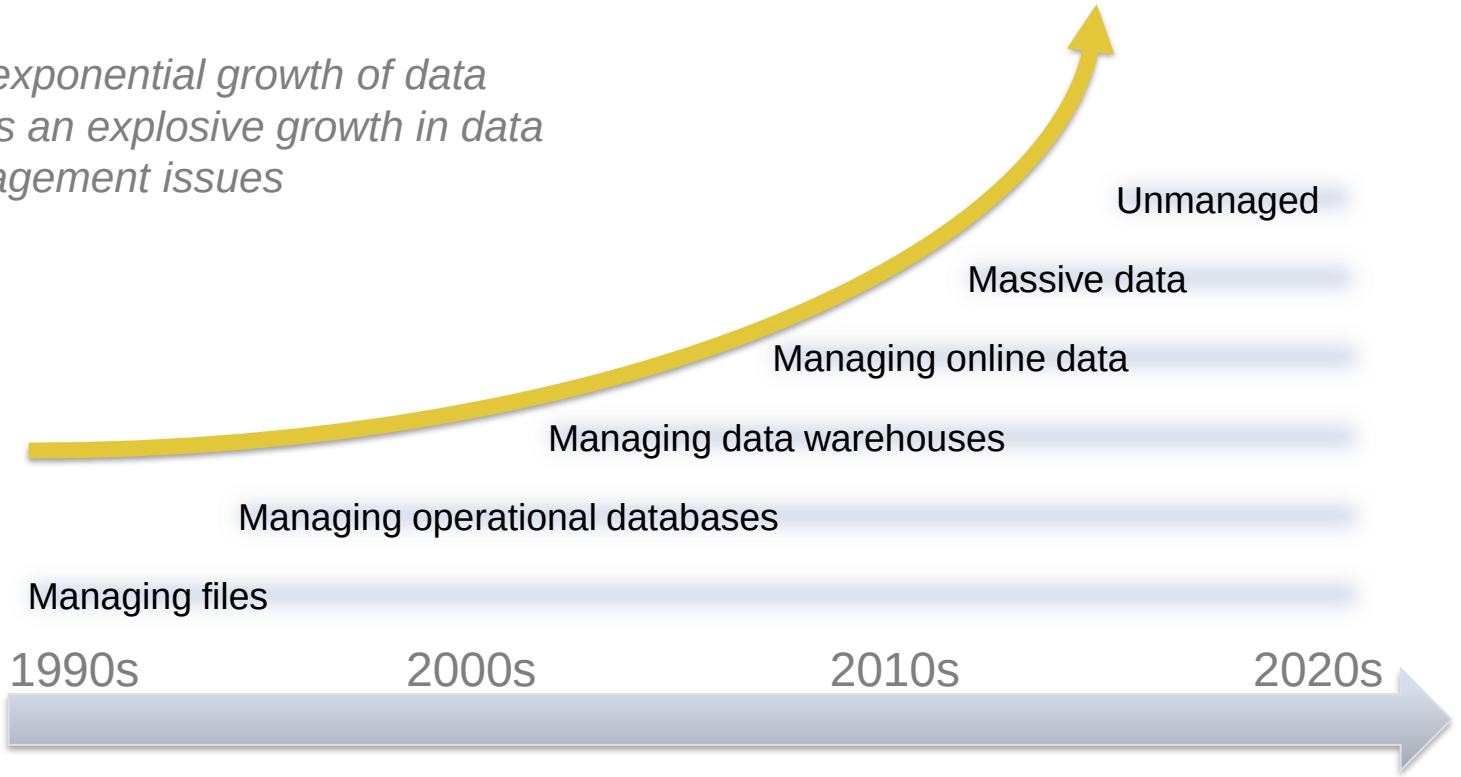
Here's Your Report!

Starting with the end in mind…



The Changing Role and Scope of "Data Management"

The exponential growth of data drives an explosive growth in data management issues



Human Valuation and Big Data

Valuation Area
What is the source of the data?
What do we know about this data and what it means?
What's in the data; how representative is it?
How was the data obtained and managed?
Who else is using this data?
How consistent is the data with other data?
How good was the process used to create this data?
How current is the data?
How much is it changing?

Implications of Today's Big Data

We have to know and manage a lot ABOUT our data, not just the data itself

What is the source of the data?
What do we know about this data and what it means?
What's in the data; how representative is it?
How was the data obtained and managed?
Who else is using this data?
How consistent is the data with other data?
How good was the process used to create this data?
How current is the data?
How much is it changing?

Lineage / Pedigree
Business Context
Quality
Reasonableness
Consistency
Pervasiveness
Socialization
Controls
Currency
Volatility

Why is this so different?

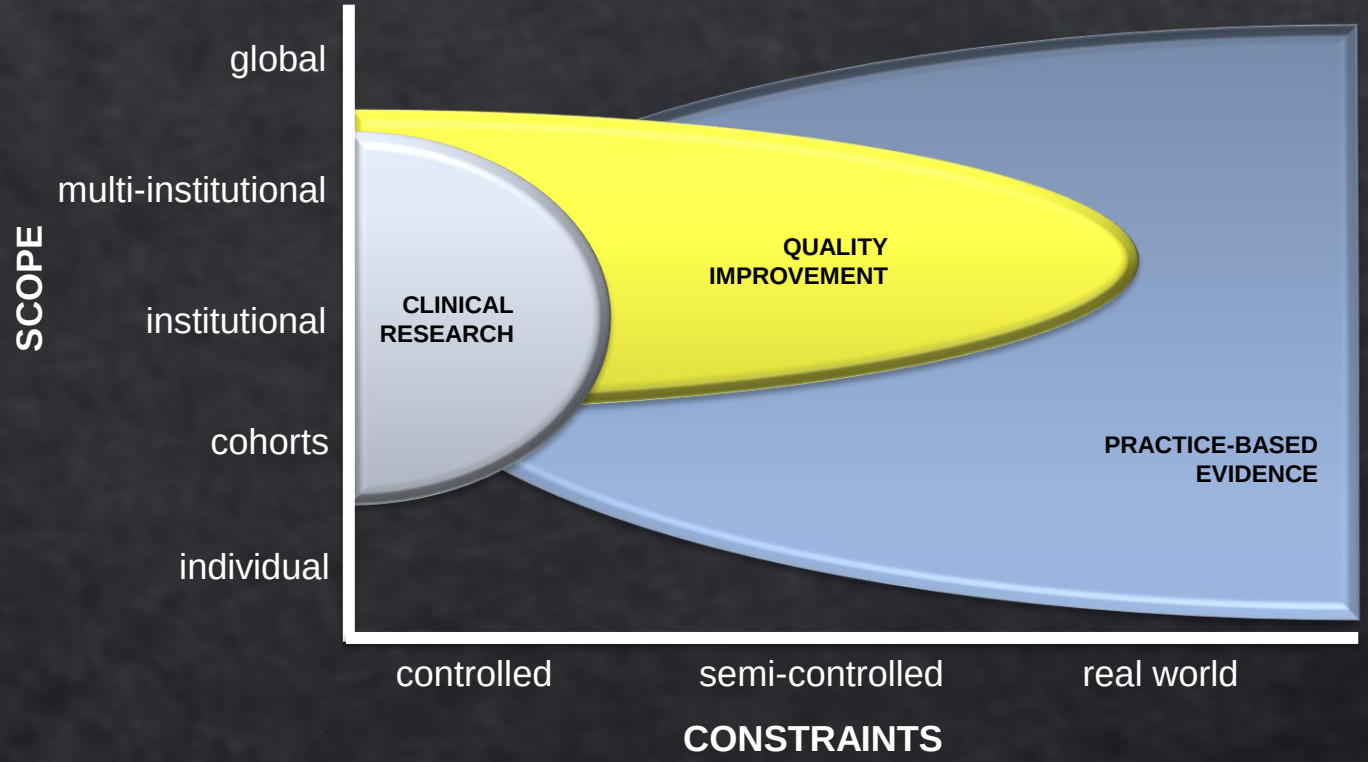


Image via Flickr user horiavarlan

Dismantling the Hype

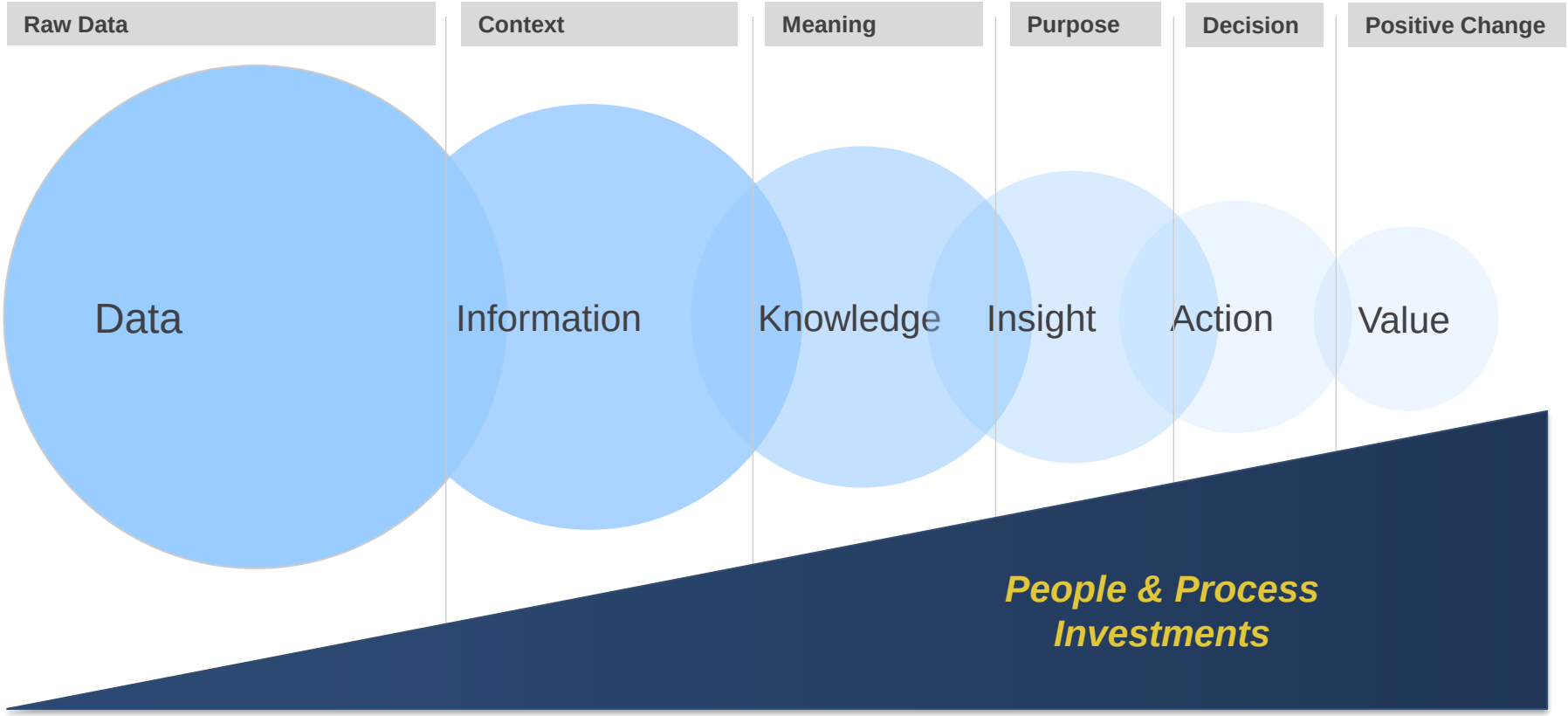
The Buzz

- A technology problem
- You can fix it with text mining, machining learning, AI, Hadoop, or another buzzword
- Standards is the fix
- Sticking to one vendor is a fix
- Outsourcing helps

The Reality

- A people / process problem
- Garbage in, garbage out…but we can improve it even if we can't fix it
- Standards help a lot
- Sticking to one vendor is impossible, but does help
- No vendor knows your business better than you

Harnessing Big Data is an Organizational Competency



Designing for Reusability

Image courtesy of http://www.flickr.com/photos/wonderlane

Reusability in Analytics: Products!

PROJECTS

- Stakeholders are project area experts
- Effort is focused on predefined questions
- Work is relevant to project team
- Timeline is project driven
- Data definitions are project specific
- Data structured for single use
- Little-to-no analytical code reuse
- Release available to project stakeholders

PRODUCTS

- Stakeholders are functional experts
- Questions are not predefined
- Work must be relevant to multiple customers
- Timeline is engineering driven
- Data definitions are enterprise-wide
- Data is structured for broad re-use
- Analytical models are built for multiple projects
- Release available to entire enterprise

Investing in Competencies

DATA GOVERNANCE	Data Policies and Standards	Data Roles & Stewardship	Governance & Decision Making	Master Data Management (MDM)	Asset Provisioning, Management & Certification	Data Quality Management
BUSINESS OPERATIONS	Data Strategy Formalism	Consulting and Guidance	Analytics Competency Development & Staffing	Data Operations Management and Controls	Knowledge Management	Community Engagement
ENGINEERING OPERATIONS	Engineering Management	Architecture Design and Management	Data Model Engineering & Management (Domains)	Analytical Model Engineering & Management	Lifecycle and Quality Management	Metadata Management

Centralized vs. Federated Capability Development

Like many academic medical systems, UNC HCS has a diversified, empowered culture

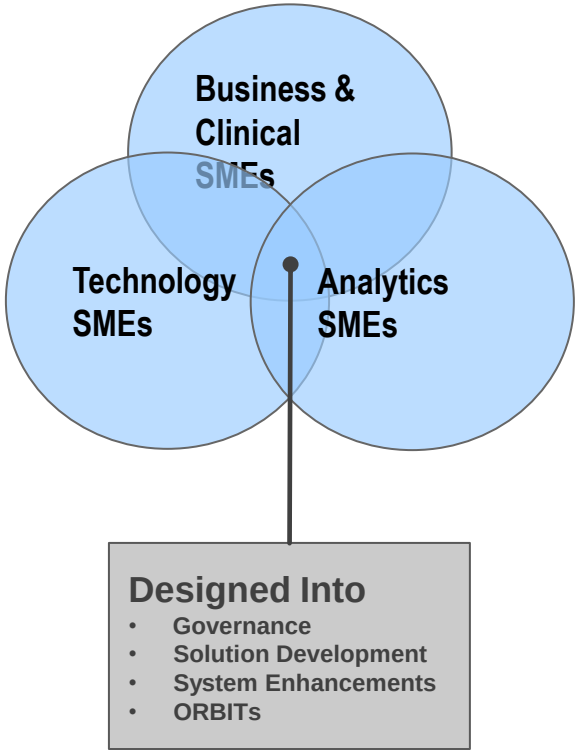
- Organizational units with deep subject matter experts (SMEs)
- Spirit of research, innovation, and entrepreneurship

Any journey with analytics must respect those cultural norms

- Fully centralized → "ivory tower", bottlenecks, loss of institutional context and SME
- Fully federated → no economies of scale, impossible to establish a single source of truths

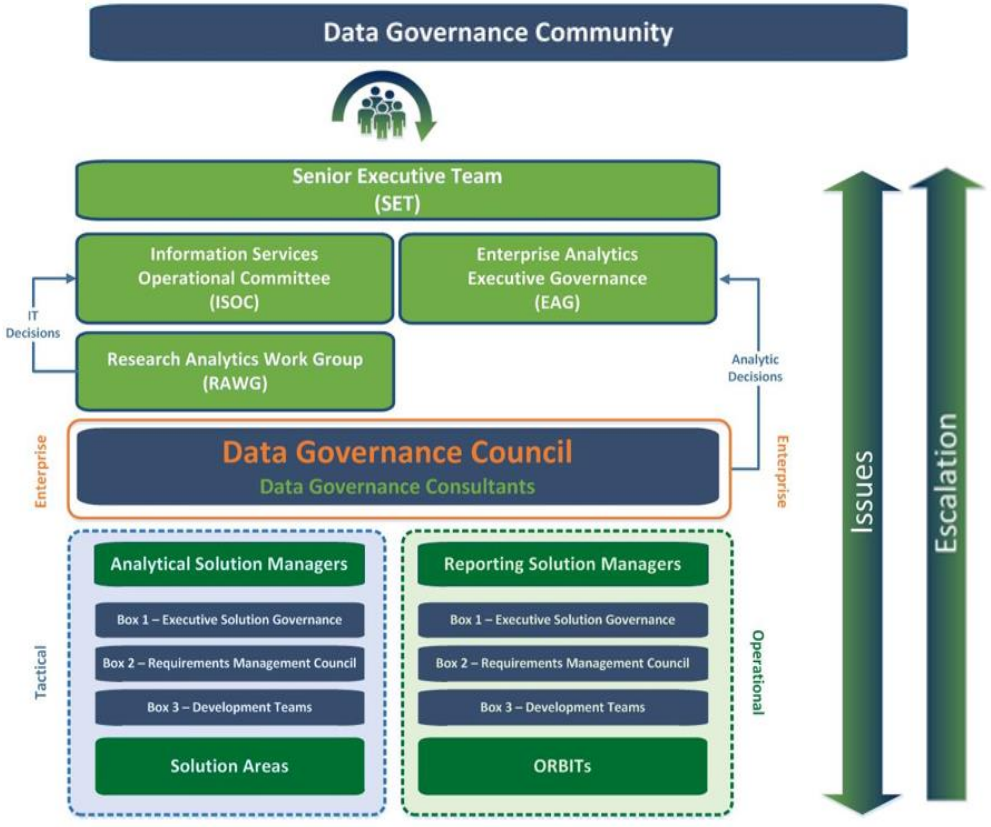
We opted to pursue a hybrid model

- Centralize building reusable assets
- Federate the use and extension of those assets
- Bring federated SMEs into all build-related work
- Help federated users be more effective with data / analytics

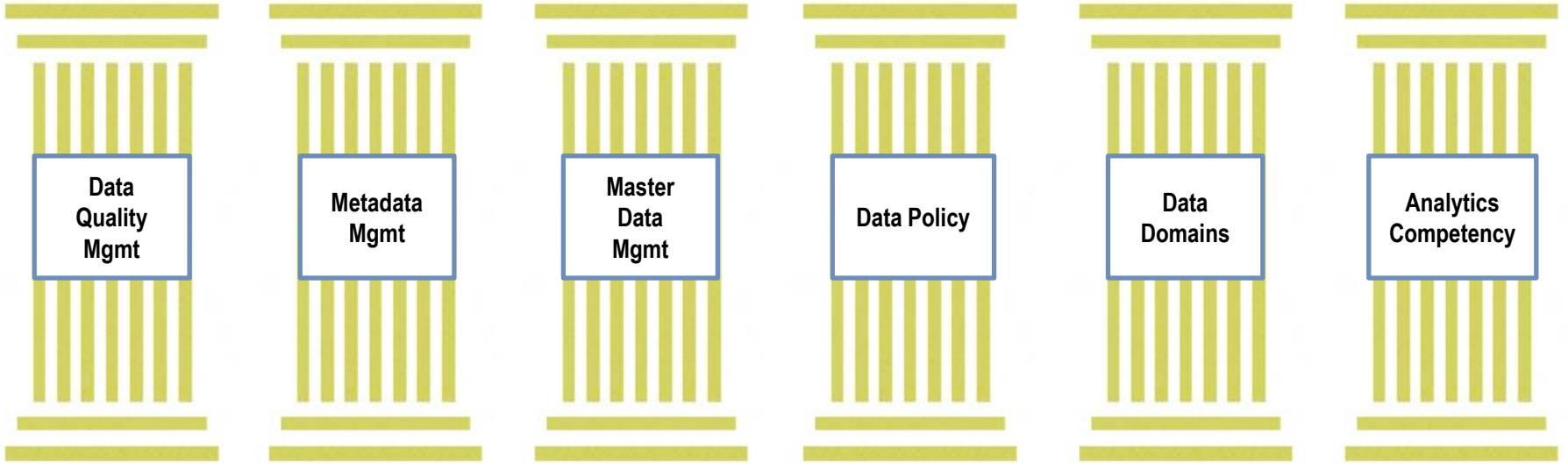


Data Governance

DATA GOVERNANCE	Data Policies and Standards	Data Roles & Stewardship	Governance & Decision Making	Master Data Management (MDM)	Asset Provisioning, Management & Certification	Data Quality Management
BUSINESS OPERATIONS	Data Strategy Formalism	Consulting and Guidance	Analytics Competency Development & Staffing	Data Operations Management and Controls	Knowledge Management	Community Engagement



Data Governance Program Pillars



- Data Profiling
- Transparency of results
- Quality thresholds
- Prioritized remediation efforts

- Application inventory
- System interfaces inventory
- Business glossary
- Application data dictionaries
- Core reports inventory
- Data release inventory

- Source of truth for key shared data
- Common terms and groupings for integrated analytics

- Assign decision rights and accountabilities
- Permissions
- Standardize data movement and access approval methods
- Further define regulation
- DG Maturity

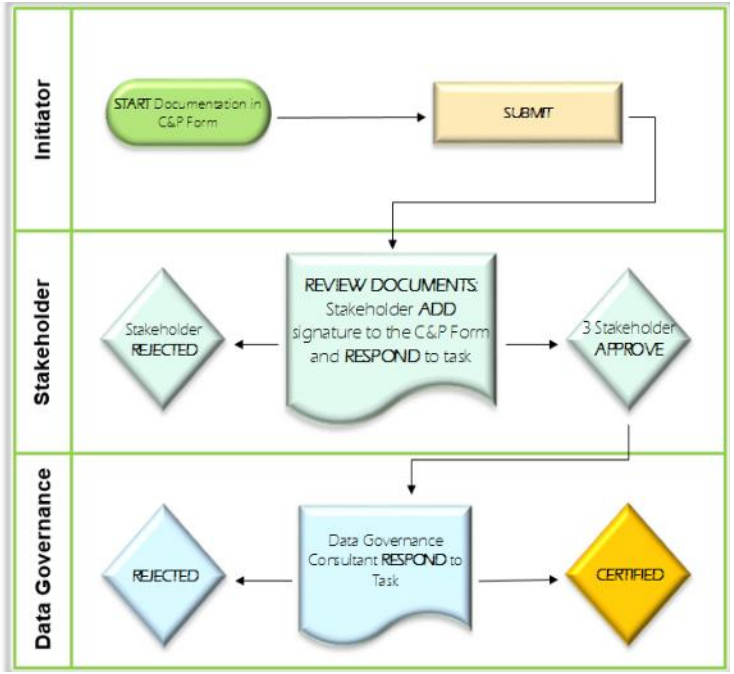
- Group data into unique domains
- Nominate Information Owners
- Data Steward selection and engagement

- Tool inventory
- Remove duplication of efforts and solutions
- Clearly prioritize analytics initiatives
- Drive progress toward self- service analytics

(Knowledgent, Mikol)

Asset Certification

- **Certification is a review process of key elements and tools used for decision making across the enterprise. Certification provides:**
  - Reliability
  - Traceability
  - Standardization
  - Documentation
- **Certification Levels**
  - Gold
  - Silver
  - Bronze
- **Certification Drivers**
  - Initiator
  - Stakeholders
  - Data Governance



UNC HEALTH CARE SYSTEM

Community Knowledge & Culture

Knowledge Management

Community Engagement

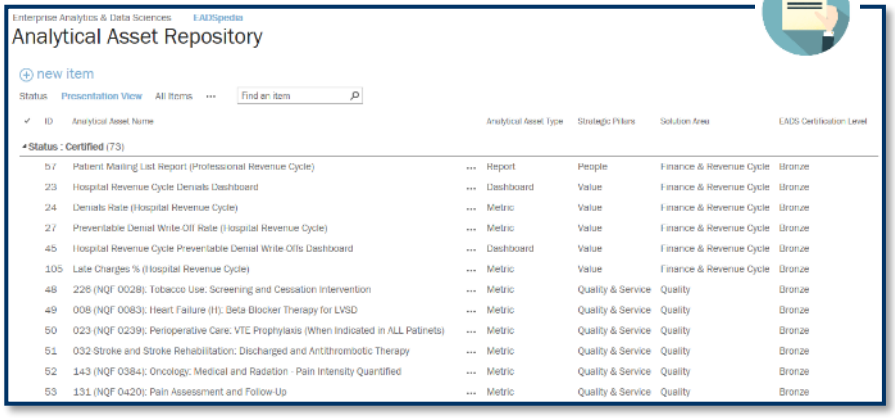
Welcome to EADSpedia!

EADSpedia is intended to be a **one-stop shop for curated information around data, reporting, & analytics at UNC Health Care**.
The success of EADSpedia relies on engagement from the entire analytics user community to provide content and share insights.
**Please be patient with us and forgive our dust as we work diligently to build out this resource!*

Certified Analytical Assets

Review the analytical assets that have been submitted for Certification and their current status.

If you have a request for an item to be certified, you can submit it here.



Analytics Community

Be a part of the community that helps drive content for EADSpedia through knowledge sharing, networking, and discussion.

If you're an analyst or ofter wear the "report-building" hat for your area, come be a part of the community!

Documentation

Search below or use this link to find information about a dashboard, report, Webl Universe, training, or something else.

Search...

ENTERPRISE ANALYTICS & DATA SCIENCES

HEALTH CARE

A Word on Ethics

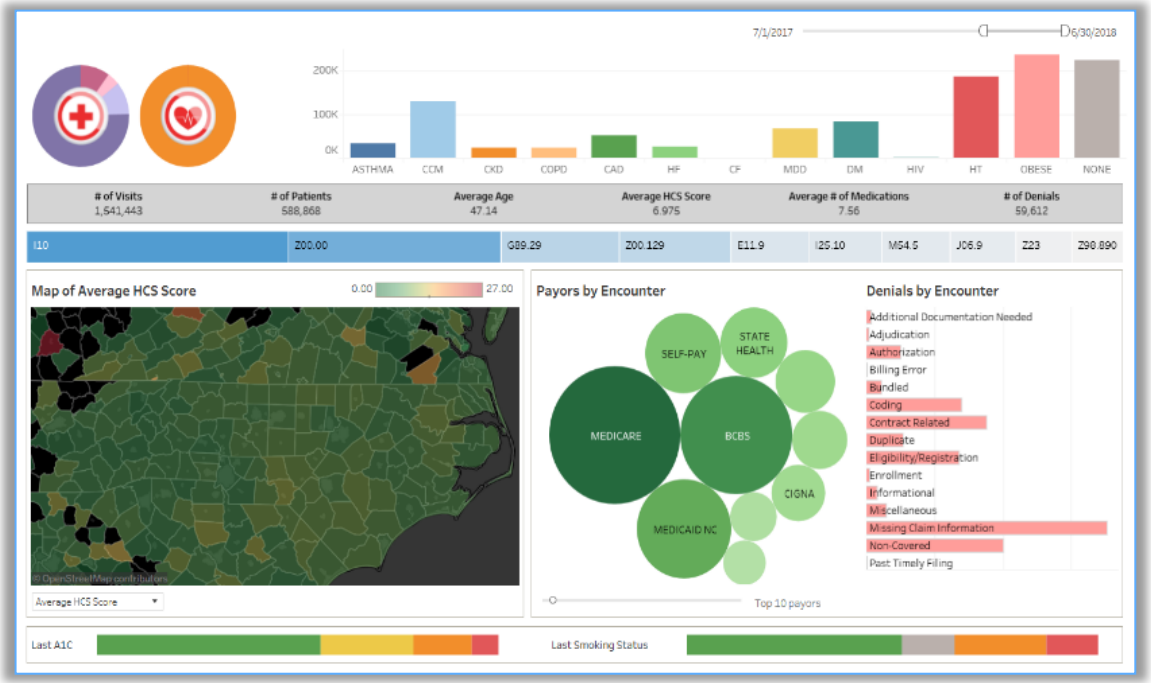
Data availability velocity > policy development velocity

Obtaining multidisciplinary perspectives on a big data opportunity up front is critical to ensuring the right questions are being asked

Questions to ask as soon as possible:

- Who are the stakeholders for this work?
- Who has granted consent for this work, when, and why?
- What are potentially negative outcomes of this work, and who would be impacted (e.g., sponsors vs. owners vs. stakeholders)?
- If we told patients and/or physicians we were doing this, what would they think?
- Where is the line between “improvement” and “research”?
- What controls can be used to mitigate potentially negative impacts of this?

Example: Utilization Modeling

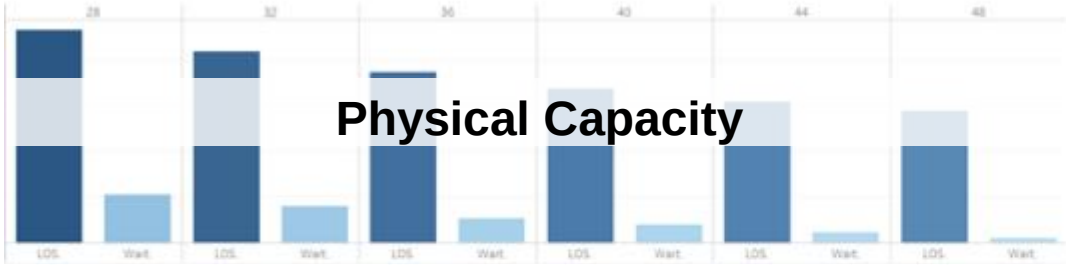


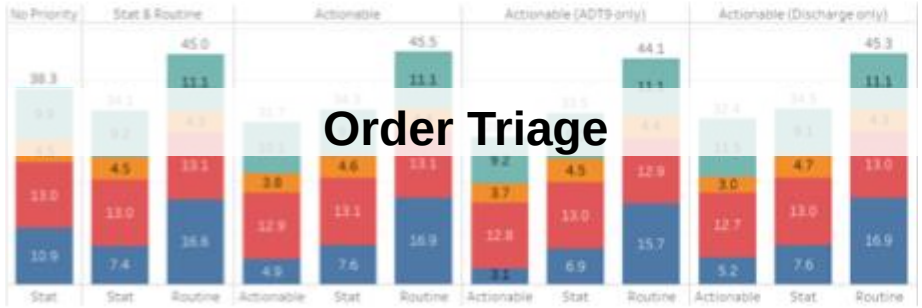
Common views of health care system utilization through shared semantic representation

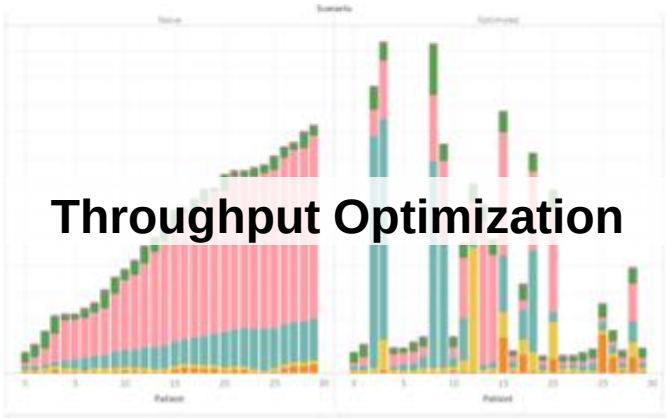
Requires data domains through a more sophisticated data strategy linked to data governance

Example: Patient Throughput Optimization

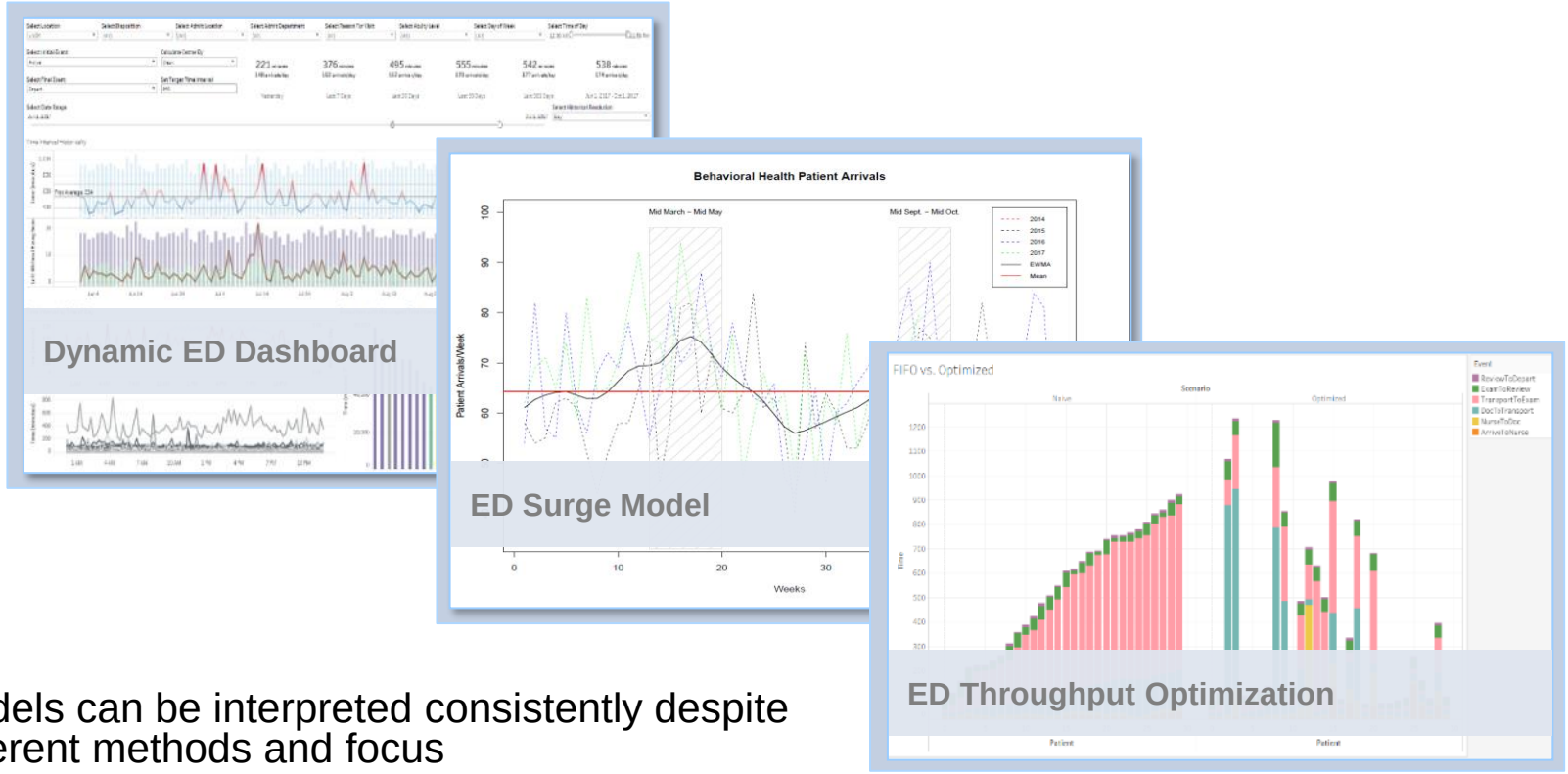
Using discrete event simulation to drive operational efficiencies







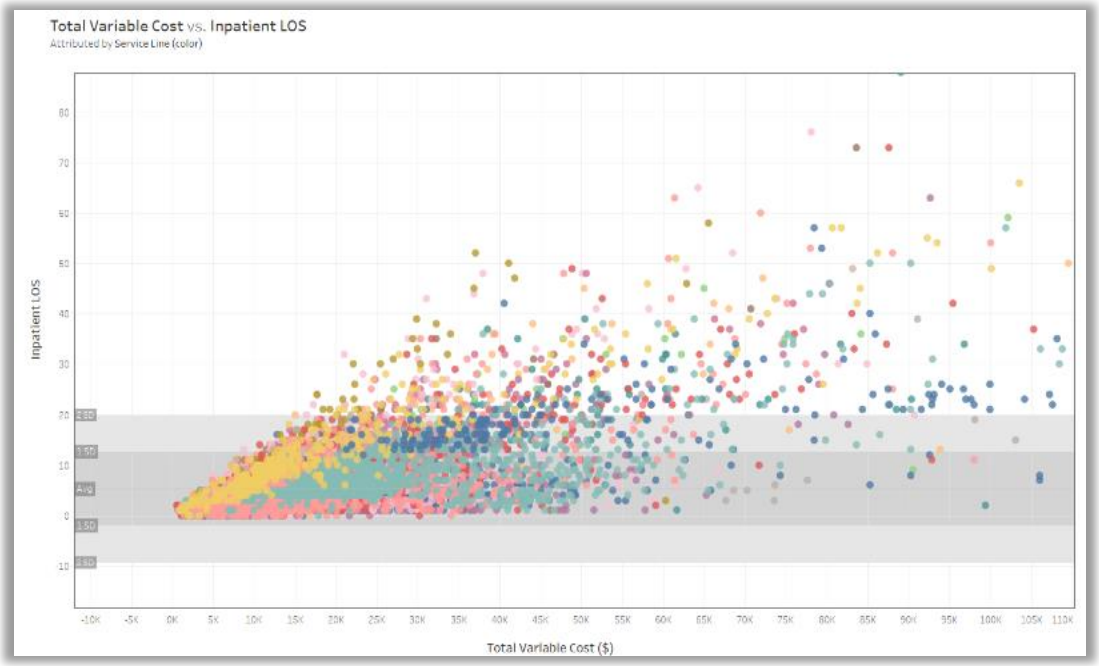
Example: Emergency Operations



Models can be interpreted consistently despite different methods and focus

Example: Care Variation

Mapping clinical context to big data



- Individual diseases have disease-specific models
- Performance is defined against system-level standards
- Analytics are used to normalize comparisons

Image courtesy of http://www.flickr.com/photos/strangrthancandy

DATA Issues
Storage
Structure
Timeliness
Semantics & Language
Validity
Reliability
Triage
Pedigree

INSIGHT Issues
Innovation
Health Outcomes
Profitability
Productivity
Translational Science
Customer Intimacy
Risk
Value

Image courtesy of http://www.flickr.com/photos/jdhancock

Summary

1. **Deriving value from big data is about a lot more than traditional "data management"**
   - What we know about our data →
     What we know from our data
2. **Data governance is one of the key domains required to effectively operationalize big data**
3. **Routine value creation from big data is dependent on growing and transitioning enterprise capabilities**
   - Reusable designs and assets
   - New process development
   - Business and clinical engagement
4. **Governance programs need to be ever mindful of ethics considerations**
   - New data use is sometimes unplanned data use

Questions

Jason Burke
System VP and Chief Analytics Officer

Email: Jason.Burke@unchealth.unc.edu
Twitter: @jaburke
LinkedIn: jasonburke