

The Changing Role of Data & Data Quality for Community Health Centers

David Hartzband, D.Sc.

Director, Technology Research

RCHN Community Health Foundation

Introduction

- The role of data at community health centers is changing
 - Some of these changes are predictable, but some of them are surprising
 - All of them will need to be planned for & acted on in order for CHCs to continue to perform productively, that is to continue to provide the highest quality care to their patients sustainably
- This presentation looks at these changes & tries to provide ways to prepare for them so that CHCs can leverage data to their best advantage

- The new data landscape
 - What data are we talking about?
 - What forces, trends, regulations are affecting this data?
 - What are some of the consequences & implications of this new landscape
- Amount of data
 - More than we think
- Data Appreciation & Awareness
 - What skills are needed to deal with all this data?
- Data Culture
 - How can organizations prepare?
- Data Quality & Quality Assessment
 - How can we know how good our data are, and what to do about gaps and weaknesses?
- The evolution of the data infrastructure
 - What changes will we need? In what timeframe?
- What to do about data

What data are we talking about?

- EHR
 - Clinical
 - Demographic
- PMS
 - Insurance
 - Add'l Demogr.
- Labs
- Registries
- Other Clinical
- Financial
- Claims
- Public Health
- Pop Health
- Macroeconomic
- Microeconomic
- Geolocational
- Social Determinants

& Other, Other, Other...

Influences on Data Change:

- Post-MIPS changing regulatory environment
 - APM: ACO, Medical Home, etc.
- Need for deeper analytics
 - Population Health, Predictive Analytics, Personalized Medicine, Genomics
- Integration of emerging tech
 - Big Data Analytics, Machine Learning, Natural Language Understanding
- Data Volume
 - Ever increasing (see slide on data volume)
- Data Diversity
 - More different types of data (as on previous slide)
- Storage models & products don't support data requirements
- Failing current infrastructure
 - Need for new infrastructure models
- Poor interoperability
 - No comment
- Non-support of actual clinical & admin workflows
 - Current apps (EHR, PM etc.) do not support data or workflow requirements well
- Organizational & cultural impediments for HIT evolution
 - Many examples of technological & organizational "conservatism" that hold back overall system (technical & non-technical) improvements

Some Possible Consequences:

- HIT infrastructure & Apps become impediments to productive use
 - Misalignment and failure to meet short-range functional and operational requirements make infrastructure and applications much less effective long term
- Measurable quality decline
 - Follows from lack of effective infrastructure and apps
- Realized revenue declines
 - As above
- More time, effort & expense required to maintain & operate system(s)
 - Maintenance effort takes time & resources (\$) away from actual work of the CHC
- More vendor dependence as system components require more attention
 - Follows as a direct consequence

- Both the existing infrastructure that supports HIT & the apps that provide functionality can only be stretched so far before they become problematic
 - Network speed and reliability need to improve
 - Storage capacity needs to substantially improve
 - Interoperability needs to improve considerably beyond what the ONC has specified
- Apps need to align with actual workflows to become both more usable & especially more useful
- Much deeper and broader analytics need to be provided in order to support new requirements for population health, personalized medicine, etc.
- The role and capabilities of emerging technologies such as AI and machine learning need to be determined & integrated

- Data Culture is the social, organizational & personal context that includes both data appreciation and data awareness as major principles
- An organization with a positive data culture has aligned its use of data with organizational goals & strategy
- Data culture starts at the top and permeates all aspects & levels of the organization
- Data cultures are collaborative, especially with respect to the analysis of data & interpretation of results
- Data cultures are **NOT** technology based, but based on data appreciation & awareness

Data, Data & More Data

- Most healthcare organizations currently have 1-5 TBs (1TB=10¹² bytes) data under management including EHR, other clinical, administrative & financial data
- Some have substantially more, up to 40-45 PBs (1PB=10¹⁵ bytes)(Kaiser Permanente)
- Consensus is for healthcare data to grow 40%-60% per year, *i.e.* to double every two years.
- This means that a typical healthcare organizations will have ~30 TBs of data in 5 years & many may have 100+ TBs in this time.
 - Note that the print collection of the Library of Congress (32M discrete books, etc) is in the range of 10-15 TBs while the entire collection including image, audio & video is about 50PBs.
- The management & use of this data becomes problematic with current (& near-future) storage & application technologies requiring a move to new architectures.
- We'll return to this when we discuss infrastructure

- Most people are not mathematicians or data scientists! Data appreciation starts not with data, but with what data tell us:
 - Stories – large scale analysis is a story about how something works, about relationships among things & influences, this is more interesting* than:
$$\alpha = \bar{y} - \beta \bar{x}, \beta = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
 - Solutions to real-world problems
 - Games
- These approaches foster the appreciation of data as important in our work & personal lives
- Beyond this, people who work with data at all levels must develop data awareness
 - This is the ability to know: 1) what data are available; 2) what data are relevant to a specific problem; 3) in general, how to analyze that data, & 4) how to interpret the results of the analysis
- This awareness is essential for being able to use data in effective & productive ways

*To most, but not all people ☹ (general linear regression)

Data Culture Dos & Don'ts

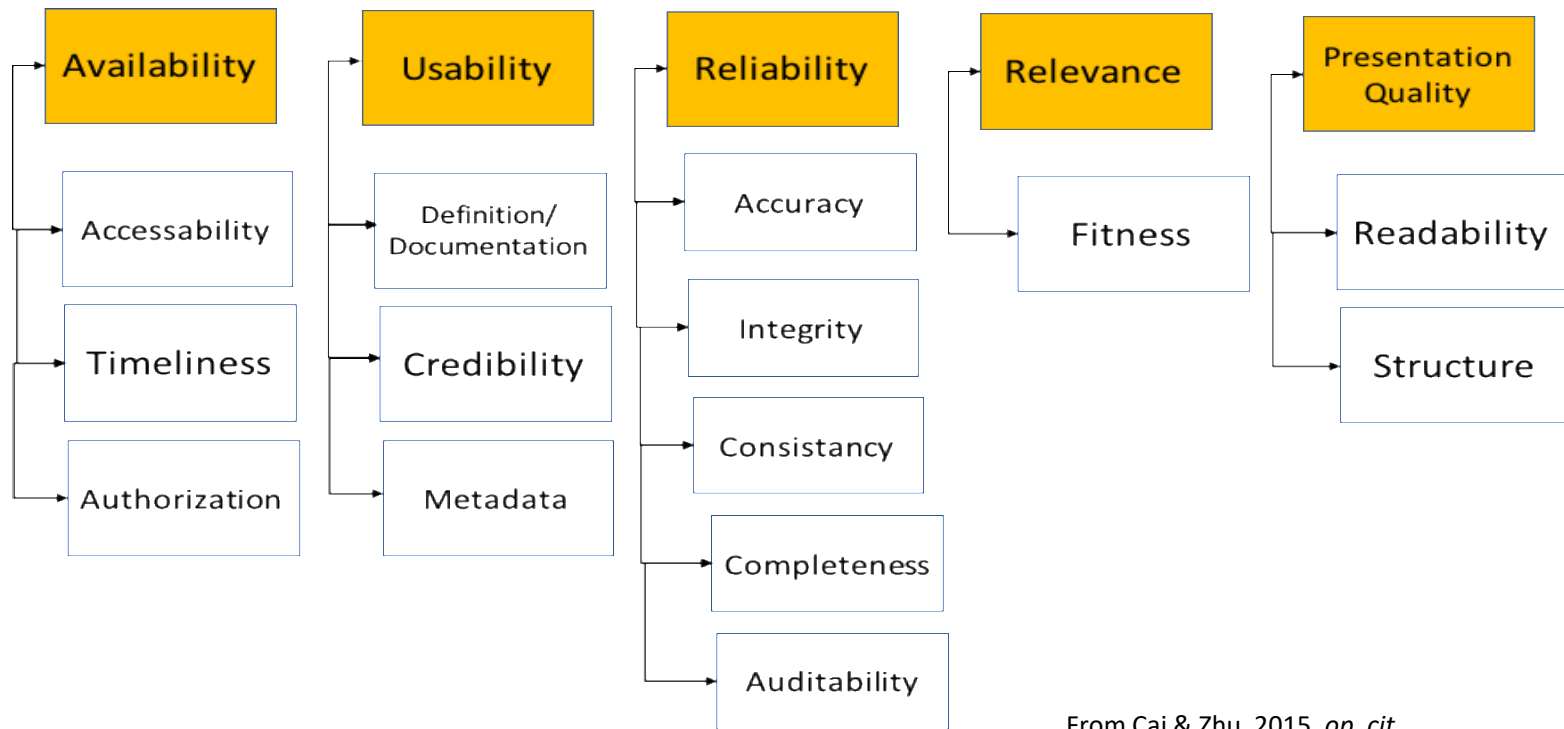
Do	Don't
Promote data awareness throughout the organization, have program & project champions emerge	Make development of measurements or Key Performance Indicators (KPIs) the first priority
Elicit executive (CEO) support & participation in data project planning & projects	Base success of projects on evaluation of KPIs
Work to align use of data (reporting, analytics...) with organizational strategy	Commit to a specific technology or application before the strategic issues are understood
Use cross group collaboration to plan & interpret analytic results	Appoint program & project leads based on organizational hierarchy
Ensure that technology can support data projects (acquisition, storage, analysis...)	Delegate the development of a data culture to an outside entity (except as a consultant to the internal lead)
Evaluate success based on accomplishment of strategic goals	

- Data is considered to be of good quality if it is “fit” for its intended use in clinical treatment & operations decision making
- Data Quality has been studied literally for centuries. Amazon lists >7000 current books on the topic with 92 listed just for healthcare data. A Google search on the topic provided 217,000,000 returns
- The five dimensions of good quality data have recently been described by Cai & Zhu* (see next slide)
- Quality is generally assessed by measuring aspects of these five characteristics
 - The most effective way to do these measures is by comparison of data values from different sources derived from the same data set
- Hartzband & Jacobs** describe an assessment that compares data values from separate data sources derived from the same data set. - This “Level-Up” exercise is further described on slide 14

*Cai, L and Zhu, Y 2015 The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, 14: 2, pp. 1-10, DOI: <http://dx.doi.org/10.5334/dsj-2015-002>

** Hartzband, D. & F. Jacobs. 2016. Deployment of Analytics into the Healthcare Safety-Net: Lessons Learned. Online Journal of Public Health Informatics * ISSN 1947-2579 * <http://ojphi.org> * 5(3):e203, 2016

Dimensions of Data Quality



From Cai & Zhu. 2015. *op. cit.*

- Each of the five primary dimensions are immediately recognized in (healthcare) data sets
- Secondary characteristics can be important in judging data quality & potentially fixing issues with quality

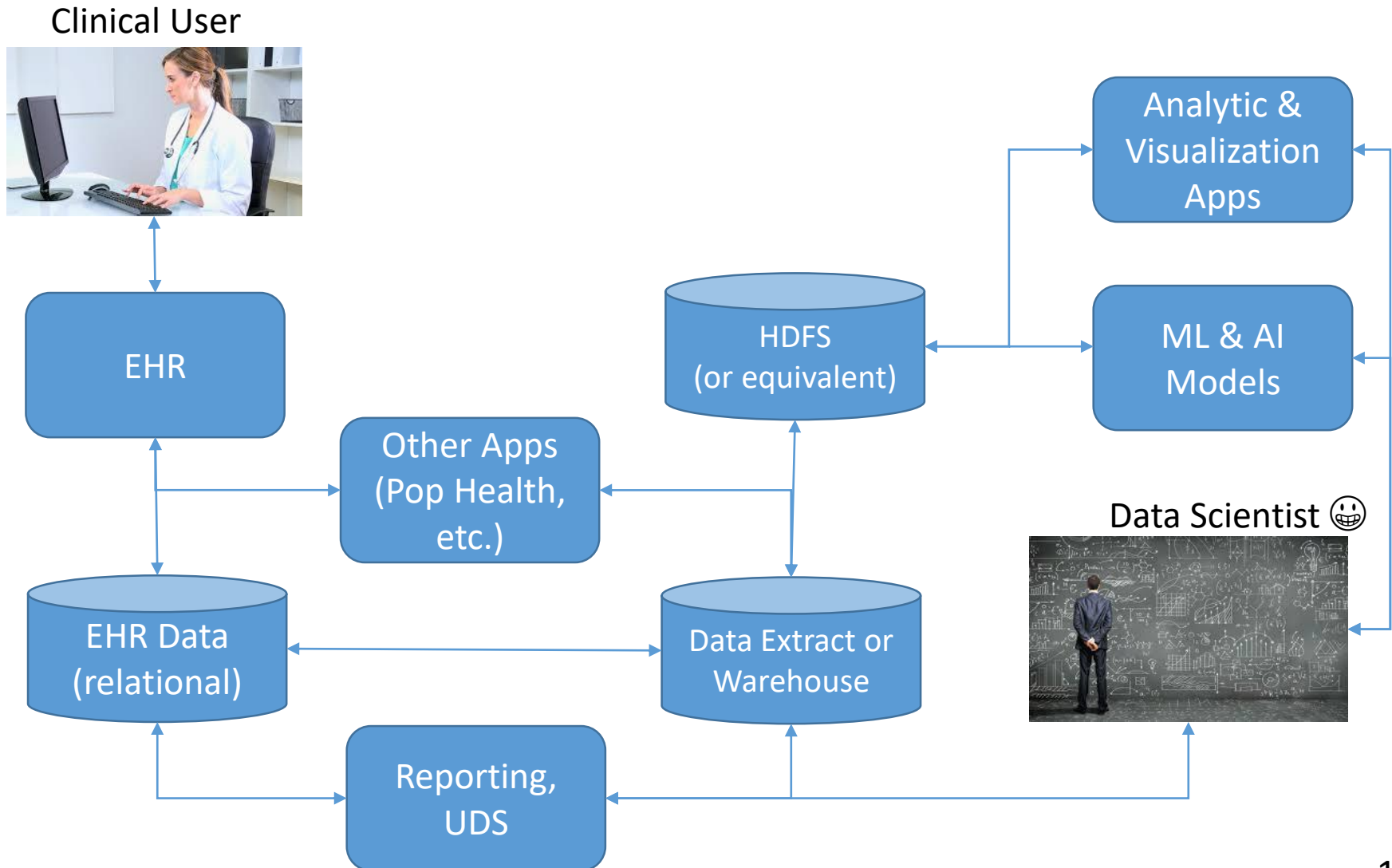
- Level-Up is an analytic technique that compares data from different data sources that are derived from the same base data.
- In its initial form, it compared data from an EHR's underlying database with data from an analytic data source that had been loaded from the EHR.
- The comparison with EHR data can be made from the EHR's database to a variety of secondary data sources including: 1) a data extract or data warehouse derived from the EHR data; 2) claims or other billing data; 3) data derived from clinical notes, data derived from external sources (site logs, registries etc.)
- The comparison is done by the applications of standardized queries (usually in SQL) to both data sources & then comparing results
 - Often the comparison can be made with UDS results
- Differences in results may be the result of anomalies in the data & can be characterized into a small number of categories & evaluated to determine if there is a “fix” for the issue (see next slide)

Data Issue	Possible Solution	Comments
1. Deviation from standard definitions	Remediate using standard definitions	UDS definitions used for all reports & analysis
2. Missing &/or omitted data	Attempt to recover data from other sources	Claims data, provider notes, site logs etc.
3. Incorrectly entered data	As in 2.	Development of data awareness may help
4. Data values not entered into searchable field	Natural language application may assist	Often from imported data
5. Errors related to structure/complexity of EHR	Simplify workflows for: data capture, work with vendor to improve EHR	Easier to correct data at capture than at clinical use or analysis
6. Errors related to migration of EHR system	Maintain original database for report gen & quality	Work with vendor(s) to ensure correct migration
7. Errors related to cultural or organizational bias	Work to uncover bias, process to advise staff	Progress must be reviewed

- We want all of the data used by the CHC to be of the highest quality possible
- This is a goal that we are constantly working towards, but...
- We **know** that some of our data is not as good as we'd like or need it to be
 - CHCs don't control all of the data they use, but being aware of the quality of external data helps us to understand its limitations & effect on analytic results
- Current approaches to improving data quality rely on governance techniques, especially during data collection and entry, but these techniques take time & resources & do not apply to existing data sets
- Quality effort going forward will focus on these methods to the extent possible, but will also emphasize working backward from anomalies in analytic results to identify and "fix" data quality issues.
 - Once identified by backward analysis, steps can be taken to prevent the specific error type in the future
- Data awareness will allow results of analysis to be interpreted even if there are known or unknown quality issues

- The HIT infrastructure supporting many health care institutions today, including CHCs, is not up to supporting the necessary reporting & analytic function over the next 3-5 years
 - Substantially more data and more different types of data will need to be managed & used
 - New types of applications (both architectures & function) will need to be supported
 - New technologies will become increasingly important (natural language processing, advanced analytics, AI-ML, etc.)
 - Interoperability needs to be greatly improved
- Current relational (DB) and server-based infrastructure is less & less able to support this evolution
- A new infrastructure that emphasizes massive data use, a much larger user base and new technologies supported natively needs to be developed & adopted
- We have about 5 years to do this
 - By that time data volume & variety as well as analytic needs will force this

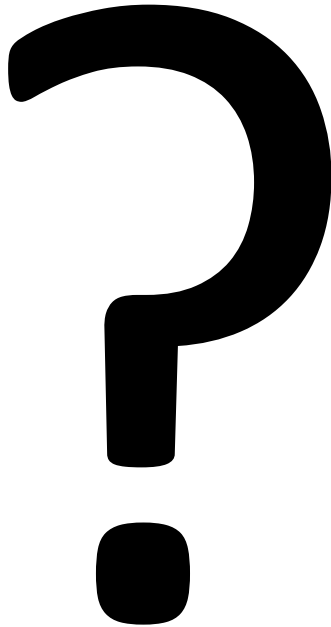
One Possible Interim Picture



What To Do...

People who generate and use data at the CHC - starting with C-Suite - should:

- Facilitate and build a positive data culture for the CHC
 - Be aware of trends and implications of the “new data landscape”
 - Be aware of the totality of data used by the CHC both internally produced & especially externally sourced
- Understand the measurement and implications of data quality for internal & external data
 - Have a program to measure data quality & to ameliorate those issues that can be addressed
- Understand the role and potential limitations of the data infrastructure at the CHC
 - Have a plan to evolve the infrastructure as necessary
- Remember – Culture is not enough, culture needs to drive iteration & change



David Hartzband, D.Sc. (math)

dhartzband@rchnfoundation.org

dhartz@mit.edu

617.501.4611 (m)

617-401-2503 (CIC office)

617-714-3964 (fax)

www.rchnfoundation.org

LinkedIn, Twitter:dhartzband

blog:<http://posttechnical.blogspot.com/>