Article published online: 2021-10-11

Perspective: Data Governance—Making Data Great

Keith F. Woeltje[1,2]

[1] Institute for Informatics, BJC HealthCare, St. Louis, Missouri, United States
[2] Washington University School of Medicine, St. Louis, Missouri

Address for correspondence Keith F. Woeltje, MD, PhD, BJC HealthCare, 8300 Eager Road, Suite 400A, St. Louis, MO 63144, United States (e-mail: keith.woeltje@bjc.org).

ACI Open 2021;5:e80–e83.

Introduction

Health care delivery in the United States is undergoing massive changes. Although the pace at the local level can often be imperceptible, the overall momentum is relentless. Multiple factors are driving these changes—the rate of cost increases for care in the U.S. is unsustainable, the outcomes of care are worse than other countries that spend less, and a large portion of the population does not have real access to ongoing care, leading to enormous societal costs. On the health care delivery side, there has been a shift from stand-alone care delivery through stages of collaborative and value-driven health care toward a more accountable care/population management approach (►Fig. 1).

Politicians, policy experts, and health care administrators long recognized that changing models of health care would require significant amounts of data. This was a major factor in the founding of the Office of the National Coordinator for Health Information Technology (ONC), and later the Health Information Technology for Economic and Clinical Health (HITECH) Act which brought about the Meaningful Use program.[1] With the increased use of electronic health records (EHRs), the availability of electronic data in health care has risen dramatically.

Organizations increasingly see their data as a strategic asset. To most effectively leverage this asset, data from diverse sources (e.g., revenue cycle, supply chain, quality measures) must be combined for analysis. The use of data from diverse sources has always occurred, but in many organizations, this was often accomplished through establishment of locally curated data collections. Unsurprisingly, organizational leaders found they could get different answers for the same question depending on which group they asked.

Because data definitions can be quite nuanced, a team can become reluctant to share data with the rest of the enterprise. There is a concern that others would not understand "my data." This can then lead to users on other teams to be creative in finding sources for the data they need. By acquiring data sets from diverse acquaintances across the organization, analysts may be using data with somewhat unclear provenance (often many steps away from the original sources) for important organizational reports that drive executive decision making. This can exacerbate the issue of getting different answers from different groups.

Technical approaches to improving the situation include establishing a "single source of truth" for enterprise data, for example, via an enterprise data warehouse or a data federation approach. This can help with some of the notions of "my data" versus "our data" in an organization, but does not in and of itself ensure that the data are well characterized and of high quality. As health care organizations struggle to use their data effectively, they realize they need an approach to addressing these issues. Whether they formally label it as such, what the organizations are grappling with is data governance.[2,3]

Data Governance Program

BJC HealthCare is a 15-hospital health system covering portions of eastern Missouri and southwest Illinois, with headquarters in St. Louis, MO. BJC is partnered with the Washington University School of Medicine (WUSM). Established in 1994, for most of its history BJC largely operated as a federation of somewhat independent hospitals which suffered all the data woes previously mentioned. The oncoming changes in care delivery and financing described above has led BJC to work more as an integrated system. Initial efforts at data governance were prompted by a desire to consolidate multiple repositories of clinical data. Shortly thereafter, the efforts were spurred on by a decision by BJC and WUSM to implement a shared EHR that would replace the multiple hospital and ambulatory EHRs used by both organizations. The early focus was on defining data domains needed for the EHR consolidation, including a master-data management

received
July 16, 2020
accepted
April 20, 2021

DOI https://doi.org/10.1055/s-0041-1730383.
ISSN 2566-9346.

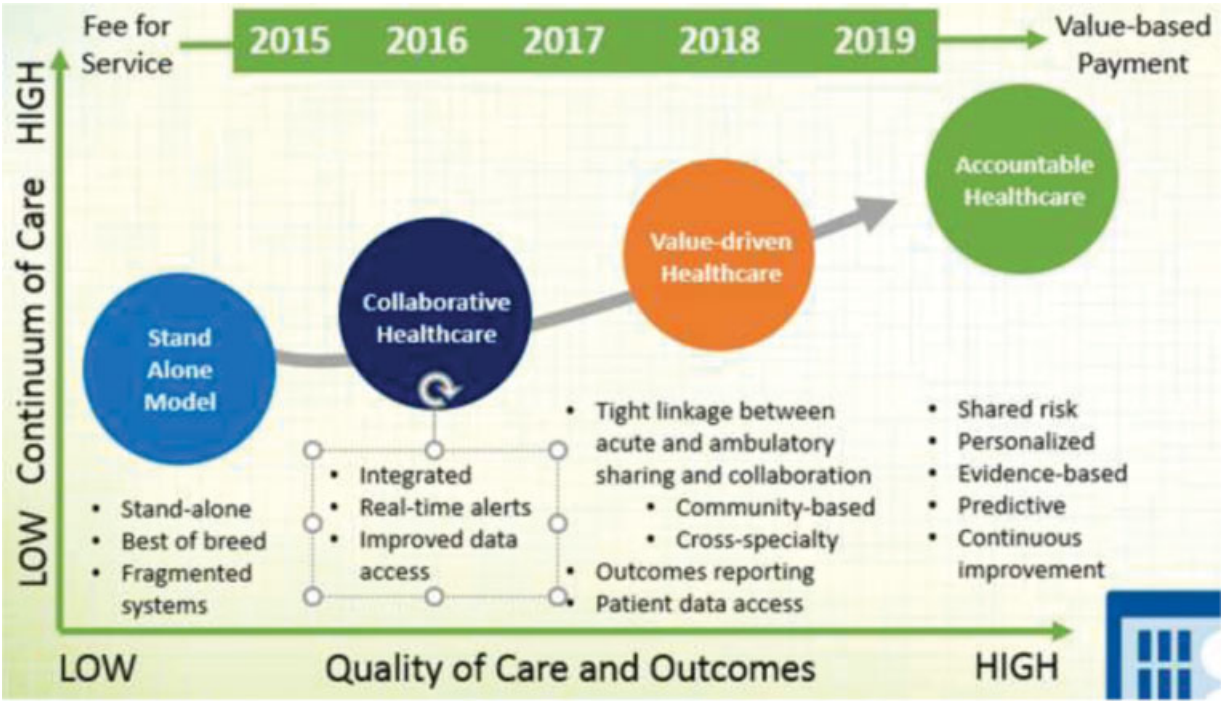

Fig. 1 Evolution of care delivery.

(MDM) program to establish a single information on providers/staff that the EHR could rely on. Executive sponsors, data trustees, and data stewards were named to further these efforts.

The program did not really move much beyond data documentation management (e.g., data glossary, data dictionary) related to data for the EHR during the implementation time period. Admittedly, this was a complicated set of related use cases. Progress was also stalled due to a serious illness within the program leadership. Following the EHR implementation there was a desire to reinvigorate the program, with a focus on establishing organizational data governance policies.

Framework for Data Governance

As part of the efforts to reenergize the program, we recognized that we were hampered by the lack of a conceptual framework for our data governance efforts. We had no common reference for describing the work we were doing. Working with consulting partners, we established a basic framework entitled the "Data and Information Lifecyle

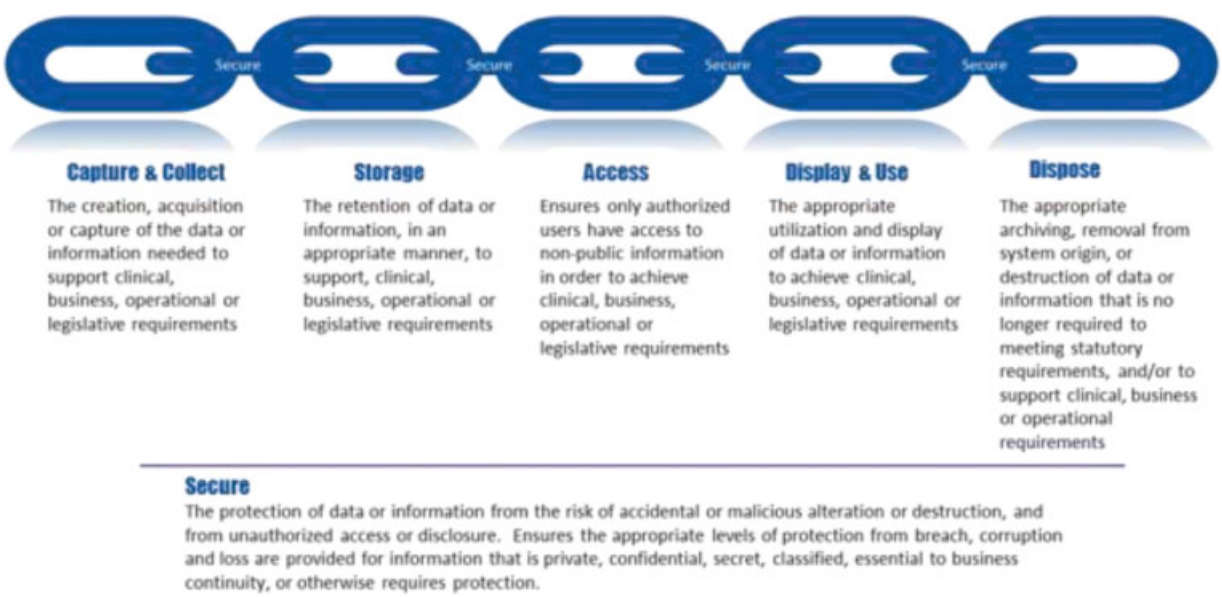

Fig. 2 BJC Data and Information Lifecyle framework.

Framework." This framework consisted of six key phases that data pass through (►Fig. 2).

Capture and Collect

The first step in any story about data is the capture of those data. For the staff involved in this key step, understanding the downstream consequences of their actions may help promote attention to detail and data quality. Organizational efforts here can be focused on automation and eliminating to the degree possible error-prone manual entry steps.

Storage

This phase includes all data storage, including primary source systems and consolidated data repositories. Also included in this phase would be the tools and processes for moving and transforming data between storage locations.

Access

To ensure appropriate data use across the enterprise it may be practical to readily provide appropriate role-based access to the organization's enterprise data repositor. This ensures that analysts and other users are not forced to seek out alternative data sources to do their jobs.

Display and Use

This is the key element in the data and information lifecycle. It ensures that the organization's data can be used effectively for analysis and subsequent decision making that all these efforts are undertaken. Just as there is value in providing a consistent data platform, there is also value in defining a standard set of analytics and reporting tools for the organization.

Dispose

Although the cost of storing data continues to fall, this element is a reminder that there are limits to how long some data are useful. An organization must decide what those limits are, and then establish policies for appropriate disposal of data at the end of their useful life. Regulations may also define how long certain data must be retained. This element also covers long-term archiving of data that are no longer needed for operational use, but which need to be kept for legal and regulatory reasons.

Secure

In the first version of our framework, security issues were called out in each element. After many discussions we determined that it really should be called out as its own element to emphasize the importance of security concerns in the modern health care data environment.

The Patient's Data Story

To further help provide a context for this framework, we created a representation of how these elements applied to a patient's story, in the context of the three standard elements of any change model: people, process, and technology (►Fig. 3). This representation helps take a conceptual framework and relate it to roles and activities that nontechnical staff may be more familiar with. It is an illustrative tool that allows the foundation for a common vocabulary and shared meaning for data discussions and decisions to be cultivated.

Policies

The data governance program also recognized the need for organizational policies to support and drive the discipline of governance. Establishing a reference framework provided a way

Fig. 3 Data and information lifecycle in the context of the patient story.

to assess the current state. In that assessment we were not surprised to find that there were hundreds of system and local hospital policies addressing the major elements of data and data governance. Gaps in policy coverage were also identified.

Rather than attempt to reconcile myriad disparate policies that had various levels of granularity, we chose to start with a top-level data governance policy, and then define additional policies using the new framework. These new policies would supersede existing policies. Our highest-level data policy simply establishes the data and information lifecycle framework as the organizational standard, and sets very high level organizational expectations. As an example:

The key data governing rules for the capture and collect phase must include:

- Documentation standards that include the minimum clinical and business data sets and elements to be captured and collected as well as timeframes for collection.
- Annual training and communication programs that informed employees and providers of their responsibilities related to minimum data collection standards, patient's data rights, and protecting the organization's information. We are bolstering our culture of data accountability.
- Systems are designed to ensure data collected is correct via hardwired standards, system designs, and workflow to pay attention to simplified processes that drive standards.

Subordinate policies have been written in some key areas. For example, BJC has an active information security group which had been in the process of revising system policies. That policy development was brought under the data governance umbrella. Additional subordinate policies are in development, driven by organizational needs.

Data Governance Model

As mentioned, during our initial data governance efforts we had established executive sponsors, data trustees, and data stewards. An executive sponsor is a member of the C-suite and provided overall budgetary and strategic approval of the program initiatives. A data trustee is a senior organizational leader who acted as the owner of data and could make decisions about data use within a particular data domain. Examples of data domains include "provider," "patient," "medication," "procedure," and "supply." For many data domains there are more than one trustee; for data domains used within our shared EHR we have both BJC and WUSM trustees. A data steward is typically a more front-line staff member who works on a daily basis with data from their particular domain in a technical role.

Our initial data governance committee was simply the aggregate of the data trustees. That group proved to be too large to be an effective decision-making body; in practice decisions were made by the director heading the data governance program (who reported to the chief medical informatics officer [CMIO]) and simply presented to that group for vetting. The loss of the director in that role was one catalyst for prompting a refresh of the program. Technical aspects of data governance (e.g., maintaining data dictionaries) and MDM were moved into the system information technology (IT) organization. There was strong organizational consensus that data governance per se was not an IT function, but an organizational operational function.[2] Our data governance committee was reconstituted as a much smaller group at the vice president/executive director level representing key organizational constituencies (e.g., finance, legal, IT, quality). The committee is co-chaired by the system CMIO and the Chief Supply Officer. The data governance committee in turn is overseen by a group of our system's senior executive leadership, including the Chief Clinical Officer, Chief Information Officer, and Chief Financial Officer.

The purpose of the data governance committee includes:

- Develop enterprise strong policies and procedures that describe the ways to manage data.
- Define the rules of engagement for data.
- Promote an enterprise way of thinking about data (i.e., Data & Information Lifecycle) and increase communications about data and data activities.
- Identify the resources who manage data and hold them accountable to each other.
- Ensure alignment between BJC and WU related to data governance.
- Ensure that data governance subcommittee efforts support BJC clinical and business objectives/priorities.

Conclusion

Our organization has struggled to use data effectively as a strategic asset, and to effectively describe and integrate data from different groups within the system. This struggle is not unique to BJC. To address these issues, we have developed a data governance program that is led by the business with cross-functional support from our technical teams. Initial efforts were limited in scope, but very effective and instrumental in an enterprise EHR launch. More recently there has been more energy in broadening the scope and effectiveness of the program. Establishing a clear conceptual framework for our efforts has been incredibly helpful. Undoubtedly our program will continually mature; for now we believe we have established a solid foundation for this future evolution.

Conflict of Interest
None declared.

References

1 Wachter R. The Digital Doctor. New York: McGraw Hill Education; 2015
2 Healthcare Information and Management Systems Society (HIMSS) Practical Steps to Data Governance. 2016. Accessed March 11, 2019 at: https://www.himss.org/ResourceLibrary/GenResourceDetail.aspx?ItemNumber=47974
3 Hripcsak G, Bloomrosen M, FlatelyBrennan P, et al. Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting. J Am Med Inform Assoc 2014;21(02):204–211