I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

Data Governance: A Quality Imperative in the Era of Big Data, Open Data, and Beyond

BARBARA L. COHN*

I. INTRODUCTION

The world is awash in data, nuanced and characterized by variations as diverse as dialects. Data is everywhere, embedded within the "fabric of our daily lives"[1] in ways both imperceptible and conspicuous. Technology is changing the way data is generated, collected, maintained, and utilized. In the era of big data,[2] there is a penchant to ascribe value to the sheer velocity, volume, and variety of data captured, stored, and generated. Realizing meaningful value from data, however, is much more complex than the speed, capacity, or the data itself. While data is a catalyst for innovation, data governance is a catalyst for quality, and value is derived from well-governed quality data. Relevant, timely, consistent, reliable, and

* Barbara L. Cohn is the New York State Chief Data Officer, Office of Information Technology Services. The views expressed in this essay are those of the author and do not represent the views of, and should not be attributed to, the NYS Office of Information Technology Services.

[1] Executive Office of the President, "Big Data: Seizing Opportunities, Preserving Values," May 2014, at 1, http://www.whitehouse.gov/ sites/default/files/docs/ big_data_privacy_report_5.1.14_final_print.pdf.

[2] Ibid., 4. The common three V's of "Big Data" were defined by Doug Laney in a 2001 report entitled: *3D Data Management: Controlling Data Volume, Velocity, and Variety*, http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Although Doug Laney's definition has been the most oft cited definition to date, a literature review of the term 'big data' will reveal varied expansions on the original 3Vs (e.g., http://dataconomy.com/seven-vs-big-data/).

accurate data is an expectation and not achieved serendipitously. Data governance institutionalizes the processes and protocols by which organizations harness the value of their data to advance business goals. Its importance in supporting and informing an organization's decisions cannot be overstated.

II. QUALITY DATA IS AN IMPERATIVE THAT DATA GOVERNANCE MUST SUPPORT

Big data technology is driving exponential growth, consumption, and reliance upon data and the information derived from data. "This explosion of data from web-enabled appliances, wearable technology, and advanced sensors to monitor everything from vital signs to energy use to a jogger's running speed"[3] makes quality data an imperative that data governance must support. Maximizing the derivative value of data is foundational to data governance. It is pivotal to innovation and the generation of new insights – social, economic, scientific. At its most basic application, data is a convenience. At its most significant, it is of a critical nature.

In this essay I will describe what data governance is and the general principles that should guide it. I will sketch some case examples of effective governance, as well as highlight the impetus for the growing importance of data governance in the era of big data irrespective of context or setting. It is important to note, however, that data governance is neither new nor limited to big data.[4] On the contrary, although often raised in the context of big data, data governance is a necessity regardless of size, regardless of format, regardless of medium.[5] This essay is not to advocate a specific governance framework. Rather, it is to underscore the relevance and

---

[3] Ibid., 4.

[4] Ancient infographics have also contributed to maximizing value from data in terms of greater understanding. An effective governance framework provides a forum for reconciliation, decision, and communication. As Dr. Johanna Kieniewicz, lead curator of a show at the British Library which highlighted the history of data visualization, noted, "[w]e are in an era of big data, but there was also an explosion in data back then, particularly related to vital statistics and climate. They too were trying to reconcile how to work with the information and communicate it to the public." Carey Dunne, "16 of Science's Best Infographics from Ancient Greece to Today," *Fast Company*, March 3, 2014, http://www.fastcodesign.com/3026917/16-of-sciences-best-infographics-from-ancient-greece-to-today.

[5] In addition to being represented by numbers, data is also represented by graphics, images, visualizations, etc. Mark Mosley and Michael Brackett, eds. *DAMA Guide to the Data Management Body of Knowledge*. (Technics Publications, LLC, 2009), 2-3.

significance of having a set of core principles that establish: (i) a strong foundation that can be built upon; (ii) a solid infrastructure within which to maximize performance and confront and leverage challenges; (iii) an effective framework that creates opportunity, business value, and insight; and (iv) an organizational structure that can evolve in a dynamic environment – all the while maintaining basic precepts of sound data practice. Meaningful governance requires a sustained and broad-based effort to effectuate positive outcomes and change. This essay will also address the need to take account of both the algebraic and the human algorithm,[6] and why a keen awareness of how data is transformed into actionable knowledge[7] is integral to understanding the complex nature of the "data" in data governance.

III. WHAT IS DATA GOVERNANCE?

Narrowly defined, data governance is a framework which formalizes the roles, functions, and procedures within which an organization's data is well managed and enabled as a strategic asset.

Apropos, in the era of predictive analytics, it is history[8] that provides the greatest insight into the essential elements of effective and sustainable governance. An exemplary template is the United States Constitution. The Constitution sets forth core principles that lay the foundation for a sustainable governance structure and an adaptable framework for a new nation. The Framers "sought not only to address the specific challenges facing the nation during their lifetimes, but to establish the foundational principles that would sustain and guide the nation into an always uncertain future."[9] The Framers understood the need for the document to be a living charter,

---

[6] Mark Little, "Finding the Wisdom in the Crowd," *Nieman Reports* (Summer 2012), http://www.nieman.harvard.edu/reports/article/102766/Finding-the-Wisdom-in-the-Crowd.aspx.

[7] David Weinberger, "The Problem with the Data-Information-Knowledge-Wisdom Hierarchy," *Harvard Business Review*, February 2, 2010, http://blogs.hbr.org/2010/02/data-is-to-info-as-info-is-not/.

[8] "Data that is centuries old from collections like ours is now being used to inform cutting edge science…The British Meteorology Office still uses information from log books on East India Company clipper ships to test their climate models, the idea being that to understand weather patterns of the present, they need to understand patterns of the past." Dunne, see note 4.

[9] Geoffrey R. Stone and William P. Marshall, "The Framers' Constitution: Toward a Theory of Principled Constitutionalism," *American Constitution Society for Law and Public Policy*, (September 2011), 1-2. https://www.acslaw.org/sites/default/files/Stone_Marshall_-_The_Framers_Constitution_Issue_Brief_1.pdf.

the broad principles of which would be enriched and given concrete meaning over time by the judgment and experience of future generations.[10] They were deliberate to ensure that key principles would themselves be preserved and remain constant over time. They also astutely recognized and understood that "the application of those principles must evolve as society changes and as experience informs our understanding."[11]

Broadly applied, data governance provides a schema for collaboration involving an organization's people, processes, and technology[12]…*and data*. It sets forth an organization's vision, as well as its policies, protocols, and standards in support of attaining maximum value from data. It assures filters and drives compliance to guard against "garbage in, garbage out." It provides a forum for decision-making based upon trusted data. At its core, "[a] data governance program is not an application that can be purchased, installed, and implemented with a specified end date, but a process that, over time, affects the culture and the way an organization conducts business."[13] It is a continuous quality improvement and performance process.[14] Data governance thus requires a holistic understanding of an organization's business, the data at its disposal, and their interrelationships. The value of IT and business working together cannot be over-emphasized, and its centrality is one of the most important components of an effective data governance structure. It is a prerequisite. Their integration "establishes confidence and credibility in the enterprise's information."[15]

"Data governance is a relatively new term, and many organizations continue to pioneer new approaches."[16] Notwithstanding the uniqueness of organizations, there is a commonality of purpose that is shared, as well as common characteristics based upon basic tenets and

---

[10] Ibid., 2.

[11] Ibid., 2.

[12] Martha Dember, "7 Stages for Effective Data Governance," *Architecture & Governance Magazine* 2, no. 4, http://www.architectureandgovernance.com/content/7-stages-effective-data-governance.

[13] Ibid.

[14] See Mosley and Brackett, *DAMA Guide to the Data Management Body of Knowledge*, 38.

[15] See Dember, note 12.

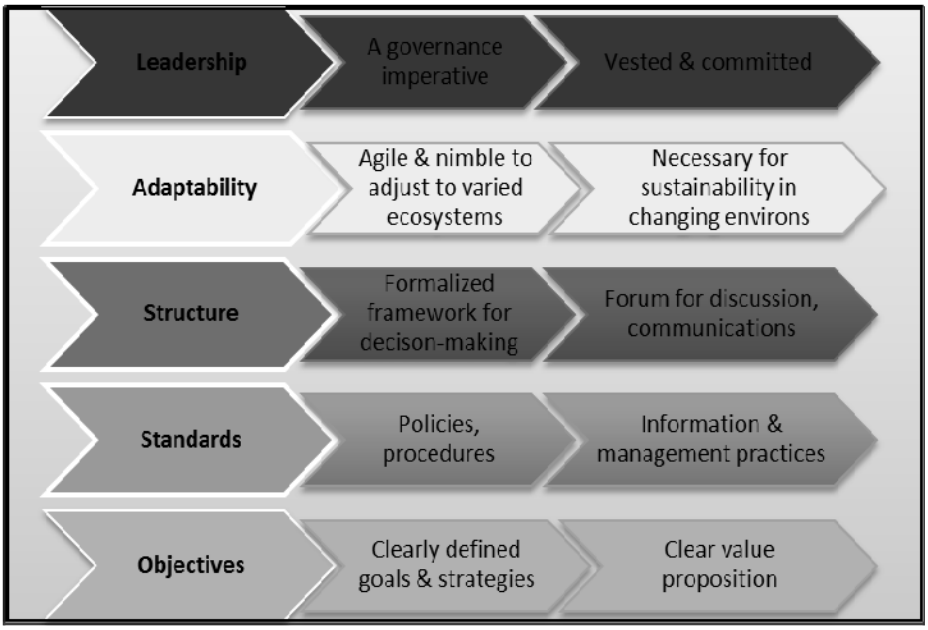
[16] See Mosley and Brackett, note 14.

principles.[17] The most effective data governance frameworks include a set of core elements, highlighted in Figure 1, which establish a solid foundation and common denominators in a global society where the norm is constant change. The five core elements of an effective governance framework are captured in the handy mnemonic LASSO - Leadership; Adaptability; Structure; Standards; Objectives.

Figure 1: Data Governance Core Elements

- Leadership

Strong executive leadership is one of the most critical elements for success. A vested and committed executive is an essential component to motivate and drive the highest levels of quality and performance. It is a governance imperative. Without strong leadership, an organization will drift and data governance will fail.



- Adaptability

Adaptability is a necessity for sustainability in meeting both internal and external change. Effective data governance must be nimble to adjust to varied eco-systems and changing environments. It

[17] Ibid.

must be sufficiently agile to make course corrections as needed, "address challenges, and take advantage of the opportunities they present."[18] Without flexibility for rapid adaptation, governance will falter, processes will stagnate, and outcomes will wane. It is difficult to foresee and prepare for disruption; however, change is constant and organizations must be able to adapt.[19] Organizations must be acutely aware and conscious of the environment within which they operate and of the reach and impact of changing circumstances and tangential relationships.[20]

- Structure

A data governance framework must provide a forum for decision, a forum to build cooperation and determine priorities, a forum for communication. Without structure, optimization will not be realized and data-driven support will languish. An effective structure establishes a clear framework for discourse and resolution. A solid structure helps assure organizational awareness, decisiveness, and accountability.

- Standards

Policies, rules, and procedures facilitate consistency. Standards (e.g., metadata, common data elements, classification protocols) facilitate interoperability, as well as help achieve the highest state of accuracy, completeness, and reliability in application. Standards are necessary to provide a structure that will: (i) bridge disparate cultures within an organization; (ii) bridge diverse data vocabularies and mixed semantics; (iii) bridge differing data sources, syntax, taxonomies, sizes, and complexity; (iv) reconcile inconsistencies; and (v) verify hypotheses. Standards unify an organization to ensure data quality, integrity, usability, and consistency, without which data cannot and should not be relied upon. Standards drive quality, performance, and data management best practices.

---

[18] Ibid., 60.

[19] "Harvard Business School professor Herman "Dutch" Leonard highlights the importance of governance adaptability," Illinois Department of Human Services, "Establishing Governance for Health and Human Services Interoperability Initiatives," *A Report of the State of Illinois Interoperability and Integration Project*, (2013): p. 37. http://www.acf.hhs.gov/sites/default/files/assets/establishing_governance_for_hhs_handbook_508compliant_final.pdf.

[20] Ibid.

- Objectives

Objectives provide clarity of purpose. Objectives must clearly define the goals and strategies of an organization to help ensure a common, clear, and consistent approach. Objectives provide effective direction for oversight, especially where cultural norms are being re-engineered and transformed. Without a clear value proposition, organizations will straggle and drift without purposeful direction, and the value of data will be minimized.

IV. THE RISE OF DATA GOVERNANCE

Data governance is among the hottest topics both for IT practitioners and in boardrooms by reason that organizations and individuals rely upon data more and more for everything they do. The utilization and increasing influence of data is conspicuous in every sector including, but not limited to healthcare, finance, manufacturing, urban planning, media, transportation, engineering, energy, education, environment, and medical research. With that reliance comes an expectation that the data upon which we rely will be of the highest quality, relevant, accurate, consistent, and complete. On the most mundane level, we expect when using a restaurant app to make a dinner reservation that our table will be ready when we arrive. We expect when using a navigation app while driving that the data populating the app is accurate and will get us to our destination. That same expectation, all the more so, applies in life-and-death situations. When a new drug is released to market, the public rightly expects that the drug will be safe based upon accurate, reliable, and trusted data from years of clinical trials. As David Brooks noted: "The rising philosophy of the day…is data-ism."[21]

The world today contains an unimaginably vast amount of data – sometimes structured,[22] sometimes unstructured,[23] and sometimes

---

[21] David Brooks, "The Philosophy of Data," *The New York Times*, February 4, 2013, http://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html?_r=0.

[22] Structured data is "data that resides in fixed fields within a record or file…Relational databases and spreadsheets are examples of structured data. Although data in XML files are not fixed in location like traditional database records, they are nevertheless structured, because the data are tagged and can be accurately identified." "Encyclopedia, Definition of Structured Data," *PC*, http://www.pcmag.com/encyclopedia/term/52162/structured-data.

[23] "Unstructured data is all those things that cannot be so readily classified and fit into a neat box: photos and graphic images, videos, streaming, instrument data, webpages, pdf files, emails, blog entries, power point presentations, wikis, word processing documents."

semi-structured.[24] It is widely consumed, and the collection and analysis of data is multiplying exponentially. Our technical capacities for handling data, Brooks explained, have created the expectation that there is value in measuring everything that can be measured.[25] Data has generated high hopes for future benefits yet to come[26]. Predictive and real time are fast becoming the expected norm, with the growing potential and capacity for big data analytics and "data science"[27] to impact our daily lives and decisions.[28]

To meet these expectations, organizations are increasingly recognizing the necessity of data governance to help society realize data's incalculable potential benefits. If the tenets of sound data governance are rigorously and consistently applied, we will see data's potential in:

- Advancing commerce, science, and research through the discovery, analysis, visualization, and utilization of data in ways never before possible;
- Inspiring creative and cooperative problem solving;
- Generating new insights that can improve the lives of citizens;
- Spurring economic growth and new industry;

"Encyclopedia, Definition of: Unstructured Data," *PC*, http://www.pcmag.com/encyclopedia/term/53486/unstructured-data.

[24] Semi-Structured data is "a type of structured data, but lacks the strict data model structure. With semi-structured data, tags or other types of markers are used to identify certain elements within the data, but the data doesn't have a rigid structure. XML and other markup languages are often used to manage semi-structured data." Vangie Beal, "Definition of semi-structured data," *Webopedia*, http://www.webopedia.com/TERM/S/structured_data.html.

[25] See Brooks, note 21.

[26] Ibid.

[27] See, Gil Press, "A Very Short History of Data Science," *Forbes,* May, 28, 2013, http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/.

[28] Executive Office of the President, see note 1, at 39.

- Optimizing information;
- Maximizing business insight and business value;
- Unlocking innovation;
- Increasing productivity; and
- Creating opportunity.

V. THE OBJECTIVE: FROM RAW DATA TO ACTIONABLE KNOWLEDGE

"Since the first censuses were taken and crop yields recorded in ancient times, data col-lection and analysis have been essential to improving the functioning of society."[29] Data has been the catalyst behind every major scientific revolution.[30] Having data, however, is not the same as knowing what to do with data.[31] The objective of data governance is to help people and organizations move from raw data to actionable knowledge.[32] A keen awareness of how data is transformed into actionable knowledge is therefore integral to designing any effective data governance scheme.

Effective data governance helps minimize the challenges to producing data that is accurate and reliable, meaningful and consequential. Data cannot be assumed flawless, and it does not interpret itself. Drawing conclusions from data for taking action involves more than mathematical calculation. When David Brooks wrote about the rise of data-ism, he astutely identified an issue that is often overlooked: the wise use of data requires a balance between seeming reliance on the data itself and "intuitive pattern recognition."[33] The structure of a governance framework needs to be mindful of tendencies to ignore one over the other and, as such, provides a forum to reconcile differences and make determinations.

---

[29] Executive Office of the President, see note 1, at 1.

[30] Jonathan Shaw, "Why 'Big Data' is a Big Deal: Information Science Promises to Change the World," *Harvard Magazine*, March-April 2014, http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal#.UwTOnV1_5yM.twitter.

[31] Ibid.

[32] Weinberger, see note 7.

[33] Brooks, see note 21.

For the transformation of data into knowledge is simultaneously an art, as well as a science. Knowledge is the result of a "complex process that is social, goal-driven, contextual, and culturally-bound."[34]

VI. DATA GOVERNANCE – CASE EXAMPLES

Data governance provides the necessary structure that will guide entities to realize the potential value of one of their most valuable resources. Just as "organizations are unique in structure, culture, and circumstances,"[35] so too are data governance approaches. Some are established by formal document or all but self-evident in the transparent workings of a firm or agency; others are not as easily discernable but deeply embedded in the daily operations of an organization. In the three examples that follow, the fundamental core principles of data governance are present.

A. Open NY

Successful governance requires a significant confluence of purposes and goals that data governance must harness and that are shared across an enterprise. This is exemplified in the implementation of the New York State Open Data initiative.

New York State's Open Data portal, data.ny.gov, has set a new standard in Open Data. Its purposeful design and content are promoting a new consciousness and enterprise level of understanding regarding data and the application of data as a strategic asset. Ever cognizant of the end user and their relationship with data, the development and implementation of data.ny.gov achieved a Quality-by-Design[36] Open Data portal. It is distinctive and cutting edge in terms of maximizing the public's understanding and utilization of data. It promotes standardization to facilitate reuse and advance interoperability.[37]

---

[34] Weinberger, see note 7.

[35] Mosley and Brackett, *DAMA Guide to the Data Management Body of Knowledge*, 60.

[36] The New York State Open Data platform is a "Quality by Design" web-based public data portal that catalogues data and enables data to be discoverable. Open NY (https://data.ny.gov/) provides access to machine-readable data that can easily be retrieved, downloaded, sorted, searched, and re-used; it provides access to data for end users to process, analyze, and utilize. New York State Open Handbook, Nov. 6, 2013, 5, http://nys-its.github.io/open-data-handbook/OpenDataHandbook.pdf.

[37] Ibid. "It begins the process of standardizing the state's data, which will make it easier for both government workers and the public to discover and use the data. This, in turn,

Exchanging data within, between, and outside organizations can be a very complex process. For data to be interoperable, it must be organized in a standardized way. Metadata provides the essential tools for discovery and reuse. In furtherance of this objective, Open NY provides context[38] for each published dataset requiring, at a minimum: a core metadata scheme,[39] a data dictionary, overview document(s), and supplemental source documents and links, where applicable. A global platform, accessed in more than 180 countries and territories and over 9,000 cities worldwide, Open NY enriches raw content with context in order to maximize end user utility and analysis -- i.e., the transformation of data into actionable knowledge. This enrichment allows for the development of new insights that can improve the lives of citizens and improves the flow of information within and outside government. New York State has established itself as a trusted and respected leader in this sector.[40] Its sustainability and success are attributable to its development and implementation using a solid governance framework.

Concurrent with the launch of the State's Open Data portal in March 2013, Governor Andrew M. Cuomo issued Executive Order No. 95[41] which clearly defined goals, objectives, structure, and processes. Among the many provisions, the Executive Order required the appointment of a Data Working Group (DWG) to serve in an advisory role, as well as the appointment of a data coordinator for every covered state entity. The Executive Order served to advance several critical governance requirements. First, a forum for discussion, communication, and decision support was established in the DWG.

---

advances "interoperability" so the data can be more easily shared and analyzed." http://nys-its.github.io/open-data-handbook/OpenDataHandbook.pdf.

[38] "Data is the representation of facts as text, numbers, graphics, images, sound, or video. Information is data in context. Without context, data is meaningless; we create meaningful information by interpreting the context around data. Information contributes to knowledge. See Mosley and Brackett, DAMA Guide to the Data Management Body of Knowledge, 2.

[39] "Data Documentation and Metadata," University of Minnesota Libraries, https://www.lib.umn.edu/datamanagement/metadata; "Dublin Core Metadata Initiative Metadata Innovation," http://dublincore.org/.

[40] Three months after data.ny.gov was named among the best nationally by the Center for Data Innovation, in November 2014, data.ny.gov was a finalist for the U.K. Open Data Institute (ODI) Data Publisher Awards celebrating high publishing standards and use of challenging data. The UK ODI sought nominations from around the world for pioneers and champions inspiring the use of Open Data.

[41] Exec. Order. No. 95, (March 11, 2013), http://www.governor.ny.gov/executiveorder/95.

Second, the Executive Order articulated roles and responsibilities in the appointment of data coordinators. Third, the combination of the two created a robust network through which policies and standards are disseminated, coordination and collaboration between business and IT is advanced, and implementation of the initiative and achievement of objectives are assured. Fourth, the Executive Order required the development of an Open Data Handbook, which includes guidance for the covered state entities regarding the identification and prioritization of publishable data, as well as models and standards. Supplemental publication guidance further reinforces these standards detailing requisite formats and documentation content. The New York State Open Data Handbook has received national recognition and has become a leading "must read" guidance document for any Open Data initiative.[42] Together, the Executive Order and the Open Data Handbook incorporate the core elements of effective data governance:

- Leadership – a strong and vested executive sponsorship;
- Adaptability – explicit recognition that the Open Data Handbook may be amended from time to time; agility of the Open Data portal;
- Structure – established by Executive Order and operationalized in implementation;
- Standards – detailed in the Open Data Handbook and supplemental guidance documents;
- Objectives – set forth in the Executive Order and the Open Data Handbook.

Data.ny.gov further epitomizes these propositions with its unprecedented breadth and depth of high quality data and agency participation. Its foundation represents a true collaboration between business and IT; between government and its citizenry.

---

[42] "The state of New York has published a very comprehensive handbook on open data that should be in the reference library for any open data initiative," National Association of State Chief Information Officers, "States and Open Data: From Museum to Marketplace - What's Next," (Kentucky, 2014), 10, http://www.nascio.org/ publications/ documents/ NASCIO_EAOpenData_May2014.pdf.

B. The Library of Congress

In 2014, technology is not only fundamentally changing the way data is generated and accessed, but also how data is maintained and collected. Created in 1800,[43] the Library of Congress serves as an example of how a long-established institution is adapting to a dynamic environment in the age of the digital revolution. The Library is the largest library in the world, its collection vast and ever growing.[44] According to the Library's own statistics, it "receives some 15,000 items each working day and adds approximately 12,000 items to the collections daily."[45] The volume of physical additions is dwarfed, however, by the influx of digital material. For example, in 2010, the Library became an archive for Twitter[46] and, in 2013, received a 133 terabyte file containing Twitter's digital archive of pubic tweets from the time the company was founded in 2006 - approximately 170 billion tweets.[47] Today, the Library continues to receive an ongoing stream of public tweets to archive, which grows this one collection by about 500 million tweets per day.[48]

How does the Library integrate new holdings, including digital holdings, into its metadata schema and expand access to its collections?[49] The Library belongs to several international organizations, including the Program for Cooperative Cataloging

---

[43] "History of the Library," Library of Congress, http://www.loc.gov/about/history-of-the-library/.

[44] The Library has "more than 158 million items on approximately 838 miles of bookshelves. The collections include more than 36 million books and other print materials, 3.5 million recordings, 13.7 million photographs, 5.5 million maps, 6.7 million pieces of sheet music and 69 million manuscripts." "Fascinating Facts," Library of Congress, http://www.loc.gov/about/fascinating-facts/.

[45] Ibid.

[46] "Update on the Twitter Archive At the Library of Congress," Library of Congress, January, 2013, http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/; http://www.loc.gov/ today/pr/2013/ files/ twitter_report_2013jan .pdf.

[47] Library of Congress Is Archiving All Of America's Tweets, January 22, 2013, http://www.businessinsider.com/library-of-congress-is-archiving-all-of-americas-tweets-2013-1.

[48] Ibid.

[49] "Program for Cooperative Cataloging," *Library of Congress*, http://www.loc.gov/aba/pcc/.

(PCC),[50] with the aim of expanding access to library collections by "providing useful, timely, and cost-effective cataloging that meets mutually accepted standards of libraries around the world."[51] In furtherance of this aim, and its mission to encourage free and open exchange of data, the PCC is guided by a governance document, which outlines roles and responsibilities and provides for a structure that is sufficiently agile to permit it to adapt to changing needs.[52] It is a living document[53] and embodies all of the core elements of an effective data governance framework.

Complementing the governance document is a strategic directions and actions document that is updated annually. One of the key strategic directions cited for 2014 relates to guidance for metadata creation in the digital environment.[54] This goal is firmly aligned with the objectives outlined in the governance document regarding metadata which emphasize the "emerging variety of new information resources, and the influence of the PCC to facilitate meeting the needs to allow users to effectively search the entire information discovery environment."[55]

C. *Storyful –Data Verification and Finding the Wisdom in a Crowd*[56]

In 2012, Mark Little, the CEO of the social news agency Storyful, authored an article published in Neiman Reports entitled, "Finding Wisdom in the Crowd," [57] in which he struck a cautionary note about validating and verifying data: "Technology will play its part but don't underestimate that human algorithm."[58] He describes a purposeful work environment where journalists work together with engineers,

---

[50] Ibid.

[51] Ibid.

[52] "PCC Organization & Governance," *Library of Congress*, http://www.loc.gov/aba/pcc/about/pcc-org.html.

[53] Ibid. See Governance Document revised February 2014, "Program for Cooperative Cataloging Governance Document," Library of Congress, last revised February 26, 2014.

[54] Ibid. See Mission Statement and Strategic Directions Through 2014.

[55] Ibid.

[56] Little, see note 6.

[57] Ibid.

[58] Ibid.

comparable to business working with IT. He uses the phrase "human algorithm"[59] to sum up Storyful's hybrid approach and underlying philosophy,[60] believing that decisions and the broadest solutions are best informed by a "combination of "automation and human skill," [61] interpreters of nuance. Little describes checklists that the editorial staff utilize to validate and rate the sources of their video and image data.[62] It is journalism in its oldest and purest form:[63] a checklist to identify and validate the who, what, where, when, why, and how of a story – applied to data[64]. The elements Storyful checks in validating data are not dissimilar to many of the elements found in a core metadata scheme. Examples of the validation questions include: (i) "do weather conditions correspond with reports on that day; (ii) do accents or dialects heard in a video tell us the location; (iii) are there other accounts affiliated with this uploader and how can they help us identify location, activity, reliability, bias, and agenda?"[65] Taken together, the collaboration between the journalists and the engineers, the validation methods, and the decision making process provide visibility into the approach to data governance Storyful employs, even though the organization's processes are not formally so labeled.

In a smartphone and a 140 character society, news is a 24/7 cycle. Individual devices capture world events around the clock; crowdsourcing takes place even in the most remote areas. News is no longer a passive event for the audience;[66] the community has become an active participant. Storyful recognizes that the overabundance of data and the explosive amount of related content generated daily demand "new protocols to shape collaboration."[67] Such protocols are the essence of data governance.

---

[59] Ibid.

[60] Ibid.

[61] Ibid.

[62] Ibid.

[63] Ibid.

[64] See Mosley and Brackett, note 5.

[65] Little, see note 6.

[66] Ibid.

[67] Ibid.

VII. CONCLUSION

In 1939, T.S. Eliot wrote a play called *The Rock*. Seventy-five years later, two popular lines from the poem serve as a refrain and a reference point for vigilance in the era of big data: "Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?"[68]

The T.S. Eliot quote aptly serves as a driving force to motivate effective data governance. The transformation of data into actionable knowledge is not formulaic; it is both qualitative and quantitative.[69] Judgment, observation, analysis, and experience help inform our understanding. Implementing a solid data governance framework maximizes the individual elements essential to that transformation: discovery, analytics, theory, and application – which handily form the acronym "DATA." Each element is critical to yielding the greatest return on investment from an organization's information assets.

Sound data governance furthers the promise of extraordinary discovery in the era of big data, open data, and beyond. It assures essential checks and balances. It serves as a catalyst for consequential outcomes derived from quality data, which is pivotal in: identifying new trends and patterns; fostering innovation in the scientific and business communities; and enhancing our lives as citizens and consumers. Effective data governance is imperative to achieve the highest state of organizational and community health, well-being, and advancement.

---

[68] T.S. Eliot, "The Rock" (London, Faber & Faber, 1934).

[69] "The forward edge of science, whether it drives a business or marketing decision, provides an insight into Renaissance painting, or leads to a medical breakthrough, is increasingly being driven by quantities of information that humans can understand only with the help of math and machines. Those who possess the skills to parse this ever-growing trove of information sense that they are making history in many realms of inquiry. "The data themselves, unless they are actionable, aren't relevant or interesting," is Nathan Eagle's view. "What *is* interesting," he says, "is what we can now do with them to make people's lives better." Shaw, see note 29.