Data Quality Improvement Through OODA Methodology

(Completed Research Paper)

Arun Sundararaman and Sree Krishnan Venkatesan
Health Analytics Solution Factory, Accenture Technology, India
Arun.sundararaman@accenture.com Sreekrishnan.venky@accenture.com

The topic of Data Quality (DQ) in the field of Information Management has been extensively researched. DQ methodology continues to be an area of significant importance within the study of DQ. While few established methodologies exist for defining, measuring and managing DQ, the need to develop new methodologies for a holistic management of DQ continues to grow. The Observe-Orient-Decide-Act "OODA" framework provides great potential for adoption in the study of DQ. This Paper introduces the OODA framework and discusses different proposed models that may be considered for implementation of the same as a methodology in the study of DQ. This paper covers a detailed discussion on practical work conducted to address DQ challenges faced by a large Health Insurance Organization. Evolution of different maturity models of OODA for DQ along with the results are presented and analyzed. This paper recommends research directions for adoption of the OODA Methodology for DQ.

Additional Key Words and Phrases: Observe Orient, Decide, Act, data quality, methodology, big data quality, maturity model, data quality dimensions, assessment.

1. INTRODUCTION

Commenting on poor quality of data (and its impact on confidence levels of policy decisions), Y.V.Reddy, former Governor of Reserve Bank of India, India's Central Bank, once said "everywhere around the world the future is uncertain; in India, even the past is uncertain!"[1]. Looking at this quote from the data quality (DQ) perspective and its impact on decision making, be descriptive or predictive, we could state that "with good quality of data the future (predictive analytics) could be uncertain; without adequate DQ measures, even the past (descriptive analytics) can also be uncertain".

Data is the nerve center of decisions and actions in any organization (profit, not-for-profit, commercial, Government or any other forms of organizations or business entities). Existence of good levels of Data Quality (DQ) influences success or failure of organizations [2]. A recent research publication estimates that poor quality data costs US # 3 Trillion per year [3]. Stated in non-monetary terms, some of these impacts include estimates, such as: 50% — the amount of time that knowledge workers waste in hunting for data, finding and correcting errors, and searching for confirmatory sources for data that they don't trust or 60% — the estimated fraction of time that data scientists spend cleaning and organizing data,

While people often tend to consider DQ as synonymous merely with data accuracy, DQ is much more than simply data accuracy. Past research in Information Quality (IQ) point out that Data and Information Quality can be conceived as a multi-dimensional concept with varying attributes depending on the individual researcher's viewpoint. Most commonly, the term "Data Quality" is described as data that is "Fit-for-use", which implies that it is relative, as data considered appropriate for one use may not possess sufficient attributes for another use.

DQ is subjective in nature and therefore if assessed independent of the business objective for which the data is intended to be used, it is likely to lead to incorrect results. Just as every metric is defined to provide a definite perspective of any selected subject or as every tool is designed to be used for specific purposes, DQ needs to be defined and approached with purpose in mind; the quality of business outcomes can be a common and relevant purpose in the study of Decision Support Systems (DSS). There are several issues that remain unresolved with respect to the relationship between information quality improvements and organizational outcomes. More and more companies are recognizing that data is a key organizational resource, and all kinds of business data are used increasingly in strategic information systems in decision support. The ability of an organization to make accurate strategic decisions is greatly weakened when the DSS/data warehouse contains inaccurate data. This necessitates development of DQ Assessment Methods that assess and measure DQ. However, despite a decade of research and practice, only piece-meal techniques are available for measuring, analyzing, and improving DQ in organizations. There are several issues that remain unresolved with respect to DQ and the relationship between DQ and organizational outcomes.

Research on DQ have continued to be focused on traditional sequential phases e.g. Assessment Phase, Improvement Phase and most of these are based on process redesign approach [4]. Further methodologies published so far adopt either data-drive techniques or process-driven techniques, whereas, a comprehensive DQ approach needs combination of both these techniques. The limitation with these approaches is the lack of focus on quality of business outcomes and the associated research challenge lies in the need to develop a framework that is focused on quality of business outcomes that result from the quality of decisions that are derived from use of Information from Decision Support System (DSS), i.e. development of a model to measure DQ from the quality of business outcomes.

Further, published report indicates that from a historic perspective correlation exists between quality dimensions and the evolution of ICT technologies [4]. Therefore, broad objectives of this research work are as follows:

- Explore new methodologies for DQ that addresses the iterative nature of data flows into modern IT Systems
- Methodologies that are not sequential in nature, but, that can provide for inter-linked phases
- Methodologies that are capable of being both process and data driven
- Address evolving DQ requirements and challenges arising from BigData platforms and Information Systems built on these platforms that carry the inherent volume, variety and veracity challenges.
- Domain related aspects of DQ

This research aims to address these challenges. We developed an approach based on OODA framework and explored its applicability to DQ management.

This paper introduces an "Iterative OODA Framework" for DQ and presents the results from actual work conducted using the said framework to address real-life DQ challenges. The paper also suggests future research directions to extend and/or modify this framework to manage other DQ challenges.

2. LITERATURE REVIEW

2.1 Study of Data Quality Assessment frameworks

Existing literature provides a wide range of techniques to assess and improve the quality of data, such as record linkage, business rules, and similarity measures. Recent research by Stvilia et. al., 2007 [5] has focused on defining methodologies that help select, customize, and apply data quality assessment and improvement techniques. According to Batini et. al., 2009 [4] DQ assessment methodology may be defined as a set of guidelines and techniques that, starting from input information describing a given application context, defines a rational process to assess and improve the quality of data. In the cited work, Batini et. al., 2009 summarize different perspectives that can be used to analyze and compare DQ methodologies, listed below:

1. Phases and steps that compose the methodology (includes assessment / measurement)
2. The strategies and techniques that are adopted in the methodology for assessing and improving DQ levels
3. The dimensions and metrics that are chosen in the methodology to assess DQ levels
4. The types of costs that are associated with data quality issues
5. The types of data that are considered in the methodology
6. The types of information systems that use, modify, and manage the data that are considered in the methodology
7. The organizations involved in the processes that create or update the data that are considered in the methodology
8. The processes that create or update data
9. The services that are produced by the processes that are considered in the methodology

Table 1 provides a list of existing DQ Measurement / assessment methodologies

S No	*Acronym*	*Name of the Methodology*	*Author*	*Reference*
1	TDQM	Total Data Quality Management	Wang	Wang et. al. 1998
2	DWQ	The Datawarehouse Quality Methodology	Jeusfeld	Jeusfeld et. al.,, 1998
3	TIQM	Total Information Quality Management	English	English, L., 1999
4	AIMQ	A Methodology for Information Quality Assessment	Lee	Lee et. al., 2002
5	CIHI	Canadian Institute for Health Information Methodology	Long and Seko	Long. J et. al., 2005
6	DQA	Data Quality Assessment	Pipino.L, Lee.Y and Wang.R. Y.	Pipino, L. et. al., 2002
7	IQM	Information Quality Management	Eppler	Eppler, M., 2002
8	ITSAT	ITSAT Methodology	Falorsi, P., Pallara, S., Pavone, A., Alessandroni, A., Massella, E., and Scannapieco, M.	Falorsi et. al., 2003
9	AMEQ	Activity-based Measuring and Evaluating of product information Quality (AMEQ) methodology	Su and Jin	Su, Y and Jin, Z 2004

10	COLDQ	Loshin Methodology (Cost-effect Of Low Data Quality	Loshin	Loshin, 2004
11	DaQuinCIS	Data Quality in Cooperative Information Systems	Scannapieco, M., Pernici, B., and Pierce, E	Scannapieco et. al., 2005
12	QAFD	Methodology for the Quality Assessment of Financial Data	Amicis, De F. and Batini, C.	Amicis and Batini, 2004
13	CDQ	Comprehensive methodology for Data Quality management	Batini, C., Cabitza,F., Cappiello,C. and Francalanci, C.	Batini et. al., 2008
14	HDQM	Heterogenous Data Quality Methodology	Batini Carlo1, Barone Daniele1, Cabitza Federico1 and Grega Simone	Batini et. al., 2011

Table I. Existing Data Quality Assessment Methodologies

In a recently published work Ge et. al., 2011 [6] have dealt with the difficulties associated with assessing information quality. Through their work, while they acknowledge that research provides several approaches to measure information quality and many case studies constantly illustrate the difficulties in assessing information quality, they reveal that even though several IQ assessment frameworks have been proposed, in practice, organizations are still facing difficulties when implementing these assessment frameworks. Through a wide-ranging literature survey, the authors further reveal that most existing frameworks are too generic to be used for assessment purposes or merely remain at a theoretical stage. Hence, they emphasize the need to address the limitations of some IQ frameworks, and to develop a practical IQ model based on valid and reliable measurements. Summarizing the findings of their work the authors conclude that IQ is a complex and multi-dimensional phenomenon, which has yet not been fully understood and that this nature of IQ causes challenges to measure IQ and may explain why current frameworks have their limitations.

Literature to study several approaches and methodology for DQ assessment has been found to be multidisciplinary in nature. These studies may be grouped as those focused on attributes to identify software quality (ISO Model), Users' DQ problem detection, information credibility and a host of DQ related problems. While all these methodologies have merits, most of them focus on narrow specialties, whereas, all-encompassing DQ Methodology viz., Information Value Methodology evolved to study DQ [7].

Similarly, in a more recent published work, 15 high level categories of research methods have been identified for DQ research [8] and conclude with an anticipation that DQ research will continue to grow and evolve and recommend that new forms and methods of DQ research is needed to develop techniques for managing and improving the quality of data. To summarize, while there exists a strong need for focused methodologies for study of DQ, limited research has been conducted in this direction.

2.2 *OODA Framework*

The OODA framework was initially defined by John Boyd and the OODA loop consists of four main steps: Observe, Orient, Decide and Act [9] as represented in Figure 1.

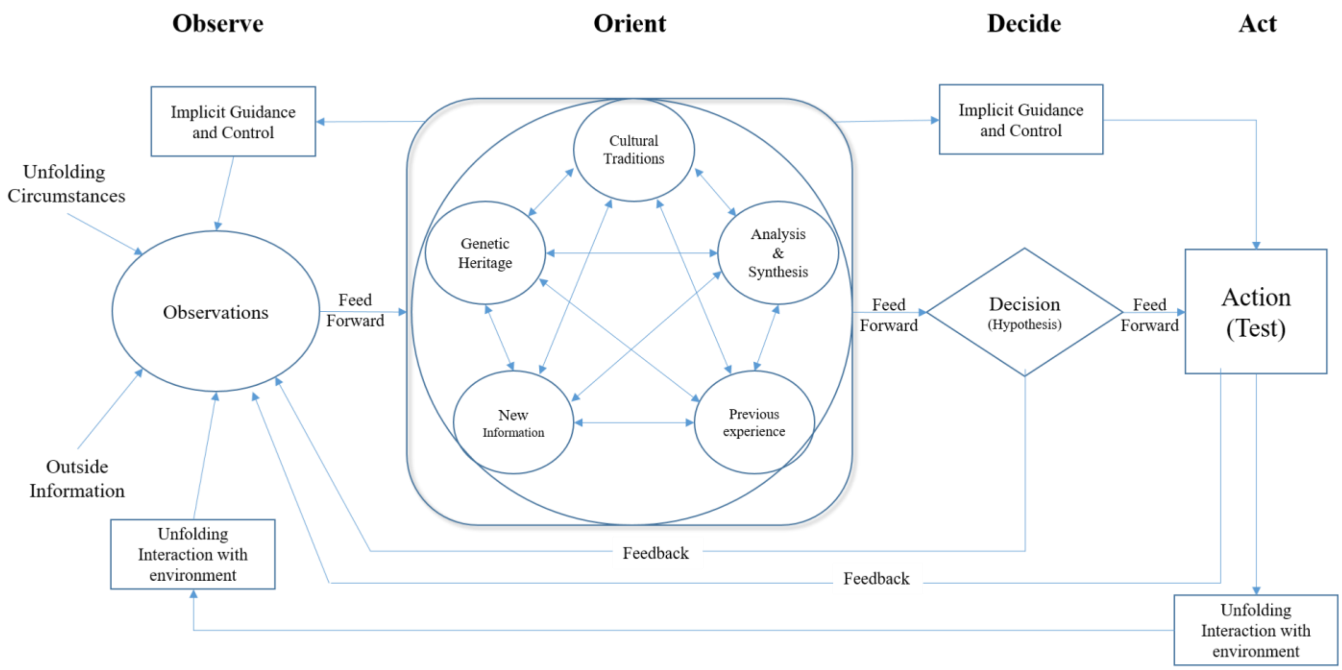

Fig. 1. OODA Loop Framework

It may be observed that the OODA framework does not address an important attribute which is extremely critical for analysis and decision making i.e. time. The initial version of OODA was defined in a sequential paradigm i.e. discrete and consecutive steps of collecting information, processing information, decision making and initiating actions. This approach to decision making lacks the dimension of multi-processing, that is today implemented in real-time systems, besides approaching decision making as a sequential event instead of a parallel process. To address these gaps, Kannisto argues that it is possible to create and further develop a conceptual model based on the base OODA and to create a general concept to support a wider range of organizations in their attempt to gain better level of situational awareness (SA) and to support decision making [10].

Another published work [11] states that any organization and its goals can be described as several OODA loops that can be cycled iteratively. The faster users can cycle through an OODA loop, the more competitive the organization can be when it comes to correcting problems and improving performance. The said work concludes that the combination of OODA loops with Key Performance Indicators (KPI) can render an organization agile and competitive when using Business Intelligence and emphasizes the importance of DQ to achieve faster loops. Use of OODA and improving the speed of the OODA loops for better DQ Management and decision support was reinforced in the work of Sarah North et. al [12].

With this motivation, gaps in the current DQ methods as listed earlier and the need to explore new methodologies for DQ research, the authors carried out work that involved multiple iterations of DQ measurement, analysis, decisions and cleansing actions. This work was carried out using the OODA framework to address DQ issues of a healthcare organization. This work was carried out to address a real-life DQ problems faced by a large Healthcare Organization [Blue Cross Blue Shield Association]. Due to confidentiality purposes the authors of this work are constrained from referring the name of the client. This work was carried out in Accenture's Health Analytics Solution Factory over a period of 4 months and the results reported in this paper are based on the said work.

3. OODA METHODOLOGY FOR DATA QUALITY

There exists limited research on the adaptability of OODA for DQ. There exists a notion that OODA methodology both misses essential elements of data quality practice and is overly complex and that OODA methodology offers a complex way of approaching error correction. The OODA Framework does not intend to replace the data quality techniques or error correction frameworks but rather attempts to overlay such techniques in a methodology so that the DQ issues can be handled efficiently in a time bound manner. An organization would approach DQ correction using tools and techniques at their disposal in a strategy that is phased in its approach as depicted in the multiple waves of maturity. Boyd's OODA "loop" provides an effective framework for igniting creativity and initiative throughout an organization and harmonizing them to achieve the organization's goals. To achieve this, organizations must evolve their own practices suitable for their people and their competitive environments [13].

3.1OODA Loop in Data Quality

In practice, OODA Loop for DQ may result in multiple waves of actions triggered by one more a set of related observations. This is even more likely in addressing DQ in DWH / BI Systems, since data in DWH is typically structured, modeled and stored in a manner that it would impact multiple functions, support multiple analytical models etc. Moreover, considering the dynamic nature of incremental data feeds to the DWH on a daily or in many cases intra-day basis an approach that supports inter-connected or parallel loops may be considered ideal for practice. A conceptual representation of recommended adoption of OODA for DQ in DQH environment is depicted in Fig 2.

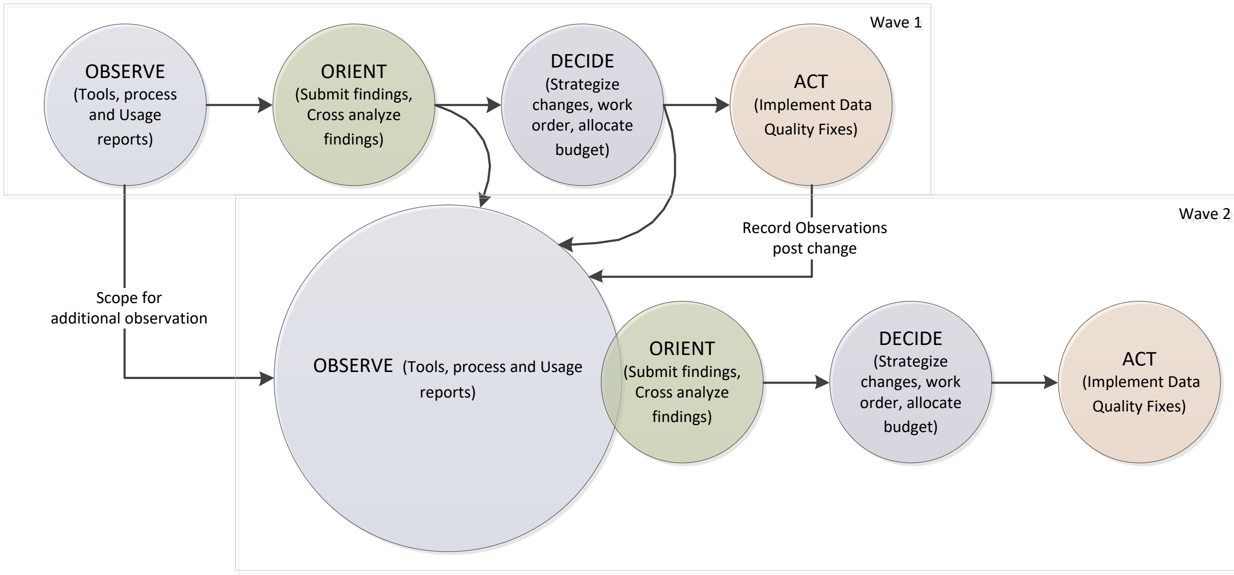

Fig. 2. OODA Loop Framework applied for study of DQ

3.1.1 Observe

This phase involves utilizing tools, process and people to observe DQ issues through the data changes needed in the organization's existing data. Tools, such as monthly or weekly reports, dashboards are utilized to identify the changes required to data thus indicating DQ issues. Issues logged by end users of the data are also a good source to observe DQ issues. Profiling tools help observe the data at periodic intervals and notify potential DQ issues, such as anomalies, out of range values, invalid values, etc. Observations are also made based on mandated reports sent to government agencies or external vendors and feedback received from those external agencies. Observations can be reactive or proactive.

3.1.2 Orient

A focused team is set up to analyze DQ issues to measure DQ and identify the severity of DQ issues identified in the observe phase. A Data governance team is established to deeply analyze the root cause and nature of DQ issues observed. Orienting multiple stakeholders brings a common understanding of DQ issues and multiple stakeholders are involved to agree on the root cause of DQ issues.

3.1.3 Decide

Tactical and operational decisions are made on the nature of fixes required to address DQ issues. Examples of such decisions include one-time clean-up of history data or changes to Extract-Transformation-Load routines that load data to the DW or having the source systems (from where data to the DWH is sourced) to implement data validation or data capture rules etc. The other key decision relates to the size of the team that needs to act on the fixes; this may depend on decisions around fixes and/or the extent of the application systems that need to be modified for affecting the fixes.

3.1.4 Act

Decisions listed in the previous phase are implemented and the original teams that were part of the observation are engaged to validate that the decisions and actions result in closing the identified DQ issues. All these actions are implemented along with stringent validations measures to ensure that only approved changes are made either to the data or data processing routines.

3.2 *OODA Loop in Data Quality – Maturity Models*

Since adoption of OODA as a framework to address DQ is in its early stages, organizations may go through 4 levels of maturity in effectively utilizing the full potential of OODA concepts. Velocity of OODA loops, extent of overlap and volume and value of DQ issues identified are the key dimensions that define this maturity. The framework for this methodology suggests an iterative approach that reflects advancing levels of organizational maturity and experience.

3.2.1 MATURITY MODEL 1

This is typically the most common model and a direct adoption of OODA framework, as depicted in Fig 3.Typical characteristics of this model are as listed below:

- Traditional Observation techniques are used. End users of data find anomalies or an external vendor observes deviation.
- Existing correction processes are used and data is not used until it is corrected using SDLC process of Analysis – Design – Build – Test – Deploy.
- DQ mechanisms are not inbuilt and are reactive to triggers in the existing system.

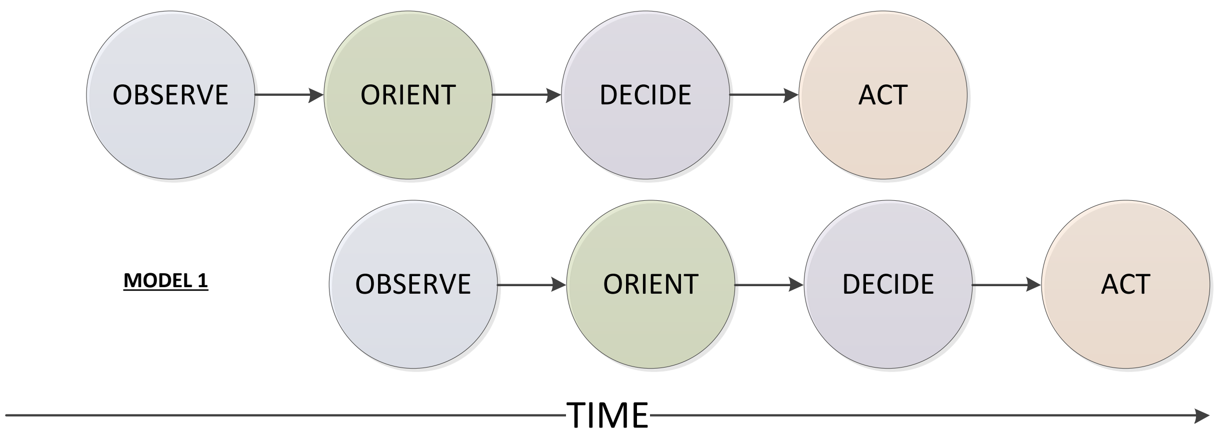

Fig. 3. Model 1 of OODA Loop Framework for DQ

3.2.2 MATURITY MODEL 2

This model introduces overlaps between phases and is typically best suited to clean up data right after a major shift in organization data sources like, new source, changes to existing source, etc. A view of this model is depicted in Fig 4. Typical characteristics of this model are as listed below:

- Observation is a result of an existing set of actions that caused changes to the data.
- Reaction is faster because of pre-defined timelines or a dedicated project to monitor DQ issues.
- Phases overlap because of the need to deliver quicker turn around in cleaning up DQ issues.

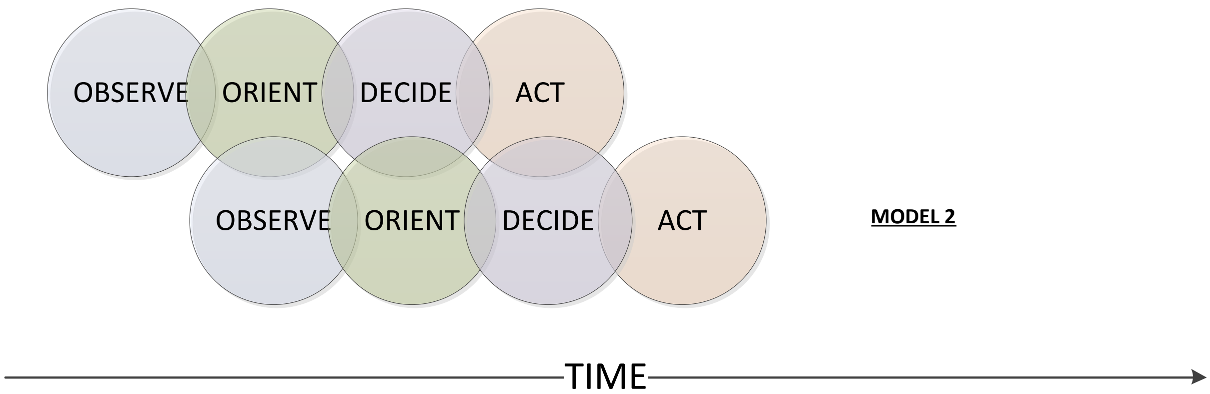

Fig. 4. Model 2 of OODA Loop Framework for DQ

3.2.3 MATURITY MODEL 3

This is an extension of Model 2 and introduces extensive phase overlaps and rapid timelines to address DQ issues. This model is depicted in Fig. 5 and the typical characteristics are listed below:

- This typically follows a Model 1 or Model 2 implementation where patterns of DQ issues are observed that require a major attention.
- In most cases a parallel rebuild of the databases, which are perceived to be of lower quality, are done and is expected to be completed at a shorter period.
- Changes are observed, decisions made in near real time basis leading to extensive phase overlaps and faster correction timelines.

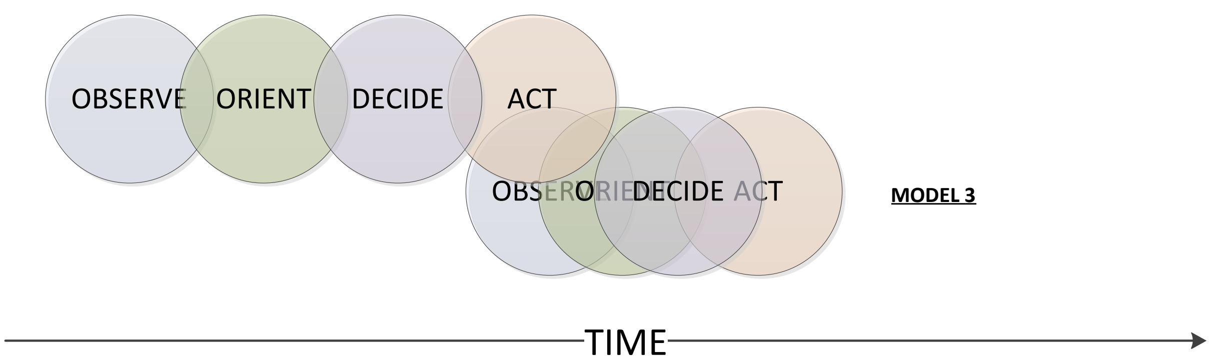

Fig. 5. Model 3 of OODA Loop Framework for DQ

3.2.4 MATURITY MODEL 4

Observe and Orient functions as a regular ongoing activity, not necessarily driven by DQ issues, are the hallmarks of this model. This model is depicted in Fig. 6 and the typical characteristics are listed below:

- Utilization of advanced tools and techniques, such as Machine Learning and Artificial Intelligence algorithms trained to identify data anomalies.
- Large volumes of data ingested by source systems.
- Constantly reviewed and Decision Support systems are trained to make decisions on Data flows.
- Phase overlap becomes a natural process because of Proactive Nature of Observations.

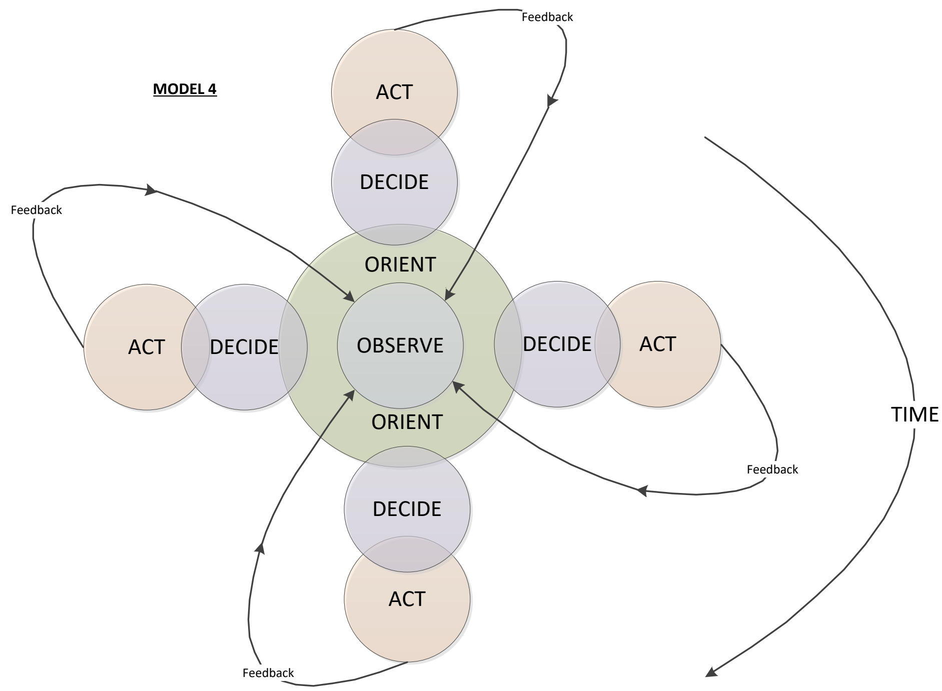

Fig. 6. Model 4 of OODA Loop Framework for DQ

3.3 OODA for DQ – Details of research work

As stated earlier, the underlying work for this paper was a carried out in a research setting within a practice work carried out for a large Health insurance organization, referred to as client in this paper. The client had an enterprise data warehouse (DWH) comprising data from several source systems and data organized in dimensional models for reporting requirements of various subject areas, such as Membership, Provider, Claims, etc. Over a period, some of the main source systems witnessed major changes in their functionality or new applications were implemented. For example, few years back the client implemented a leading commercial software application for managing their claims processing requirements. Over a period, the users of DWH started reporting higher levels of concerns in the quality of data and their ability to rely on the data contained in DWH. For nearly 2 years, these DQ issues were being handled at a slow pace and attention and with a traditional "waterfall" like approach. This comprises wave 1 of the work. Since this approach did not provide the necessary speed and impact to DQ, users of data heightened their concerns on the quality of data and hence DQ problems starting attaining other dimensions, such as accuracy, timeliness, reliability etc. This triggered the need to set up a focused team on DQ, a team that did not exist earlier with the client. The authors of this paper had significant roles to play in this DQ assignment covering data profiling and analysis, DQ measurement, root cause analysis, solution design and implementation. This DQ assignment was carried out over a period of 9 months and comprised a total team of 10 resources working full time on this assignment. These 9 months may further be broken into 2 iterations, named as Wave 2 and Wave 3. In summary, this work is represented as 3 waves in Fig. 11. Waves 2 and 3 of this work witnessed a planned shift in formalizing adoption of OODA framework for DQ, increasing the speed and concurrency of OODA loops in a controlled and planned manner. Thus, the work progressed from OODA for DQ Maturity Model 1 to Model 2 and Model 3 described in the previous section. 2 key factors were measured and monitored to track the success and progress of this work i.e. the velocity of the OODA loops (the speed with which the DQ problems could be analyzed, solutions proposed and implemented to release good quality data to users) and the number of attributes in the data that could be taken up for DQ improvements e.g. Member address or Diagnosis Code, Referring provider information, New Born Birth weight indicators, Diagnosis codes, APTC amount, Type of Facility Code, Payment amount, Product Ids, Surgical Admission indicators, etc). The overall flow of these waves, key activities performed in each of these is listed in Table II and the maturity model followed through this work is depicted in Fig. 7.

Wave	Observe	Orient	Decide	Act
1	End users regularly use information for their purpose but also compare them with regular patterns for anomalies.	End users submit information to the DWH team for deep dive on the anomalies observed. DWH teams analyzes and submits findings on the analysis	Teams then picks up the issues and decide on a date to fix the issue based on priorities and predefined timelines	Fixes are implemented and validated by the data users to ensure anomalies done exist anymore.
2	End users observe growing list of anomalies cross impacting multiple subject areas reflecting in growing DQ issues.	Parallel focus on observe and orient introduced as a pilot exercise.	Decisions were made on prioritization and split into multiple releases. A total of 47 DQ issues across various layers were scheduled to be delivered in 6 releases.	The project implemented fixes that would clean up ~ 25 attributes ending up correcting millions of records across layers and ensuring corrections to data processing routines that would prevent future DQ issues. Attributes cleaned up include Referring provider information, New Born Birth weight indicators, Diagnosis codes, APTC amount, Type of Facility Code, GL Code, Payment amount, Product Ids, Surgical Admission indicators etc.,
3	As Wave 2 Orientation was in play, it was observed that most DQ issues fall under any of the 4 categories viz., Business Logic changes, Technology induced issues, Latency and Source issues and Derived Attributes Observations were made that there might be additional issues uncovered by the previous Wave.	Random checks on observations revealed the issues were not very evident thus necessitating parallel approach across and speeding up the duration of the loops.	Wave 3 Decide was aligned to be in parallel with and based on that Wave 2 Act decisions.	Key Subject areas (Claims, Accounting) were rebuilt in an alternative environment and a comparison with the current data set was made to identify records required to update. Over 40 attributes across multiple layers (Staging, Foundation and Data Marts) were rebuilt and identified to be requiring updates. In addition to attributes from Wave 2, new attributes requiring DQ fix included Discharge status code, Admission codes, Pricing related codes, Coordination of Benefits, Claim line Denial Indicators apart from missing records.

Table II. Waves, key activities and maturity models

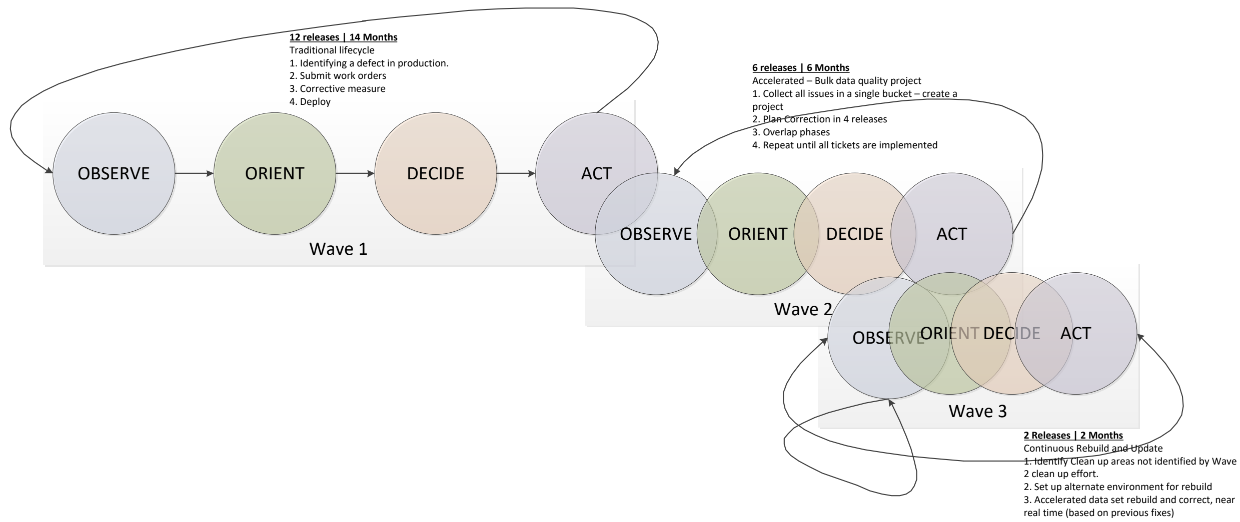

Fig. 7. Research work conducted: OODA Loop Framework for DQ

Summary of the results from the work are given in Table III.

Wave	Number of Loops	Model	Velocity	DQ Issues Identified per loop	Attributes/ DQ span
1	12	Model 1	8 weeks	~3-5	~ 3-5 attributes; few thousand rows of data
2	6	Model 2	4 weeks	~8	~13 Attributes, millions of rows across layers (staging, Foundation, Data marts),
3	2	Model 3	3 weeks	~ 14	27 Attributes and 1.5 million rows.

Table III. Results of DQ work using OODA framework

4. CONCLUSION AND FUTURE DIRECTIONS

It may be observed that adoption of the OODA framework has been helpful to address the DQ issues in a timely manner and encompassing a larger data set. Also, the results from the work demonstrate that different maturity models of the OODA framework may be adopted to address DQ issues, based on several factors, such as context of the subject, time sensitivity of DQ issues, DQ measurement etc. Another key conclusion from the result that needs highlight is that with adoption of higher maturity models (as discussed earlier), the velocity of loops (i.e. time taken to analyze DQ problems to the point in time where DQ solutions are implemented) improves, the number of attributes that are impacted by DQ increase.

Speed remains the most critical aspect of OODA i.e. any organization that can execute the Observe→Orient→Decide→Act loop fastest will be most successful in today's dynamic business environment. Data and Data Quality are essential elements of every facet of this loop and the power of emerging technologies, such as Artificial Intelligence and Predictive Analytics data and therefore data quality is emerging as the new weapon for enterprises [14]. Future research can be directed towards study of automation options in analysis, measurement and review of DQ in each phase of OODA e.g. visualization techniques for Observe, analytics for Orient (i.e. relate data sets etc.), machine learning based decisions and IoT to initiate actions. Each of these options may be applied for in-depth study of role of DQ.

This paper presented a new framework for identifying and improving DQ through cycles of application of the OODA methodology. The framework for this methodology suggests an iterative approach that reflects advancing levels of organizational maturity, and the study results presented in this work demonstrate that advancement can occur with experience. With increasing maturity, organizational goals of developing the ability to undertake rapid improvement cycles and the ability to address increased numbers of identified DQ issues are better addressed. The methodology presented is adaptable that it could be used across industries, organizational types, and organizational sizes.

The focus of this work was to explore 2 dimensions of DQ i.e. speed and volume (coverage) by applying OODA to DQ discovery process. Future work may be extended to explore other DQ dimensions such as timeliness, accuracy etc.

Woodall et. al., 2010 [15] suggest that no individual existing technique or methodology for assessing DQ is wholly suitable to assess DQ for all types of requirements due to the varying nature of requirements over time and organizational needs; the requirements may be different for every organization and even the same organization over time. While some of the DQ assessment techniques are geared towards specific application areas and are often not suitable in different applications, other techniques are more general and therefore do not always meet specific requirements. In this context, while the current work focused on adopting OODA for DQ in DWH, future work may be explored to identify specific tailoring requirements for other applications e.g. ERP or portals.

With the advances in Big Data technologies volume, velocity and veracity of data is increasing [16], [17]. With the advancement of data lake architectures, new models of data storage and approaches to aggregation of data are emerging. With these advances, the need to address DQ right at the ingestion stage assumes greater significance. Traditional approaches to "observe" are based on data profiling which are rule based approaches. Use of advanced techniques, such as machine learning for early identification of DQ issues may be explored in future work, as part of the "observe" phase. In a related work [18] efficacy of different machine learning techniques for scaling out virtual clusters for the execution of deduplication algorithms under predefined time restrictions was investigated. Similar approaches may be adopted for other DQ issues beyond deduplication and considered as part of "observe" phase of OODA.

In summary, the topic of DQ methods and research approach remains to be explored and several issues continue to remain open in the study of DQ as it pertains to impact of DQ on analysis and decision making. Study of these subjects assumes significance based on the concepts of context and comprehensiveness as discussed in published literature. The research community has a huge potential to carry out focused research in various directions (listed above) with an objective to address these open issues.

REFERENCES

1. D. Subba Rao, "Who Moved My Interest Rate: Leading the Reserve Bank Through Five Turbulent Years", 1st ed. Penguin Books, 2016, pp.54-54.
2. Haug, A., Zachariassen, F., & van Liempd, D. 2011. The cost of poor data quality. Journal of Industrial Engineering and Management, 4(2), 168-193
3. Thomas.C.Redman. 2016. Getting in front of Data : The who does what. Technics Publication.
4. Batini,C., Cappiello,C., Francalanci,C. and Maurino,A .2009*, ―Methodologies for Data Quality Assessment and Improvement‖,* ACM Computing Surveys, Vol. 41, No. 3, Article 16.
5. Stvilia,B., Gasser, L.,Michael, B. T and Linda, C. 2007, "A framework for Information Quality Assessment", Journal of the American Society for Information Science and Technology, Vol 58, No. 12, pp. 1720-1733.

6. Ge, Mouzhi; Helfert, M. and Jannach, D. (2011), "Information quality assessment: validating measurement dimensions and processes", proceedings of the 19th European Conference on Information Systems.
7. Marilou Haines and Elizabeth M. Pierce. 2013. The Information Value Methodology : How average users assess IQ on the Web, Proceedings of the International Conference on Information Quality, ICIQ 2013.
8. H. Zhu, S. E. Madnick, Y. W. Lee, R. Y. Wang, "Data and information quality research: Its evolution and future", Computing Handbook, no. 2, pp. 1-20, 2014
9. Boyd, J., 1987. A discourse on winning and losing. Maxwell Air Force Base, AL: Air University Library Document No. M-U 43947.
10. Anne Koskinen-Kannisto, 2013. Situational awareness concept in a multinational collaboration environment. Doctoral Dissertation, National Defense University, Helsinki.
11. Middelfart, M. 2007. Improving Business Intelligence speed and quality through the OODA concept. Proceedings of the ACM tenth international workshop on Data warehousing and OLAP (DOLAP '07). ACM, New York, NY, USA, 97-98.
12. Muhammad Obeidat, Max North, Ronny Richardson, and Vebol Rattanak. 2015. Business Intelligence Technology, Applications, and Trends. International Management Review, Vol 11 (2).
13. Chet Richards. 2012. Proceedings of the Lean Software & Systems Conference 2011, E. Willeke, ed., Sequim, WA: Blue Hole Press,pp. 127-136
14. Vuorisalo, Valtteri. 2016. Weaponizing your data: Take OODA to the next level. Accenture Research.
15. Woodall,P and Parlikad, A. (2010), "A hybrid approach to assessing data quality", proceedings of the 15th International Conference on Information Quality.
16. Lukyanenko, Roman. 2016. Information Quality Research Challenge: Information Quality in the Age of Ubiquitous Digital Intermediation. Journal of Data and Information Quality, Vol. 7, 1-2, 3 pages. DOI=http://dx.doi.org/10.1145/2856038
17. Sreenivas R. Sukumar, Natarajan Ramachandran and Regina K. Ferrell. 2015. Quality of Big Data in health care. International Journal of Health Care Quality Assurance DOI: 10.1108/IJHCQA-07-2014-0080
18. Dimas Cassimiro Nascimento, Carlos Eduardo Pires and Demetrio Gomes Mestre. 2016. Applying machine learning techniques for scaling out data quality algorithms in cloud computing environments. Applied Intelligence, Vol 45, pp. 530-548, Springer.
19. Canadian Institute for Health Information. 2009. The CIHI Data Quality Framework; CIHI: Ottawa, ON, Canada, 2009.
20. Batini Carlo, Barone Daniele, Cabitza Federico and Grega Simone. 2011. A Data Quality Methodology for Heterogeneous Data. International Journal of Database Management Systems (IJDMS), Vol.3, No.1, February 2011
21. Chung, W., Craig,F. and Wang, R. Y. 2005, ―Redefining the Scope and Focus of Information Quality Work‖, Information Quality. New York: M.E.Sharpe, pp. 230-248.
22. Helfert, M. and Foley, O. 2009, ―A Context Aware Information Quality Framework‖, Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology, pp. 187-193.
23. Hong Chen, David Hailey, Ning Wang and Ping Yu. 2014. A Review of Data Quality Assessment Methods for Public Health Information Systems. Int. J. Environ. Res. Public Health 2014, 11, 5170-5207; doi: 10.3390/ijerph110505170