International Journal on Emerging Technologies **11**(3): 648-653(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

A Dimensional Data Model for Healthcare Revenue Cycle Analytics and Business Intelligence

Joseph George[1] and M. K. Jeyakumar[2]
[1]Department of Computer Science and Engineering,
Noorul Islam Centre for Higher Education Kumaracoil (Tamilnadu), India.
[2]Professor, Department of Computer Applications,
Noorul Islam Centre for Higher Education Kumaracoil (Tamilnadu), India.

(Corresponding author: Joseph George)
(Received 28 March 2020, Revised 06 May 2020, Accepted 09 May 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The effectiveness of any data collected, relays completely on how we utilize it for better performance of the organization. In short, how is it being used to derive intelligence or rather Business Intelligence. Business Intelligence (BI) and Artificial Intelligence (AI) are among the top technology interests in current era. Different industries utilize BI and Data Warehouse (DW) in different forms and are on different maturity levels of the utilization. Healthcare industry being one of the most data intensive domain, demands a unique way of data handling. Maturity level of DW in healthcare domain is lower than other industries, like manufacturing or e-commerce.

Self BI enables the end user to drill down and roll up through the data to arrive at desired conclusion without the intervention of any technical support. In this paper, authors are trying to develop a dimensional data model for healthcare revenue cycle, which will in turn facilitate Self BI. An effort is made to combine the clinical and financial dimensions together to arrive in a holistic view. A phased approach which consisting of four stages is utilized in the study and the resulting model is modular in nature to join and accommodate the various other data marts, representing other areas of healthcare.

Keywords: RCM, Clinical Business Intelligence, Data Warehouse, Data Analytics.

Abbreviations: BI, Business Intelligence; DW, Data Warehouse; RCM, Revenue Cycle Management; AI, Artificial Intelligence; HIS , Hospital Information System; ERP, Enterprise Resource Planning; SSIS, SQL Server Integration Service; ETL, Extraction Transformation Loading; SQL, Structured Query Language; ROI, Return of Investment; DRG, Diagnosis Related Group; ICD, International Classification of Diseases; CPT, Current Procedural Terminology ; LOS, Length of Stay.

I. INTRODUCTION

Healthcare business is very complicated and unique in nature and the data associated with it is even more complicated [2]. In healthcare, data is growing in an exponential manner and the variety of data and the velocity rate of data accumulation are second to none. In a broader perspective, data in healthcare can be divided into two major categories: Clinical data, which is generated during the care delivery process and the Administrative data, which is complementing the same care delivery process [3]. One of the examples of administrative data is the billing information, which constitute the line items of a bill or invoice generated for the specific patient encounter.

Transforming information into knowledge for informed decision making is key to any industry to get a competitive advantage [4]. Here comes the role of analytics and a well-defined Data Warehouse, which is the key for a matured BI scenario. Better care delivery and improved operational efficiency are the primary goals of any care provider. These two will result in financial benefits automatically.

Adopting a top down dataware house model [5] for analytics in healthcare is not much feasible in many instances due to various reasons such as, considerable initial investment, extensive time required and timely ROI. Also, defining the master blue print of the enterprise data model beforehand is also not possible in most of the time [6]. Healthcare demands agile development methods for analytics with an ultimate enterprise data view once integrated together. Inter related dimensional data models or data marts, eventually leading to an enterprise DW is one of the widely accepted approach in healthcare [7]. There exists a lot of studies and models specific to healthcare data management, but most of them concentrate on the structured financial data of the business. Others concentrate on the clinical outcome mode. But to have a complete view of the business a combined view is essential.

In this research, authors are trying to develop a “dimensional data model” for the most sought-after administrative data section of care delivery, which is the “Revenue Cycle Management process”, in an agile approach.

This paper is organized in the following manner. We will start with the analysis of the RCM processes, which will lead to the analytics requirements of the domain. These requirements will facilitate the findings of the facts and dimensions needed for the design of the dimensional model. The model will be capable of agile modelling and will accommodate late arriving requirements as well. The key performance indicators will be discussed, and few Business Intelligence (BI) dashboards will be touch based in the last section.

II. RCM PROCESSES

RCM is the term used to refer the group of processes involved in patient services, which involve the clinical and administrative tasks performed for capturing, managing and collecting the patient information and the associated financial charges. In short, it is the umbrella term used to refer the complete life cycle of a patient account right from registration to bill settlement [8].
RCM is not an independent activity in care delivery. It starts from the moment when a patient calls for an appointment and completes the process cycle once there is no balance on that patient's account or once the provider receives or reconcile the claimed amount. Data flow should happen seamlessly across the EMR (Electronic Medical record), LIS (Laboratory Information System) and RCM. There are many providers who opt the RCM part to be outsources to the RCM expert firms. Irrespective of who handles the process, the provider or third-party organization, analytics of RCM data is vital for prediction and to have a better management of AR (Accounts Receivable), Payer performance evaluation etc. This provides better improvement opportunity for the organization to identify areas for improvement and revenue increase [9].
In a broader perspective, we can categorize the RCM activities into 3 categories, at least for the analytical purpose.
– Front office tasks (Appointments, Insurance eligibility check, Pre-Authorization, Verification etc.)
– Clinical care tasks
– Back office tasks (Claim submission, payments, Denial management, Reconciliation etc.) [10]
A well collaborated front office to back office activities will improve the RCM tremendously and effective communication and holistic data view are essential for a well-managed RCM. Fig. 1 shows the life cycle of the value circle of an RCM stream.

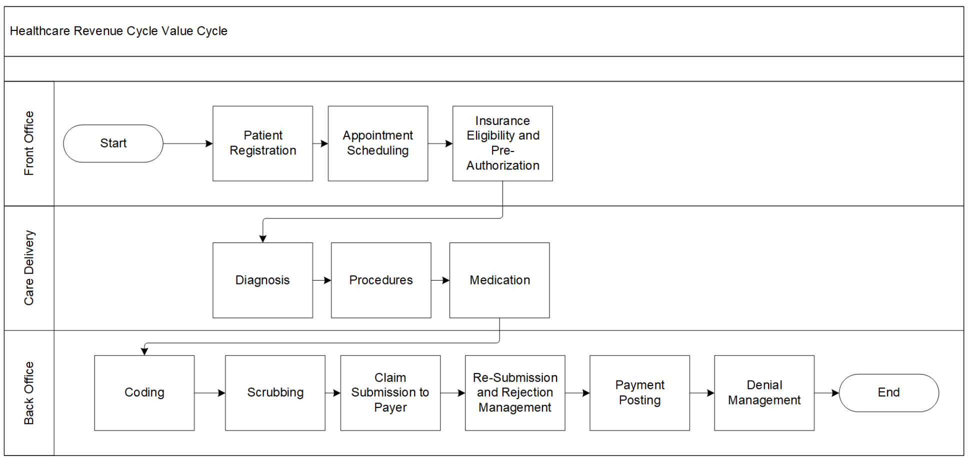

Fig. 1. RCM Life cycle.

III. MATERIALS AND METHODS

Identification of RCM Analytics Flow: Healthcare analytics needs are agile in nature. Kimball's proposal of agile Dimensional Modelling will be applied in this research work. Kimball proposes a 4 steps theory for the dimensional modeling [11]. The steps are namely the identification of the business processes, define the grain of the data to be modeled, identification of the dimensions for analysis and establishment of the facts based on the processes and the data grain. This research work is based on the Kimball theorem of Bottom Up Approach with few fine tunings.
Step #1 Business Process Identification: A process is a set of activities well-coordinated to achieve a specific objective [12]. Let us use this definition to identify the business process in an RCM life cycle [13]. In an implementation perspective, the business processes will be finally fine tuned into "Facts" and the depending angles will be turned into "Confirmed Dimensions" [14]. Below mentioned processes are based on a generic approach and it could vary upon the business nature and the environment a provider operates.
Process1 – Patient Episode (Encounter) Flow: This process includes the activities right from the appointment scheduling till the closure of the visit. This process is dependent on many factors and actors. The encounter process has a significant role in statistical analysis of volume and capacity management [15]. This also gives light into the behavior of the patient, on no-show perspective. Encounter is one of the common process across the healthcare analytics spectrum, not only in RCM. So, when we constitute an enterprise healthcare DW, encounter needs to be treated in a special way. Dependency of encounter with different dimensions are illustrated in the Fact-Dimension matrix below (Fig. 2).
Process 2 - Care Delivery Flow: Clinical Care Delivery process refers to the activities staring from consultation till the betterment of the patients. This includes the nurse assessment, physician consultation, procedures, medications etc. Every patient consultation will have (mostly) diagnosis associated with it, followed by the proposed care items(procedures). This will lead to the activities of the procedure administration and then the pharmacy activities for medication delivery. In RCM analytics every penny received or to be received need to be analyzed and mapped to clinical context.
Process 3 – Billing Flow: This includes the activities of patient billing, claim submission, payments, denial management, collections and reconciliation. There will be a contextual change in the way how we do the activities for inpatients and outpatients. Based on the health authorities' guidelines and country specifications, there could be variation in the claim submissions and receivable processes. We are considering a generic common approach in this case.

Step # 2 Define the Grain: Grain refers to the granularity of the data to be loaded into the model. Granularity depends on the analytics requirements and the extends to which the users drill down the data [16]. Revenue analysis can be stopped at the level of a bill, but when we need to have a deep dive into the individual activities of a bill, we might need to increase the granularity to the line item(activity) level [11]. In the proposed model, we are modelling both options to extend the reach of the analytics (transaction grain as well as periodic accumulating snapshot [7])

Step #3 Identify the Dimensions: The value chain of Healthcare is revolving around a set of fixed dimensions and on top of that there comes the late arriving once [17]. In this proposed research, we are concentrating only on the dimensions pertaining to the Revenue Cycle. Listed below the prominent dimensions for the RCM analysis [9].

– Date and Time
– Patient
– Doctor
– Staff
– Diagnoses
– Procedures
– Care Providing facility
– Payer
– Health plan

In healthcare analytics, there exists many more dimensions, but the above list shows only the dimensions specific to RCM.

Step #4 Identify the Facts: Based on the process analysis, its cleat that the analysis can be divided into 3 categories. Encounter specific, treatment specific and post treatment specific.

– Patient volume, no shows, departmental and individual performances are examples of the analysis coming under the encounter specific category.

– Analysis based on DRG, ICD, CPT, LOS are examples of the analysis of treatment specific.

– Billing and revenue specific examples are AR Aging, Contractual Variance, Net Collection Rate, Claim Denial Rates, Claim Rejection Rates, Collection Analysis etc.

Based on the above analytics requirements, proposed Facts are:

Claim Flow = This Fact will record every details of a claim in a consolidated snapshot fashion

Claim Activity Flow = This is bit more granular level of the above Fact. This fact will capture all details of the activities included in specific claims. This will be the most granular atomic level fact. The above fact can be treated as aggregate or generated fact from this one.

Encounter Fact = This is more or less a "Fact less Fact" representing an event, which is the foot fall of a patient at the facility.

Collection Flow = This fact will be recording the details of the collection processes of RCM events.

IV. RESULTS AND DISCUSSION

Kimball's four step methodology together with activity theory is used for the identification of the Facts and Confirmed dimensions for the subject of analysis [17]. The correlation between the resulting Facts and Dimensions are illustrated in the Fig. 2. RCM is very dynamic in nature and based on the change in the business rules and the regulations, there are possibilities of late arriving dimensions [18]. The identified dimensions will be mostly shared between multiple dimensional models, which constitute the enterprise data model and thus provides an agile environment.

The relation (star model) between these facts and the dimensions are illustrated in the Fig. 3. Detailed field list of each fact is also narrated in the same figure.

The design of the Dimensional Model: When we design the dimension model in an RDBMS it will result in a Star-Schema and when it is implemented in a multi-dimensional database will result in OLAP cubes. Here, for the representation purpose, the model is illustrated as a star schema. In the below design, Dimension tables are listed just with names, not with field list. Fact tables are provided with extensive field list too. To avoid cluttering of the diagram, connection lines (primary key – foreign key relation) are not added. Please refer Fig. 2 for connection interpretation.

The Date Dimension is referenced by all Facts. Let us take the example of Bill Activity Flow Fact. The same fact is referring the Date Dimension in multiple times on the same instance but on different values. For example, the date of encounter, date of claim submission they both refer the same Date dimension. Here Date dimension is used as a "Role Play" mode. To tackle this situation, we will make different Views of the same Date Table with appropriate names for easy reporting and joining operations across tables. The same case is happening in the case of admitting and discharging DRGs as well.

Technology used and the data flow diagram: For this research purpose Microsoft Platform is used. SSIS is used for ETL and Power BI for Visualization. Both staging database and DW are on SQL Server.

		Dimensions												
Process Area	Facts	Date	Invoice	Patient	Doctor	Other Staff	Provider	Payer	Procedure	HealthPlan	Medication	Diagnoses	IRDRG	Department
Front Office	Encounter	1	0	1	1	1	1	1	0	1	0	0	0	1
Backoffice	Bill Flow	1	1	1	1	1	1	1	0	1	0	1	1	1
	Bill Activity Flow	1	1	1	1	1	1	1	1	1	0	0	0	0
	Payment Flow	1	1	1	0	0	1	1	0	0	0	0	0	1
	Payment Activity Flow	1	1	1	0	0	0	1	1	0	0	0	0	0

Fig. 2. Facts and Dimensions relativity matrix for RCM.

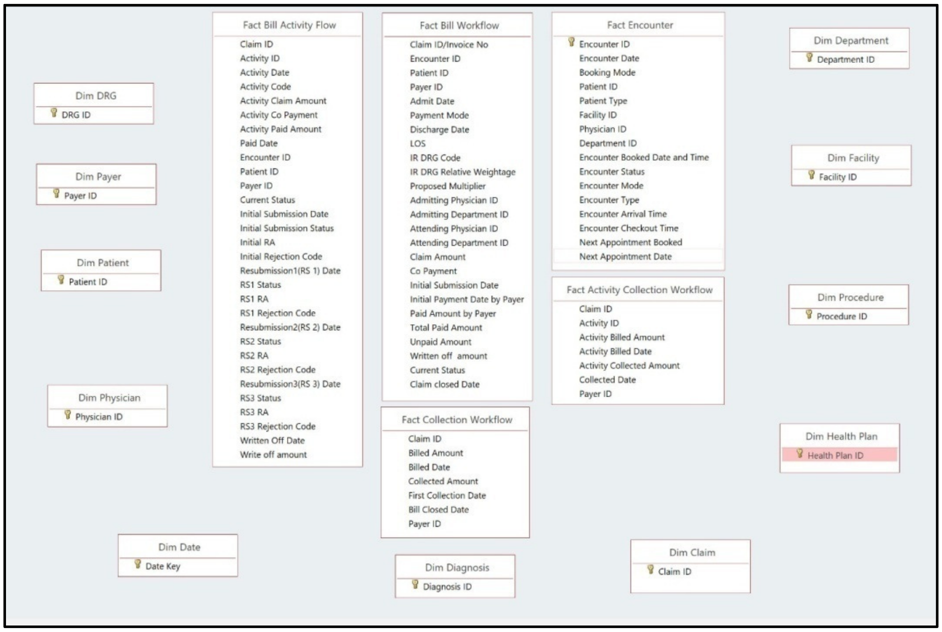

Fig. 3. Dimensional model – Star schema for RCM.

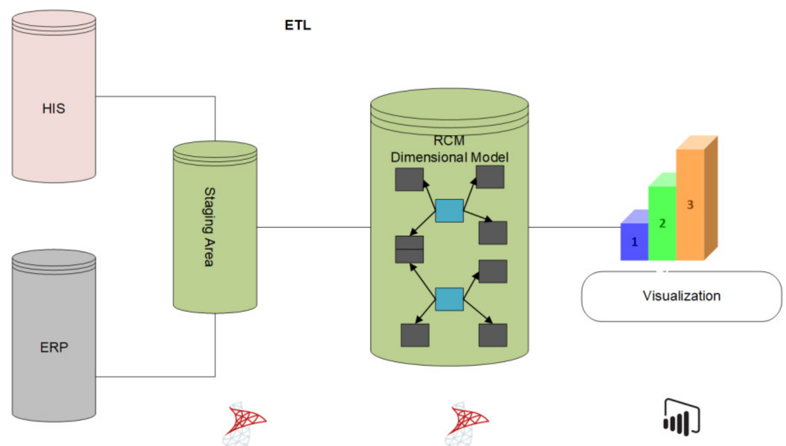

Fig. 4. Data flow representation.

Resulting Dashboards based on the KPIs: A successful DW and BI project concludes its life cycle with a well-defined presentation layer capable of self BI [4]. This should be capable of drilling down to any granular level as well as rolling up to holistic data view. This is the feature, which distinguish BI with traditional reports. As illustrated in Fig. 1, RCM processes fall under 3 major categories. Table 1 list out major KPIs and visualization scenarios built on top of the resulted dimension model of this work [19]. Table 1 shows only the limited no of parameters and the same shall be extended to various parameters based on the business requirements and the data accommodated in the Facts and Dimensions.

Figs. 5-6 shows various BI figures made from the proposed dimension model. The visualizations are created using dummy data for compliance purpose.

Table 1: Key Performance Indicators of RCM process.

Areas of interest		
Patient Encounter	**Treatment**	**Backoffice**
Patient's visit trend	Top DRGs	Revenue/Department/Physician
No Shows	Admitting complaints	Revenue/Facility
Follow up visits	Length of stay	Insurance Rejection Rate
Departmental OP volume	Inpatients/Departments	Rejection Reason
ER analytics		AR specific
Appointment mode		Payer Analysis
ER visit spread – Day/Hour		Collection Specific
Volume per physician		Denial specific
		Written off cases
		Contractual Specific

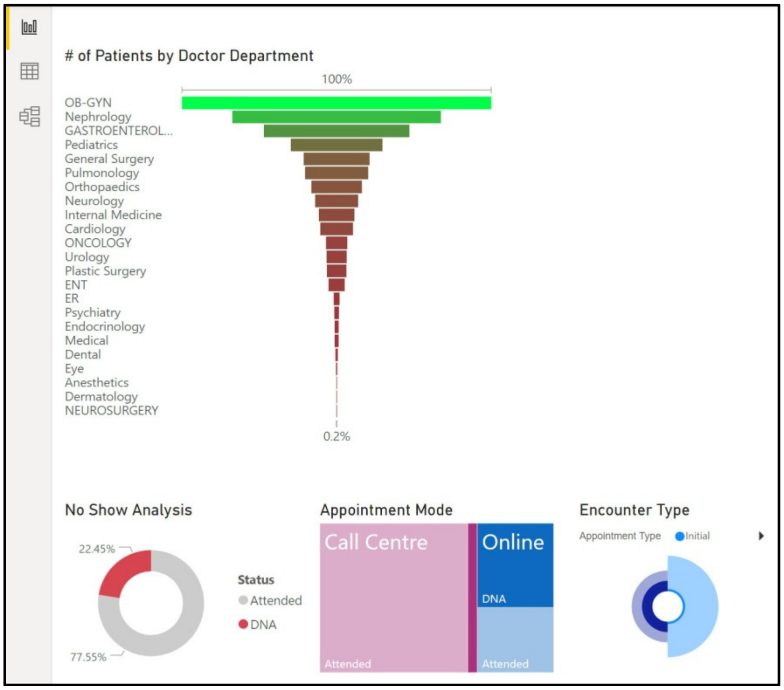

Fig. 5. Sample Dashboard – Encounter and Front office specific.

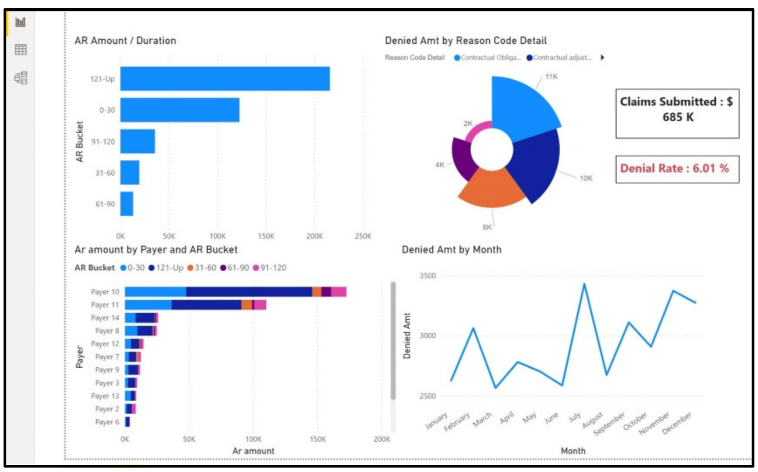

Fig. 6. Sample Dashboard – Back office specific.

Discussions: Timely availability of accurate data is the most important factor for quick decisions. With the implementation of the data warehouse and the associated BI dashboards, the finance and the revenue team has access to the consolidated and the meaningful data on their fingertip [20]. With this solution, there can be an estimated improvement of 95+% in the time to access the data. The traditional reports, which are static in nature, doesn't offer the facility of drill down and roll up. Whenever detailed information on any report is needed, it will lead to another report development. But the proposed solution, which is based on DW and Visualization features, has the ability to drill down and roll up to arrive to the root cause of any topic of interest.

Having said that, the ability to drill down and roll up will improved the root cause analysis activities and the expected reduction of time is 90+%. Let us take the example of AR. When the RCM team identifies dip in collection of payments from the payer, they normally run a handful of reports and will run through it in multiple times to arrive the conclusion. But in this scenario, referring Figs. 5 and 6, different AR bucket and the aging are enabled with drill down and the same is linked with the payers [21]. A high rejection rate value on a specific quarter can be investigated through the contributing factors which are in fact the dimensions linked to it.

A high rejection rate could be specific to a specific set of services offered at the provider. By analyzing the rejection reasons of those set of services, it can lead us to valuable conclusion such as training deficiency, process variation etc.

A simple example of the above, while testing the model with sample data is as follows. One of the physicians used to take much higher time with patients than other counterparts and still the rejection rate was high for his patients. A drill down activities revealed the fact that, the documentation methods used by the physician was entirely different from others and the best practices of documentations was not incorporated [22].

This led to the training deficiency indication. This physician was never utilizing the "favorites" features of the HIS and used to code every item one by one including medicine prescription.

Coming to the design of the model, it is modular in nature. This means the same Data Ware environment created to host the proposed Dimensional model will be used when new subjects of analysis arise. For example, when a new analysis area such as "Emergency Department (ED)" arise, the only thing we need to do is to identify the specific facts pertaining to ED. Most of the dimensions used in RCM will be reused by ED as well. The common dimensions will be shared across any number of data marts and facts, to avoid data redundancy and to enable data integrity. As and when new additions happen to the DW, ETL process needs to be adjusted and the meta data also to updated accordingly.

V. CONCLUSION AND FUTURE SCOPE

The authors are working on multiple subject areas in healthcare domain starting with RCM and the resulting models will be integrated each other. This integration finally will lead to the holistic data view and a comprehensive healthcare DW. This way the complexity of the entire project can be reduced, and the return of investment also can be materialized. On top of that, the lessons learned from every phase can be applied in the coming activities.

ACKNOWLEDGEMENTS

We would like to thank Mr. Ramachandran and Mr. Gopakumar for their support and motivation throughout the research work.

Conflict of Interest. The authors confirm that there are no known conflicts of interest associated with this paper.

REFERENCES

[1]. Ponniah, P. (2001). Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. 605 Third Avenue, New York: John Wiley & Sons.

[2]. Nelson, G. S., Technologies, T., & Hill, C. (2017). A Practical Guide to Healthcare Data : Tips , traps and techniques. *Thinking Data, 1*(1), 1–20.

[3]. Karami, M., Rahimi, A., & Shahmirzadi, A. H. (2017). Clinical Data Warehouse. *The Health Care Manager, 36* (4), 380–384.

[4]. Medrian-Medical_Management (2015). Business Intelligence in Healthcare : Strategies to Ensure Financial Success. *A Meridian White Paper, 1*(2), 1–5.

[5]. Inmon, W. H. (2002). Building the data warehouse. 3rd ed. 605 Third Avenue, New York: Wiley Computer Publishing.

[6]. Inmon, W. H., & Neushloss, D. S.(2008). DW 2.0. Burlington, Massachusetts: Morgan Kaufmann.

[7]. Kimball, R, M. (2016). The Kimball Group Reader. Crosspoint Boulevard Indianapolis, IN: John Wiley & Sons.

[8]. (n.d.). Optimizing the Healthcare Billing Cycle with RCM Analytics | Tableau Conference 2019.

[9]. Green, M. A. C. R. (2018). A Guide to Billing and Reimbursement. 14th ed. Clifton Park, NY: Delmar Cengage Learning.

[10]. Visscher, S. L., Naessens, J. M., Yawn, B. P., Reinalda, M. S., Anderson, S. S., & Borah, B. J. (2017). Developing a standardized healthcare cost data warehouse. *BMC health services research*.

[11]. Kimball, R. M. (2002). The Data Warehouse Toolkit - Second Edition. 2nd ed. 605 Third Avenue, New York: Wiley Computer Publishing.

[12]. AXELOS (2019). ITIL Foundation 4th Edition. 4th ed. Berners Street,London: Axelos.

[13]. Singh, R., Mindel, V., & Mathiassen, L. (2017). IT-Enabled Revenue Cycle Transformation in Resource-Constrained Hospitals: A Collaborative Digital Options Inquiry. *Journal of Management Information Systems, 34*(3), 695–726.

[14]. AlHazme, R. H., Rana, A. M., & De Lucca, M. (2014). Development and Implementation of a Clinical and Business Intelligence System for the Florida Health Data Warehouse. *Online Journal of Public Health Informatics*, *6* (2), 8–11.

[15]. Mugdh, M. & Pilla, S. (2012). Revenue cycle optimization in health care institutions: A conceptual framework for change management. *Health Care Manager, 31*(1), 75–80.

[16]. Inmon, Claudia, I. (2001). Corporate Information Factory. 2nd ed. 605 Third Avenue, New York: John Wiley & Sons.

[17]. Gao, Y. (2006). Dimension Identification in Data Warehouse Based on Activity Theory Dimension Identification in Data, *2*(1), 1–49.

[18]. Dale Sanders (2014). The Late-Binding TM Data Warehouse Explained. Salt Lake City.

[19]. Cheyenne Brinson, MBA, C. (2018). Six fundamental revenue cycle management metrics.

[20]. Capocchi, A. (2019). Economic Value and Revenue Management Systems. *Economic Value and Revenue Management Systems,* 41–66.

[21]. Singh, S. R., & Wheeler, J. (2012). Hospital financial management: what is the link between revenue cycle management, profitability, and not-for-profit hospitals' ability to grow equity?. *Journal of Healthcare Management*, *57*(5), 325-341.

[22]. Rauscher, S., & Wheeler, J. R. C. (2008). Effective Hospital Revenue Cycle Management: Is There a Trade-off Between the Amount of Patient Revenue and the Speed of Revenue Collection ? *Journal of Healthcare Management, 53*(6), 392–404.

How to cite this article: George, J. and Jeyakumar, M.K. (2020). A Dimensional Data Model for Healthcare Revenue Cycle Analytics and Business Intelligence. *International Journal on Emerging Technologies*, *11*(3): 648–653.