

A Systematic Review on Existing Data Dictionary

Anshul Mishra¹, Dr. Devendra Agarwal², Dr. M. H. Khan³

Research Scholar, School of Computer Application, BBDU, Lucknow, India¹

Director (Engg.) at BBDNIIT, BBDU, Lucknow, India²

Professor, Department of C.S. E., I.E.T., Lucknow, India³

Abstract: Data dictionary is an important factor for software development process and facilitates to learn the product or complete description of data object, object relationship and analysing a System. The data dictionary uses as a tool in software engineering for documentation, design and maintenance to software system. The system stores the structures, procedures, description and cross referencing about data model of project. Researchers and Experts have always claimed that data dictionary should be consider as an important factors for it have positive impact on system analysis and design in order to improve the quality of software system. This paper presents the results of a systematic review conducted to collect evidence on exiting data dictionary of system development process. In this review paper initially our aim is to find the detail view of existing data dictionary and their contribution during the software development process at initial stage. Finally paper demonstrates an existing data dictionary proposed by various experts and researchers including their contribution and limitations.

Keywords: Data Dictionary, Software Development Life Cycle, Meta Data, Repository System, Dictionary Attributes.

I. INTRODUCTION

Software development life cycle is a complex process. It has to deal with uncertain no of events, activity and causes during the development of software. Many researchers have tried to minimizing errors, detecting faults that do occur before the delivery of software. A software development life cycle thus goes through a number of basics fundamentals phases as planning, analysis, design, coding, and testing implementations. Every phases also consists a series of steps that required for the delivered the product. Mainly attention the structure approach consist the set of activity, tools as data flow diagram and data dictionary contains information about the system component [11]. An important part of any software process is the definitions of data structures, data concept, and relationship and addressing mode, they introduce the data dictionary concept. Many software process areas like environment planning, documentations, analysis, design and description of data are very important subject to get good result [8, 9]. The project manager can use the data dictionary; know about the modelling registry, vocabulary one source and repository of Meta data. The data dictionary has been portioned into different views, each characterised either by information contained in it, by its aggregation state, by its differentiated users or by users it is relevant [12]. Data dictionary always supports the software development process and facilitates the creation of better quality software. A system analysis tools provides a complete description of the logic and a cross references between the code and their specification [2].

Structure analysis concepts are powerful to express the well structured project and users needs. In general format data dictionary can decide about the structure and other

features of the whole domain concepts which are instances of the Meta classes also guide the acquisition process for the domain knowledge [1]. Moreover dictionary has been always elusive concept and its correct structure or complete design is difficult exercise.

It has been inferred from the literature survey on data dictionary that there is a heavy need to provide the level of project manager responsibility during software development phases. Hence researcher's attentions to add more features and functionalities as factors at initial stage in existing data dictionary for qualitative improvement.

II. DATA DICTIONARY

Software The data dictionary has become an important concept for many software development processes. A data dictionary is an essential part of the software system that is used to understand the data, data relationship and critical databases that contain it. The data dictionary is a key aspect to identify the Meta data, their description and units of measure which help to provide the relevant information about the software system. IBM glossary defines data dictionary as a centralized repository information system to relate about data, data origin and format of data [3].

This types of theory organized and structured as a set of data dictionary entities considered as primary or primitive entities and it is represented as a list of key attributes or fields clustered by each of the kind of data set received during the validation of service and the collection of data which are related to the different layer of validation concept [17].

According to ISO data dictionary as a reference book contains the accurate, reliable, controllable and verifiable data in data bases and user or analyst easily understands about data [14]. Many data dictionary definitions are given by experts. Gundebahar and Kus defines dictionary as handled everything about the project and worked as duty [5]. The data dictionary concept continuously improved the data management process at every phase of software development process. It developed the protocol of reference book or guaranteed for provide the help to understand accurate meaning used attributes in software process.

The data dictionary in software development process to include the data bases properties, schema and cross references. It is very hard to develop dictionary that contains all the success keys, concepts and measurements. According to Brackett dictionary information shall be included as data names, structures, data types and their formats, also include the logical view structure, data definitions which help to people for understand the data resources and supports activities of software systems [13, 16].

Structure analysis and design used as tool for listing of all the data elements, pertinent to the software system with precise rigorous meaning so that project manager have a common understanding about input output elements and their calculations. For complex project face the different quite difficulties, which time data dictionary can be treated as a data bases and full support to the developer at the level of software development process. The developer team can ask the different types of questions as “where used/ how used”, “what is location”, and “what is relation” and get answer from reference book. Experts frequently advocate that dictionary is a important consideration in software system. The most obvious is documentation; it is valuable references in any organization. The advantages are improving analyst or user communication by establishing consistent definitions of various element terms and procedures [10, 27].

Finally we can say that data dictionary is play important role in the development of data bases. The data base system has contained a maximum number of dictionaries as a standard features.

III.RELATED WORK

A data dictionary is an important excellent initial point for developing the software system. The complete domain of dictionary may be easily validated and transformed into different helpful output values. A large collection of project information is called a repository. A number of researchers accepted that analyst have created repository that may contains information about data, flows, store, records and structure of elements. A dictionary could increases automations activity of modelling, validation and auto coding to other stage [26].

A repository have also contains about project requirements details, final system deliverables information and complete information about project. The data dictionary has created at all phase as analysis, design and ultimately

documentation, due to these theory is authoritative source on how a data element is used and explained the element by the programmer in software system. Many difficult software systems uses computerised data dictionary for cross referencing all modules in the databases.

Data dictionary approaches is like a catalogue of all description of data, a single point reference of data repository and also maintain a system [6]. Data base structure may be classified as (a) Data element location (b) Data structure (c) Availability (d) Usage. These types of data structure are called the data dictionary and it is most useful tool, throughout a system development life cycle process [4, 7].

Data dictionary system has been an important research direction since 1970s with associated the data bases system as a central goal [7]. Initial software techniques an automated facility, that supports data functionality as resources context is the data dictionary. The term commonly uses as “data dictionary”, “system resources dictionary”, “directory of data definitions”, the different stage of software life cycle is a equal. To know and obtained the complete information about software system is very important issues at all stages of the software development life cycle. The analyst work facilitate to project team by providing them a summarized and conceptual system dictionary attributes that will deal with high conceptual key and summarized information regarding their developed software system and the utilized process of software [18,20].

Data dictionary uses reduces the efforts at the every stages of software development life cycle can be visualized as follows table 1 [5]. Data dictionary system increases the effectiveness and growth rate at every stage.

Table 1: [5] Impact of Data dictionary on SDLC steps

SDLC Steps	Elapsed Duration in Hours		
	With proposed Data dictionary	With passive Data dictionary	Differ ences %
Analysis	68	64	60
Development	72	81	12.5
Test	36	40	11.11
Total	166	185	10.11

Following sections systematically summarize some of the relevant important efforts made by researchers in this area.

In 2004, Karl [22] defined the Class dictionary in view of programming terminology. It have contains all the data structure which presents in a program as well as language for describing data specified by the data structures. Class dictionary makes as links between the language design, structure design and database design. The experts have primary objective of research work is improve the dictionary content and provide the help to improve the quality of programmers productivity. The objectives are summarizes as: (a) Class dictionary have contains useful

code (b) Provide facilitate to data structure, parameter and context free structure (c) Addition the Supporting parameters for improve the code. These approaches only contribute modular software efficiently.

In 2005, Paolo Atzeni, Cappellari and A. Bernstein [21] introduced multilevel dictionary approaches associated with Meta data model and schema. It describes the schema level or model level that provides a detailed description of schemas. The structure of data dictionary uses a tool under software development. It supports a variety of activity in development and reporting. The research work explains that dictionary is composed of four parts based on the contribution of two most popular features: (1) Schema level (2) Model Level. He mainly focused only on schemas generic or model generic applications.

In 2007, Martin Modell [28] proposed conceptual model of automated dictionary documentation categories and relationships. Model associated with the detailed description about software system.

1. Project content
2. Functional attributes
3. A set of process
4. A number of models
5. A number of relationships with entities

Data dictionary model should be included the critical success factors and process model components but this model has provided only the initial level help to project consultancy team.

In 2008, Ellen and Ebg [19] provided a requirement model road map that helps software development team under the efficient uses of requirements attributes at every stages of software development life cycle. This model ensures identify to policy allocated for business people, which in turn allows top level management to prepare for implementation level by updating procedures and other assets needed to enforce the policies and also defining the each data element, its data type, length, and format. It provides helps to project teams for use only data modelling tools and data dictionary applications.

In 2012, Feichiang, Periklis and Miller [9] highlighted an attribute dictionary is a set of attributes together with a set of common values of each attribute understanding unstructured or loosely structured textual descriptions of entity collections, such as product catalogs. Dictionaries provide the supervised data for learning product or entity descriptions. It presents AutoDict, a system that analyzes input data records, and discovers high quality dictionaries using information theoretic techniques. Finally Chiang's concept used only the automatic dictionary not in generalized form.

In 2012, M.Gundebahar and M.C. Kus Khalilov [5] proposed an active, automatic and composite data dictionary system for managing the complete structures and descriptions of both data bases system and applications objects including system configuration data

set. This system expedites obtaining information about data bases and applications objects, their relations. As a result it is observed that it improve the performance of software team members. However this hypothesis was not validated

In 2013, Chikh and Al. Qershi [24] highlighted to elucidate how we can provide a set of more powerful software requirements targeted to facilitate their retrieval using the Meta data knowledge. This work is define the a new concept of software requirement object as combining two kinds of knowledge as (a) Requirement Knowledge (b) Meta data knowledge. A new Metadata language is proposed in order to qualify, retrieve and reuse requirements. It has design the large scale repository system which contains the many data elements and their explanations. In this work data dictionary have been supported only requirement phase.

In 2014, Xio, Shiyong Zhi, Wu Shan and Jin Liu [25] had done a great work showing improve the quality of software system using the data dictionary learning techniques. Software development process uses the dictionary learning technique to predict software defect. It can fully exploit the class information of historical data to improve the discriminative power and provides effective solutions for the major problems in the field of software defect prediction, which are misclassification cost and class-imbalance problems. The techniques have given the solution only the unlabeled software modules.

In 2015, Biji. K. P. [23] highlighted the role of data dictionary in the data base management system, basis on structured query language or commands. It has perposed a solution for detection using data dictionary and prevention intrusion search along with ASCII values. The data dictionary can store total number of parameter and queries in applications and make a list to compare with real time generated value of parameters. The structured query language accepted the set of commands and totally depends on relationship between the data element and appropriate records. A record and relationship between the data, artefacts of software will be associated in the data dictionary, but not explain the accurate concept about data dictionary attributes and their role.

In 2015, Biku Abraham, Paul and Nebu John [15] explored an innovative idea for compress clustering data. The idea generates the data dictionary for clustering techniques. Basically dictionary helpful in two levels as (a) Primary level (b) Secondary level. It is a process of grouping similar object together. The experts explain or design the data dictionary in different perspective as follows (a) For clustering approaches (b) For compression techniques (c) For data set calculation. He focused only clustering and compression approaches dictionary.

A Survey of data dictionary shows that how do the helps to software development life cycle process. On the other hand recent development in software development process the data dictionary will provide the supports to improve quality of software and complete the user requirements.

IV. SUGGESTION

After successful completion of the systematic literature review some important critical observation are as follows. If we enhance the data dictionary at initial phase of software development process may greatly supports to software system and as well as help to project manager.

1. In order to add more features and functionalities such as security factors, quality factors, cost factors, risk factors and schedule factors much more in data dictionary at different phase of software development life cycle which have positive impact on project manager responsibilities.
2. A factors affecting to software system must be identified and then the set of factors relevant at the software development phase should be finalized.
3. Further, the no of affecting factors must be selected in software system then matching with the responsibility of project manager after includes the reference book or data dictionary.

V. CONCLUSION

Several approaches or application have been proposed in the literature for supporting the data dictionary at every stage of software development life cycle. A survey of the relevant literature shows that maximum efforts have been put at designing, maintaining, and operating to software systems need, for their solution and maximum number of accurate data which are common to many of them. The availability of a data dictionary with complete description is a very useful in the development and maintenance of software system. On the other hand the lack of data dictionary at every stage may not be compensated during subsequent development process. The above discussion our conclusion is that dictionary is an important factors that attempts to predict that how much enhance property of dictionary will be required to improve the software quality and help to project team.

REFERENCES

[1] Jurgen Jurgen Angele, Dieter Fensel and Rudi Studer, "What Could the Knowledge Engineer Learn from the Software Engineer?", Long Version, June 2003.

[2] Jussi Koskinen and Sami Kollanus, "Survey of Software Inspection Research (1991-2005)", Computer Science and Information System Reports, March 2007.

[3] ACM Digital Library, "IBM Dictionary of Computing", 10th Edition, 1993.

[4] Fabio A. Schreiber, G. Martella, "Creating a Conceptual Model of a data dictionary for Distributed Data Bases, ACM Sigmis Data Bases, Vol.11, pp 12-18, 1979.

[5] M. Gundebahar and M. C. Kuskhalilov, "An Active and Composite Data dictionary System", International Conference on Intelligent Computational System, Jan 2012.

[6] V. Raja Raman, "Analysis and Design of Information System", PNI Learning Private Limited, New Delhi, 2nd Edition, July 2011.

[7] Christos Drakoulis, "Data Dicionary / Directory System, A tool in Software Development System", Thesis Title, Naval PG School California, Sep 1985.

[8] M.Erdogan, A torun and D. Boyaci, "Revisiting the Procedure for the Vector Data Quality Assurance in Practice", International

Archives of the Photogrammetry Remote Sensing and Sapatial Information Sciences, Vol. XXXIX-34, 2012.

[9] Feichiang, Periklis, Zhuerkank and Miller, "Auto Dic : Automated Dictionary Discovery", IEEE 28th International Conference on Data Engineering, 2012

[10] Elias M. Awad, "System Analysis and Design", Galgotia Publication, Second Edition, 2008.

[11] A.Rob, "Issues of Structured Vs Object-Oriented Methodology of System Analysis and Design", Issues of Information System, Vol.5, No.1, 2004.

[12] Fabio A. Schreiber and G. Martella, "Creating A Conceptual Model of A Data Dictionary for Distributed Data Bases", ACM SIGMIS, Vol. 11, Issue 1, pp 12-18, 1979.

[13] B. Avenue, "Understanding Meta data", National Information Standard Organization, 2 April 2007.

[14] International Standard Organization, "Information technology part-1", 2nd Edition, 2004.

[15] Biku Abraham, Varghese Paul and Nebu John, "Dictionary Based Behavioural Data Compression: A clustering Approach", Middle East Journal of Scientific Research, pp 214-218, 2015.

[16] M. H. Brackett, "Data Resources Quality", Addition-Wesley Information Technology Series, 2000.

[17] Said Rahma, "Data Dictionary and Data Collection Summary Report", European Commission, DG Information Society and Media, 2010.

[18] Dumke, R., Blazey, M., Reitz, D., and Karsten, " Causalities in Software Process Management and Improvement" , Proceeding of the International Conference on Software Process and Product Management, Nov 2006.

[19] Ellen Gottesdiener and Ebg Consulting, "Good Practices for developing user requirements", The Journal of Defence Software engineering, March 2008.

[20] Abdullah, Dr, Reena Srivastava, and M. H. Khan. "Modifiability: A Key Factor To Testability". International Journal of Advanced Information Science and Technology, Vol. 26, No.26, Pages 62-71, June 2014.

[21] Paolo Atzeni, Paolo Cappellari and A. Bernstein, "A Multilevel Dictionary for Model Management", Research Gate, October 2005.

[22] Karl J. Lieberherr, "Object Oriented Programming with Class Dictionary", LISP and Symbolic Computation: An International Journal, Vol. 1, No. 2, 2004.

[23] Biji K. P., "Data Dictionary Based Mechanism against SQL Injection Attacks", International Journal of Engineering and Computer Science, Vol 4, Issue 6, pp 12453-12457, June 2015.

[24] A. Chikh and F.Al- Qershi, "A new Meta Data Language for Software Requirements", Lecture Notes on Software engineering, Vol.1 No. 3, Aug 2013.

[25] Xio, Shiyong Zhi Wu Shan and Jin Liu, "Dictionary Learning Based Software Defect Prediction, 36th International Conference on Software Engineering, Hyderabad, 2014.

[26] Vitkin, L. and Fallahi, A., "The Role of the Data Dictionary in the Model-Based Development Process," SAE Technical Paper, 2009, doi: 10.4271/2009-01-0919.

[27] Abdullah, Dr, M. H. Khan, and Reena Srivastava. "Flexibility: A Key Factor To Testability", International Journal of Software Engineering & Applications (IJSEA), Vol.6, No.1, January 2015. DOI: 10.5121/ijsea.2015.6108

[28] Martin Modell, "The Data Dictionary", A professional's Guide to system analysis, Second Edition, 2007.

BIOGRAPHIES



Anshul received the MCA degree from Dr. R. M. L. Avadh University, Faizabad, in 2008. He is currently working as an Associate Professor, in the Department of Computer Application, at Institute of Environment and Management, Lucknow. His research interests include Software testability, Software Quality Estimation, Data Dictionary.



Dr. Devendera Agarwal is currently working as Prof. & Director (Engg.) at BBDNIIT (BBD Group), Lucknow. He has over 17 years of teaching & 5 years of industrial experience. He has done his B.Tech in Computer Science from Mangalore University in 1993, M.Tech from U.P. Technical University, Lucknow in 2006, and Ph.D. from Shobhit University, Meerut in 2013. He has over 10 research papers with 4 students pursuing Ph.D. under his guidance.



Dr. M. H. Khan, Professor, Department of Computer Science and Engineering at IET Lucknow UP. Obtained his MCA degree from Aligarh Muslim University (Central University) in 1991. Later he did his PhD from Lucknow University. He has around 28 years rich teaching experience at UG and PG level. His area of research is Software Engineering. Dr. Khan published numerous articles, several papers in the National and International Journals and conference proceedings.