Enterprise Architecture Model that Enables to Search for Patterns of Statistical Information

Oleg Chertov

Applied Mathematics Department
NTUU "Kyiv Polytechnic Institute"
Kyiv, Ukraine

Abstract—**Enterprise architecture is the stem from which developing of any departmental information system should grow and around which it should revolve. In the paper, a fragment of an enterprise architecture model is built using ArchiMate language. This fragment enables to search for information in enterprises which do not work in productive industry. Such enterprises include official statistics. The proposed model embraces all three architectural levels of corresponding information systems, namely, OLTP, OLAP, and Data Mining. Particularly, the latter level enables to search for patterns of statistical information.**

Keywords—Enterprise Architecture; Pattern Search; Data Mining; Architecture Description Language

I. INTRODUCTION

Architectural issues which arise during developing complex information systems play the role of the same importance as those arising during construction of an original building. Correct architectural decisions considerably lower the risks of the whole project of system developing and maintenance, because they make it possible to efficiently use existent infrastructure and optimally plan its further progress.

At first, the term "architecture" was used in the field of IT only in relation to hardware. Later on, this term was used in relation to information systems as a whole. Only with the lapse of time it became clear that it is necessary to apply systems approach not only to developing the information system, but also to developing the whole enterprise.

As a result, term "enterprise architecture" emerged. For the first time, it was used in the report [1]. This terminology assumes relatively wide treatment of the concept of "enterprise." In particular, the concept can be applied to the architecture of state organizations and offices.

Among various definitions of enterprise architecture, we will primarily use the one proposed by Global Enterprise Architecture Organization (GEAO): "The way in which an enterprise vision is expressed in the structure and dynamics of an enterprise.

It provides, on various architecture abstraction levels, a coherent set of models, principles, guidelines, and policies, used for the translation, alignment, and evolution of the systems that exist within the scope and context of an Enterprise" [2, p. 7].

In general, when describing enterprise architecture, the following principles are frequently used:

— The level of architecture refinement is being chosen in such a way that the amount of information about a certain component is minimized; anything irrelevant to the interconnection with other architecture components is omitted;

— Architecture definition mustn't contain descriptions of the components themselves.

Usually, four main layers [3, 4] of enterprise architecture are distinguished. They are presented in Table 1.

TABLE I.   MAIN LAYERS OF ENTERPRISE ARCHITECTURE (TOGAF)

№	Layer's name	Comments
1	Business architecture	Defines business processes and organizational structures of an enterprise
2	Information architecture	Describes logical and physical data structures, authorizing access to them
3	Application architecture	For applied systems, provides a plan and interfaces of their internal interaction, interaction with external systems, sources, or users of data, defines interrelation between applied systems and enterprise business processes supported by them
4	Technical architecture	Ensures performance of applied systems at a level described in operational requirements (reliability, scalability, capacity etc.), includes hardware, system software, networks, communication protocols etc.

Such enterprises as official statistics, i.e. those not working in productive industry, obviously need to have special architectural features. In particular, statistics deals with processing and searching not for physical resources but for information, searching being arguably its main activity.

The aim of this paper is to build enterprise architecture for enabling searching for statistical information, including searching for patterns in concealed data distribution features.

The rest of the paper is organized as follows. Section II describes three possible levels of information systems in official statistics. Section III reviews past researches that have been done in the area of Architecture Description Languages for enterprise architecture. Section IV details on the proposed architecture and its considerations for, particularly, searching for patterns of statistical information. Section V concludes the work.

II. THREE LEVELS OF INFORMATION SYSTEMS IN STATISTICS

From the developing a classification of applications for processing data in statistical information systems point of view, three principally different types (levels) may be distinguished.

1) Online Transaction Processing (OLTP) systems [5] which ensure basic functionality such as entering the data and results of appropriate statistical observations and surveys, their structured (usually by means of DBMS) storage and accounting, primary and aggregated data control, dissemination of results using various predefined output tables. Examples of such systems are the systems developed and implemented under the direction of the author for processing 2001 All-Ukrainian population census data and 2004 Moldova population census data. OLTP systems allow computing calculation indicators (like living area per one household member), creating predefined output tables and reports, and processing not predefined queries. In other words, representative capacities of such systems are rather limited. Thus, if creating a required output table was not specified at system design phase, obtaining appropriate data becomes practically infeasible, because to get a value of each cell or row of a new table one needs to make separate queries.

2) Online Analytical Processing (OLAP) systems [6] which allow creating not predefined tables and carrying out other analytical research of statistical data, including searching for their distribution patterns. Example of such systems is the system developed and implemented under the direction of the author for multidimensional analysis of 2001 All-Ukrainian population census data.

With this OLAP system, the user gained a natural and comprehensive data model arranged in three multidimensional cubes, namely, for the respondents, for the households and for the administrative and territorial units of Ukraine. Cube's dimensions were data features whose intersections enabled to obtain, filter, group, and represent information. For instance, the simplest cube for the administrative and territorial units had such dimensions as area, population size, urban or rural type, center of population type, predominate nationality. The cube for the respondents had 60 dimensions, and the cube for the households had more than 20 dimensions.

A *measure* defines what information is provided by the cube. For instance, a number of respondents can be a measure, and "native language," "marital status," "center of population" etc. can be dimensions. Each cube cell contains a number of respondents with particular features. The user analyzing information from such a cube can "slice" it across different dimensions, obtain aggregated or, on the contrary, detailed findings etc.

3) Data Mining systems which perform the most cumbersome and routine analytical operation of searching for concealed patterns that might exist in the analyzed data.

Thus, in contrast to architectural decisions of existent demographic data processing systems (in the US [7], France [8], Russia etc.) based on utilizing OLTP and OLAP levels only, we propose to add a new level of Data Mining.

Discussed levels differ sufficiently; their comparative analysis is presented in Table 2.

TABLE II. COMPARATIVE FEATURES OF DIFFERENT LEVELS OF STATISTICAL INFORMATION PROCESSING SYSTEMS

Feature	OLTP	OLAP	Data Mining
Data level	Mainly primary	Mainly consolidated	Primary and consolidated
Data variability	High (with each transaction)	Low	Low
Typical operation	Data alteration	Data analysis	Searching for patterns
Reports	Predefined	Not predefined, but with a certain list of attributes	Not predefined, list of attributes can be dynamically modified
Processed data	Only current ones; historical ones are usually stored in archive	Historical and current ones	Historical and current ones
Basic structure	Table / primary key	Cube / dimension	Cluster, class, association rule etc.
Priority	Efficiency	Flexibility	Intelligence

III. DISCUSSION ARCHITECTURE DESCRIPTION LANGUAGES

According to the monograph [9], languages used to describe enterprise architecture can be divided into two large groups:

— Universal languages such as language family IDEF [10, 11], BPMN (Business Process Modeling Notation) [12], ARIS (Architecture of Integrated Information Systems) [13], UML (Unified Modeling Language) [14] and others;

— Architecture description languages (ADL).

These two groups are complementary rather than interchangeable. Being inferior to universal modeling languages with respect to detailed description of processes, ADL languages have a natural advantage in describing architectural object features modeled [15].

ISO/IEC/IEEE Standard 42010:2011 [16] provides a rather general definition of an architecture description language as "any form of expression for use in architecture descriptions." It also contains relatively mild criteria of labeling certain modeling languages as architecture description ones. So, nowadays, one term, Architecture Description Language, is being used for modeling languages of different classes:

1) (Software engineering) software architecture description languages [17], such as ACME [18] and Wright [19] developed at Carnegie Mellon University, Darwin [20], AADL [21] etc.;

2) (Enterprise activity modeling) enterprise architecture description languages, the most known among them being ArchiMate [9], DEMO [22], ABACUS [23].

When used for describing enterprise architectures, virtually all the modeling languages from the first class have the following disadvantages [9, p. 38]:

Interconnections between different levels (representations) of a model are ill-defined; models created with different representations cannot be easily integrated in future;

— Language semantics is not "transparent;"

— There are restrictions for describing architecture of either business or technological (infrastructure) level of a model.

Thus, to describe enterprise architecture model that enables to search for statistical information and patterns in its distribution, we will use the specialized language ArchiMate which was standardized in 2008 by the Open Group consortium.

IV. PROPOSED ARCHITECTURE

In [24], the concept of on-line analytical mining (OLAM) systems was proposed. Its main idea lies in creating specialized OLAP systems for enabling of optimal searching for certain predefined patterns [25].

Such systems became widely used in criminology (to find out relations between crimes and known delinquents who could potentially commit them [26]), in medicine (to search for correlations between groups of people with a certain missing part of the Y chromosome and clinical presentation of infertility [27]) and so on. However, in a general case discussed in this paper, when the structure and elements of searched-for patterns are not known beforehand, creating OLAM systems is an unacceptable approach.

Therefore, we will build enterprise architecture model assuming that we have all three separate but complementary levels, namely, OLTP, OLAP, and Data Mining.

The model described in ArchiMate consists of three interrelated levels of view, namely, business level, application level, and technological (infrastructure) level.

To build required model at its business level, we will formally describe a process of searching for information. At other two levels, we will represent main software systems and servers used for implementing the described process, respectively (Fig. 1). To simplify notation, we don't show ancillary services like user authentication, backup and restore, contextual access to referenced data, local network support etc. To fit the whole architecture model to one page, detailed description of business functions is presented separately in Fig. 2–4.

In Fig. 1, bold vertical dotted lines additionally distinguish levels of statistical data processing system architecture singled out in Section II. The built model illustrates complementary capacities of OLTP, OLAP, and Data Mining systems when searching for statistical information.

Enterprise architecture model in ArchiMate presented in Fig. 1–4 was built using Archi editor, version 2.2 [28].

V. CONCLUSION

In the paper, enterprise architecture model that enables to search for statistical information is built. In contrast to existent models, the proposed one embraces all three possible architectural levels of corresponding information systems, namely, OLTP, OLAP, and Data Mining. The first level allows searching for required information in output tables, or by means of not predefined queries. The second level allows searching using ad hoc queries. The third level implies preparing microfiles and searching for patterns in concealed data distribution features.

The built model will be used for developing information system for processing data of All-Ukrainian population census to be held in this year.

REFERENCES

[1] E. N. Fong and A. H. Goldfine, "Information Management Directions: the Integration Challenge," National Institute of Standards and Technology (NIST), Special Publication 500-167, September 1989.

[2] K. D. Willett, Information Assurance Architecture. Boca Raton, FL: CRC Press, 2008.

[3] A. Josey, R. Harrison, P. Homan et al., Togaf version 9.1 Enterprise Edition: A Pocket Guide. Wilco, TX : Van Haren Publishing, 2011.

[4] FEA Consolidated Reference Model Document version 2.3. Retrieved from http://www.whitehouse.gov/omb/assets/fea_docs/FEA_CRM_v23_Final_Oct_2007_Revised.pdf

[5] G. Weikum and G. Vossen, Transactional Information Systems: Theory, Algorithms, and the Practice of Concurrency Control and Recovery. San Diego, CA: Morgan Kaufmann, 2001.

[6] E. Thomsen, OLAP Solutions: Building Multidimensional Information Systems, 2nd ed. N.Y.: John Wiley & Sons, 2002.

[7] 2010 Decennial response integration system (DRIS) contract. U.S. Census Bureau. Retrieved from

[8] http://www.census.gov/procur/www/2010dris/index.html

[9] G. Desplanques and G. Rogers, "Strengths and Uncertainties of the French Annual Census Surveys," Population, vol. 63, № 3, pp. 415-439, 2008.

[10] Marc Lankhorst, Enterprise Architecture at Work: Modelling, Communication and Analysis, 3rd ed. Berlin, Heidelberg: Springer-Verlag, 2013.

[11] Information Technology — Modeling Languages — Part 1: Syntax and Semantics for IDEF0, IEEE/ISO/IEC 31320-1-2012. Geneva, New York: IEEE : ISO : IEC, 2012.

[12] R. J. Mayer, M. K. Painter, and P. S. deWitte, IDEF Family of Methods for Concurrent Engineering and Business Re-engineering Applications. Columbus, OH: Knowledge-Based Systems, 1992.

[13] T. Allweyer, BPMN 2.0 — Introduction to the Standard for Business Process Modeling. Norderstedt: BoD, 2010.

[14] A.-W. Scheer, ARIS: Business Process Modeling, 3rd ed. Heidelberg, Berlin: Springer-Verlag, 2000.

[15] G. Booch, J. Rumbaugh, and I. Jacobson, The Unified Modeling Language: User Guide, 2nd ed. Boston: Pearson Education, 2005.

[16] M. M. Lankhorst and G. I. H. M. Bayens, "A Service-Oriented Reference Architecture for E-government," in Advances in Government Enterprise Architecture, P. Saha ed. Hershey: IGI Global, 2008, pp. 30-55.

[17] Systems and Software Engineering — Architecture Description, ISO/IEC/IEEE 42010:2011. Geneva, New York: ISO : IEC : Institute of Electrical and Electronics Engineers, 2011.

[18] Richard N. Taylor, Nenad Medvidovic, Eric Dashofy, Software Architecture: Foundations, Theory, and Practice. Trenton, NJ: Wiley, 2009.

[19] D. Garlan, R. T. Monroe, and D. Wile, "Acme: Architectural Description of Component-Based Systems," in Foundations of Component-Based Systems, G. T. Leavens and M. Sitaraman, Eds. Cambridge: Cambridge University Press, 2000, pp. 47-68.

[20] R. J. Allen, A Formal Approach to Software Architecture, Ph.D. Thesis. Pittsburgh: Carnegie Mellon University, School of Computer Science, 1997.

[21] J. Magee, N. Dulay and J. Kramer, "Regis: a Constructive Development Environment for Distributed Programs," IEE/IOP/BCS Distributed Systems Engineering, vol. 1, № 5, pp. 304-312, 1994.

[22] Peter H. Feiler, David P. Gluch, Model-based Engineering with AADL: an Introduction to the SAE Architecture Analysis & Design Language. — Boston, MA: Addison-Wesley Professional, 2013.

[23] J. Barjis, "Automatic Business Process Analysis and Simulation Based on DEMO," Enterprise IS, vol. 1, № 4, pp. 365-381, 2007.

[24] K. Dunsire, T. O'Neill, M. Denford, and J. Leaney, "The ABACUS Architectural Approach to Computer-Based System and Enterprise Evolution," 12th IEEE International Conference and Workshops on the Engineering of Computer-Based Systems, ECBS '05, April 4-7, 2005. N.Y.: IEEE Computer Society Press, 2005, pp. 62-69.

[25] J. Han, "Towards On-line Analytical Mining in Large Databases," ACM SIGMOD Record, vol. 27, № 1, pp. 97-107, 1998.

[26] M. Usman and S. Asghar, "An Architecture for Integrated Online Analytical Mining," Journal of Emerging Technologies in Web Intelligence, vol. 3, № 2, pp. 74-99, 2011.

[27] S. Lin and D. E. Brown, "Outlier-Based Data Association: Combining OLAP and Data Mining," Technical report SIE 020011. Charlottesville: University of Virginia, 2002.

[28] S. Dzeroski, D. Hristovski and B. Peterlin, "Using Data Mining and OLAP to Discover Patterns in a Database of Patients with Y Chromosome Deletions," AMIA 2000 Annual Symposium, November 4-8, 2000. Los Angeles: AMIA, 2000, pp. 215-219.

[29] Archi — ArchiMate modelling. Institute of Educational Cybernetics. — Retrieved from http://archi.cetis.ac.uk/
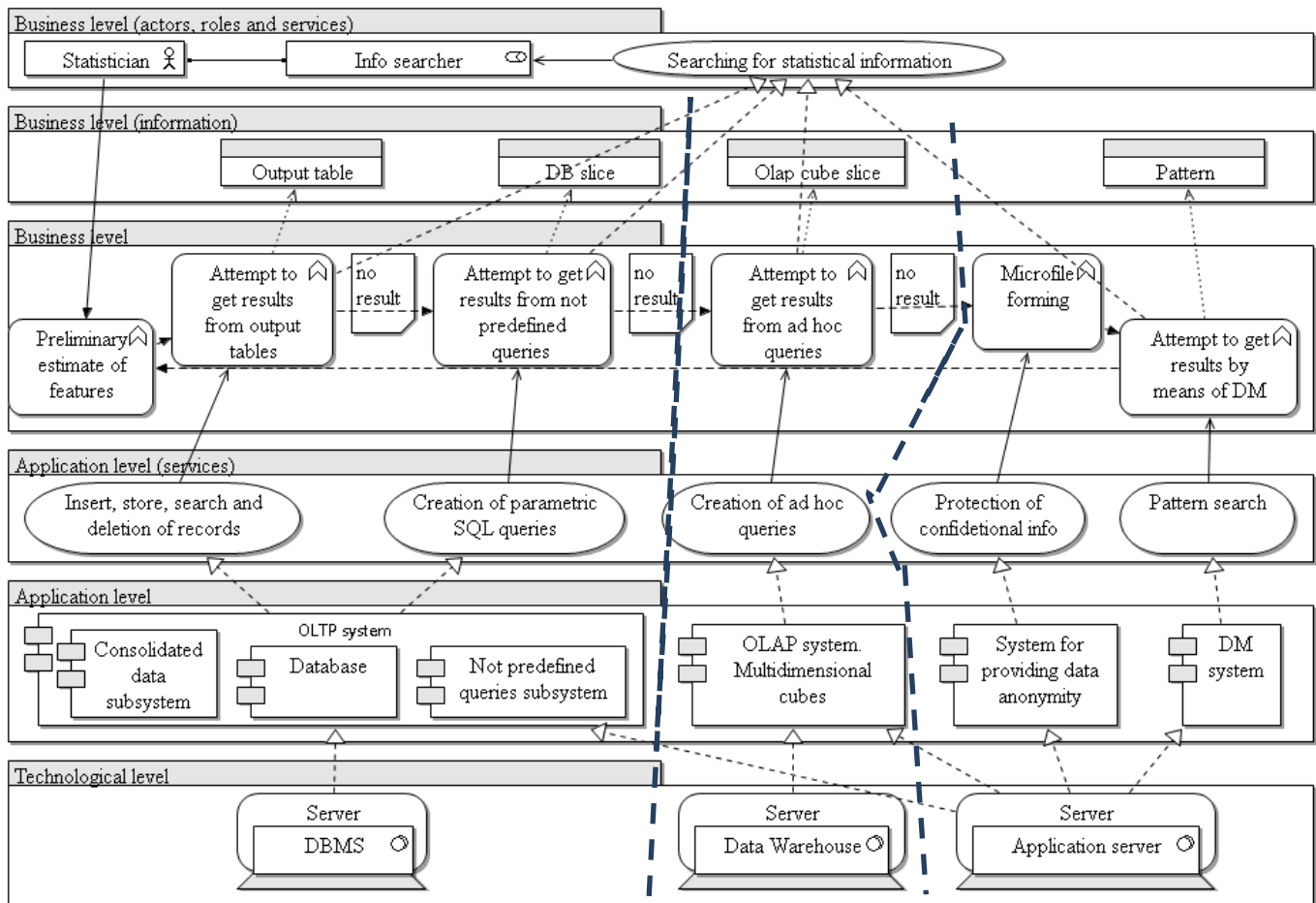
Fig. 1.   General architectural model that enables to search for patterns of statistical information.
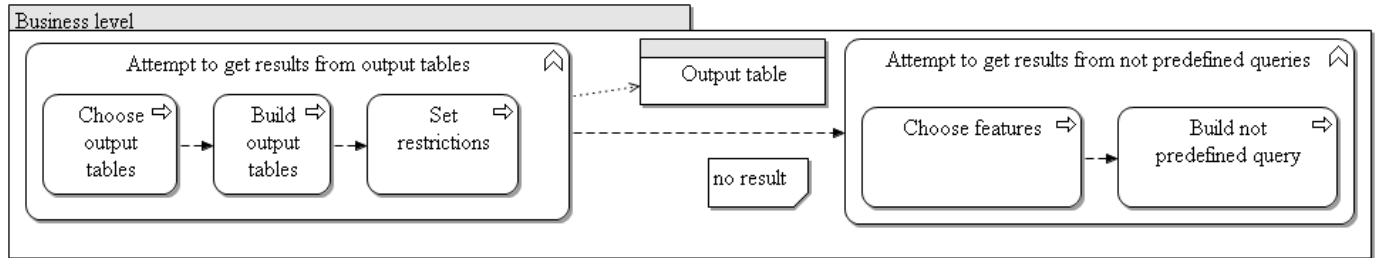
Fig. 2.   OLTP business functions for pattern search of statistical information.
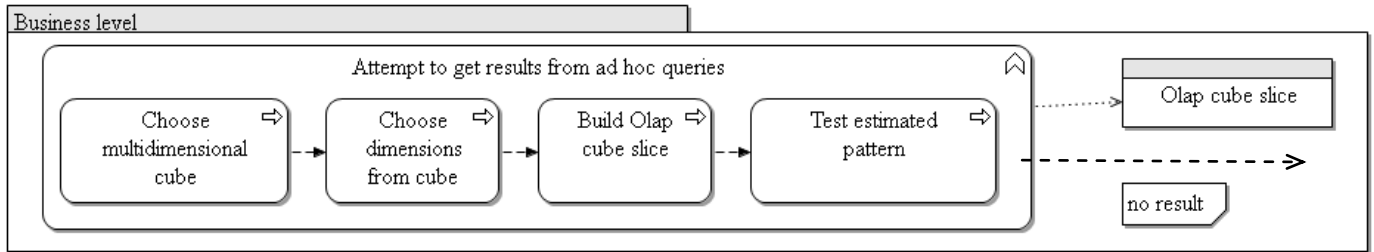

Fig. 3.   OLAP business function for pattern search of statistical information.
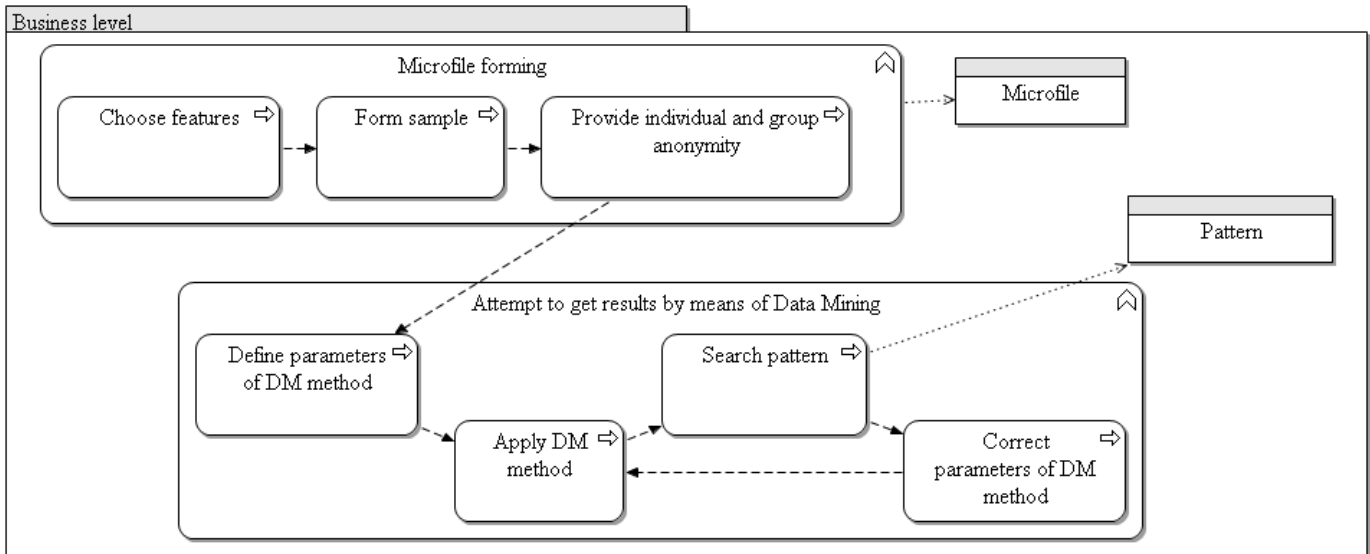

Fig. 4.   Data Mining business function for pattern search of statistical information.