

Application of complex information objects in industry management systems

Yuri Deniskin¹, Antonina Deniskina¹, Irina Pocebneva^{2*}, and Svetlana Revunova³

¹Moscow Aviation Institute, Volokolamskoe shosse, 4, Moscow, 125993, Russia

²Voronezh State Technical University, Moscow avenue, 14, Voronezh, 394026, Russia

³Moscow State University of Civil Engineering, 26, Yaroslavskoye shosse, Moscow, 129337, Russia

Abstract. The effectiveness of methods for forming a comprehensive information facility to make targeted decisions in high-tech industries is determined by the systematic organization of processes aimed at measuring costs and results in automated systems. The formation of such information facilities in terms of ensuring unity and comparability of information is considered. The main indicators are the description and language implementation of models of automated quality assurance systems, ways to ensure the unity of information and consistency of requirements, metrics of integrity and comparability of data in systems.

1 Introduction

Modern automated enterprise-level systems, especially when merged into integrated automated control systems (IASU), contain a large number of sources: databases, knowledge databases, electronic libraries, systems processing, receiving and transmitting information. In such circumstances, making qualitative (i.e. in each situation - correct, adequate, correct, timely, etc.) decisions inevitably leads to the need to share more than one of the available sources. Typically, this sharing requires not just simultaneous extraction of information, but its merging into a single information object, allowing decisions to be made at the right level. [1]. At the same time, the receipt of a comprehensive information facility must be carried out within a certain time, otherwise the information may lose its value or even lead to the adoption of a dangerous decision for the enterprise as a whole [2].

An important feature of modern systems, even if information is used only from one source of an integrated automated control system, is the transfer to critical functions that directly determine achievement end result and its characteristics. Consequently, the very fact of the system functioning becomes a necessary element of the health and safety of the enterprise. If the system does not function properly or malfunctions, it becomes impossible to achieve the results of the enterprise [3]. This necessitates the inclusion of functions that allow it to monitor the quality of its own functioning. This inclusion can take the form of a separate part of the system or additional automated systems [4,5].

The properties of automated systems impose serious restrictions on the final (based on internal processing) information. The scientific novelty of the proposed solution is the

* Corresponding author: ipocebneva@vgasu.vrn.ru

development of a comprehensive methodology for obtaining quality (compliant) data [6,7]. The methodology is based on two main approaches:

1. Extracting data of varying quality (including contradictory and poorly agreed) and data from multiple sources to an automated system, followed by the use of means to resolve, reconcile and extract reliable information.

This approach generally requires a complex system of harmonization of the information received and can be applied in cases where a priori input is "low-quality" (incomplete, untimely or erroneous). Often these are systems with a very large number of mediocre identified interconnected parameters, the process of collecting which does not allow you to get them with the specified accuracy. To make decisions in such systems, specialized algorithms are used to work in conditions of high uncertainty, which include methods based on machine learning (including neural networks), genetic algorithms, algorithms based on fuzzy logic, various heuristics, etc.

2. The use of various methods in automated systems to maintain the mutual integrity and consistency of the data at all stages of its production, transmission and processing.

This approach involves monitoring the input, storage and processing of information in an automated system in such a way that the entire processing process ensures that the data is sourced from a reliable source and were not deliberately or accidentally misrepresented.

A combined approach is also possible, using both data control and specialized methods for harmonizing it in later stages of processing.

All actions that ensure unity and comparability of information will be combined into the following four groups:

1. Maintaining information stored in various information sources in an up-to-date and coherent manner [8,9].

2. Audit the collected, stored and processed information for consistency and relevance.

3. Using algorithms that increase the unity and comparability of data.

The inclusion of tools to assess the quality of an integrated automated control system.

To ensure the unity and coherence of information, we will introduce the term "quality information." High-quality information is information that meets a certain set of user requirements. By considering different approaches to quality assurance, a set of tools and methods can be identified or based on them to ensure that information is consistent and consistent[10,11].

It should be noted that quality characteristics for real automated systems should be viewed not as separate characteristics, but as metacharacteristics, which are intersecting sets, formed from smaller ones Properties. By highlighting the set of "unity" and "coherence" properties that define each case, an approach can be formed to achieve them. It can also be assumed that, due to the intersection of sets, many of the methods used and developed will lead to an improvement in several properties of information at once.

In general, an attempt to determine the quality of information leads to a rather complex multidimensional model containing both subjective and objective metrics. In one automated system, the list of characteristics can vary depending on the task and vary over time.

In this regard, it is appropriate to consider only the most common of them, for which there are metrics and tried-and-tested methods of implementation. These characteristics include: relevance, correctness, completeness, integrity, compliance.

2 Experimental

To understand how to ensure unity and comparability of data, we will define the overall processing process within which it must be guaranteed. Due to the orientation of this study on data processing processes in quality management in high-tech industries, it is advisable to make the most of the existing proven processes from areas close to close requirements

for the complexity of the processes performed, their repetitiveness and security requirements. For example, a civil aviation area where a fairly general process of processing information is used in automated systems[12,13].It is expected to consist of the following stages:

1. Check the accuracy and reliability of the original data.

2. Save data on an electronic medium that can be used throughout the process.

Data/information control. It is made to verify that the data/information after electronic transmission was obtained by the following recipient without any distortion or change, regardless of the source of such distortions (intentional or accidental). Ways to control - calculating control amounts, incorporating electronic signatures, etc.)

4. Encryption (if necessary) of electronically transmitted data to ensure that its integrity is further protected.

5. Transfer data to the next recipient.

6. Check the recipient of the completeness and integrity of the data.

Using or publishing the data received by the recipient.

During the data/information control phase, the Bayes method can also be used, which uses an appraisal base - two blocks of information, one made up of distorted information and the other from a regular one. Each block calculates the frequency of each distortion, after which a weight score (0 to 1) is calculated, which characterizes the conditional probability that the information with this distortion is unreliable. Weight values close to the $\frac{1}{2}$ scales are not taken into account in the integrated calculation, so distortions with such weights are ignored and removed from information dictionaries.

According to the method proposed by Paul Graham, if the information contains n distortions with weight estimates $W_1 \dots W_n$, the estimate of the conditional probability that the information will be distorted, based on data from the evaluation blocks, is calculated according to the formula:

$$S = Pw_i / (Pw_i + P(1-w_i)) \quad (1)$$

This formula is based on the following consideration. It is assumed that S is the fact that the information is distorted, A is the event that the information contains distortion t . Then, according to Bayes' formula, it is fair:

$$P(S|A) = \frac{P(A|S)P(S)}{P(A|S)P(S) + P(A|\bar{S})P(\bar{S})} \quad (2)$$

If it is not initially known whether the information is distorted or not, based on the experience assumed that, it follows: $P(\bar{S}) = \delta P(S)$

$$P(S|A) = \frac{P(A|S)}{P(A|S) + \delta P(A|\bar{S})} \quad (3)$$

The formula is then summarized as follows. It is $A_1 A_2 t_1 t_2$ assumed that these are events that contain distortions and. It is assumed that these events are independent. The conditional probability that information containing both distortions is inaccurate is equal to:

$$\begin{aligned} P(S|A_1 \& A_2) &= \frac{P(A_1|S)P(A_2|S)}{P(A_1|S)P(A_2|S) + \delta P(A_1|\bar{S})P(A_2|\bar{S})} = \\ &= \frac{p(t_1)p(t_2)}{p(t_1)p(t_2) + \delta(1-p(t_1))(1-p(t_2))} \end{aligned} \quad (4)$$

The generalization of the formula in case of an arbitrary number of distortions is the formula of P. Graham. It should be noted that the widely used one finds exactly the value $\delta = 1$. Although this simplifies the calculations a little, it seriously distorts reality and reduces quality. In practice, based on information dictionaries that are constantly modified, the value of S is calculated for each block of information.

The concept of "unity of information" in an automated system will be defined through a certain set of properties: general (in general), interaction of elements, descriptions, used tools of design and development and certain methods of their implementation, through the reverse (duality / multiple information).

The most common approach is that an automated system is designed to eliminate or minimize the possibility of dual information occurring at the system and architecture design level. Input and output use duality tools and methods for input and output. The main methods of ensuring the unity of information are:

The method of traceability (most commonly used, especially at the requirements level).

3 Evaluation

Analytical methods that are divided into heuristic methods and formal analysis.

The basic principles that should be based on the implementation of the automated system from the standpoint of ensuring unity and comparability of information in it are:

1. All processes used must be defined, documented and understood by the individuals, units and organizations that perform them.

2. Documentation is done using one or more models that describe both the aerospace-specific quality assurance processes and the processes common to quality management and support systems lean production. Particular attention is paid to the interaction and integration of these submodels within a single meta-model of quality management in the enterprise.

3. The connections between the key steps of the processes must be ensured and supported by a quality of service agreement.

4. The automated system is necessarily regarded as an integral part of the integrated automated enterprise management system.

The proposed approach, which ensures unity and comparability of information, is based on a model developed, which consists of a combination of several interconnected submodels, representing the description of individual processes, requirements, usage scenarios, action charts, interaction patterns, etc.

Let's highlight several models that are needed and used to implement an automated system:

1. The data processing model is a formalized description of the procedures for computational process organization and data conversion.

2. The data-sharing model is a formal description of the procedures performed on the network: transmission (including coding), switching and routing messages.

3. The data accumulation model describes the database management system and the structure of the information base. The models of the database and the knowledge base are separated. The process of transition from semantic (information) representation to physical is carried out by a three-tier system of models of the information base: conceptual (what and to what extent information should be accumulated during the implementation of information logical (structure and relationship of information elements) and physical (methods of placing data and accessing it on machine media).

4. A model of knowledge presentation. Selected according to the subject area and the type of tasks to be solved. Knowledge presentation models such as logical, semantic, functional and integral models, combining several models at once, are widely used.

Simulation in an automated system is done using the high-level sysML graphic language, which has the following properties:

Collecting, analyzing and tracking requirements

A qualitative graphic representation of the main processes performed by the automated system.

Moving from graphic description to creating their descriptions, including drawing up formal process specifications that can be used to perform different validation, audit and harmonization algorithms;

Providing, on the one hand, a description of quality management processes, and on the other - descriptions of processes;

Application in high-tech industries as part of the operation of automated systems and process modeling;

There is a fairly broad support in the means of developing automated systems.

The language used is also close to the most common language of simulation systems of any UML focus.

Here's an example of the implementation of the WSFL language(The Web Services FlowLanguage)developedby IBM.

Each thread of work that combines the use of a set of some Web services includes:

A business process that defines the set of actions that need to be done during the process,

The rules that determine the sequence of business steps

The flow of information between the various stages of the business process.

The formal WSFL model, reflecting the interaction between the work flow and external service providers, is shown in the picture (Fig. 1).

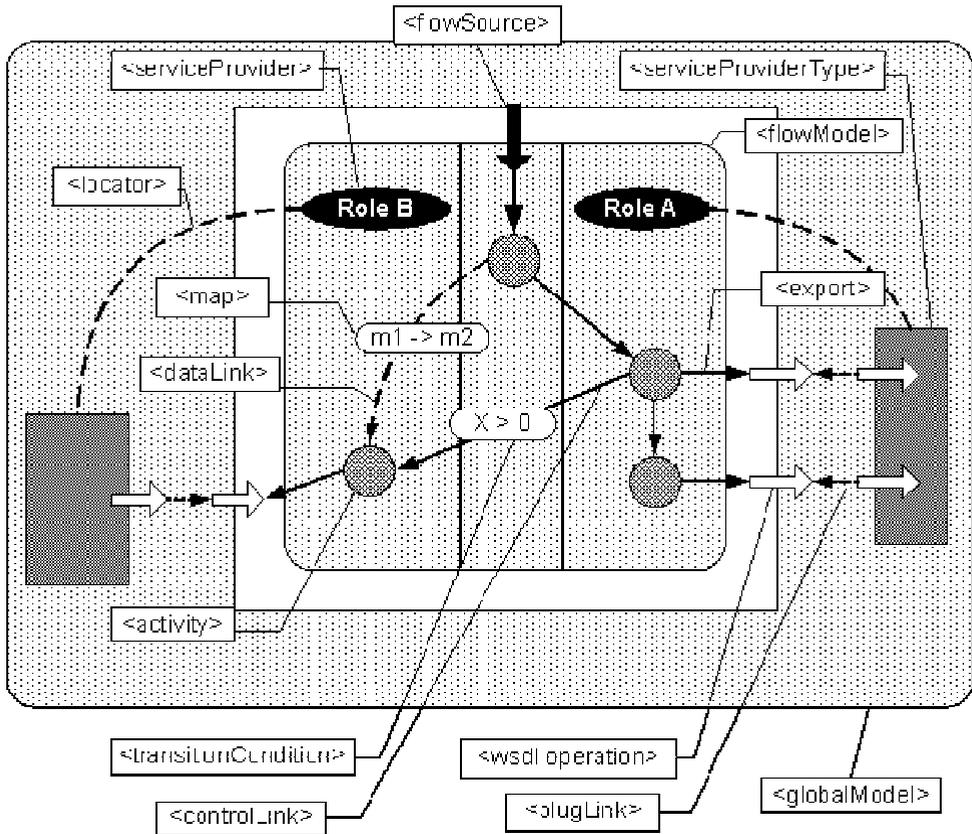


Fig. 1. Formal WSFL model.

WSFL Flow Model defines the structure of the business process: "actions" (marked in circles on the chart) describe the steps of the process, and "control links" (the element of the "controlLink") and "data flows" (the element of the "dataLink") define the sequence transitions and data flows between "actions" respectively. Each "action" is assigned to a "service provider" (the element of the "serviceProvider" that is responsible for this phase of the business process, and the links between "actions" in the thread model and "operations" are identified (element of the "wsdl:operation") service providers (elements of the "export" and "plugLink" (elements). The resulting work flow model (the element of the work flow) is shown in the center of the chart. It is broken down in vertical lines, thus reflecting the relationship of the process's "actions" with specific service providers.

With a note of the specific features of the WSFL specification:

- 1) The stages of the business process are represented by "actions" that bind to the directional graph with conditional or unconditional transitions between its tops. Each "action" is associated with some operation of the service provider responsible for its implementation.
- 2) The "action" (ControlLink) links are directed arcs of the thread model route graph. Transitions between "actions" can be conditional.
- 3) The WSFL model allows for several "actions" to be performed side-by-side. To do this, a "special type action" called forkactivity is used, from which it can be used to take several other actions at the same time. At the same time, WSFL supports a set of tools to synchronize the simultaneous stages of the process.

4) There are two dedicated types of "actions": initial (to which there are no transitions through ControlLink) and end ones (which are not related by "outgoing" transitions to other "actions" of the process).

5) A stream of data can be organized between successive "actions" to share management and other information.

WSFL and XPD languages belong to the same family of graph XML languages describing work flows, as opposed to the BPML block structure, which defines their similar functionality, although XPD gives a much greater level of abstraction (perhaps the use of human resources and the types and interfaces of automated ones are not specified).

4 Conclusions

As a result of the study, it is shown that the basic principle that should be the basis for the creation of the system from the point of view of ensuring unity and comparability of information, is its implementation as an integral part of the integrated automated enterprise management system. This will provide a synergistic effect with other components of the system. The approach that ensures unity and comparability of information is based on a model that allows to implement the relationship between fairly heterogeneous entities and to combine them in a single system of quality management. The novelty of this study is the development of a number of models related to the processing, exchange and accumulation of data, as well as the presentation of knowledge. It is obvious that the developed approach to the formation of a comprehensive information facility in the management system for making targeted decisions largely determines the systemic quality of the entire industry.

References

1. L.L. Pipino, Y.W. Lee, R.Y. Wang, *Communications of the ACM* **4**, 211-218 (2002)
2. *Manual on the Quality Management System for Aeronautical Information Services* (ICAO, 2010)
3. M. Kamalrudin, S. Sidek, *International Journal of Software Engineering and Its Applications* **9(10)**, 39-58 (2015)
4. K. Gumba, S. Svetlana, S. Belyaeva, S. Revunova, *MATEC Web of Conferences* **106**, 08023 (2017) DOI: 10.1051/matecconf/201710608023
5. E.N. Desyatirikova, O.V. Kuripta, Y.S. Stroganova, I.P. Abrosimov, *Proceedings of the 2017 International Conference "Quality Management, Transport and Information Security, Information Technologies", IT and QM and IS* **8085871**, 500-505 (2017) DOI: 10.1109/ITMQIS.2017.8085871
6. I. Pochbneva, Y. Deniskin, A. Yerokhin, V. Artiukh, V. Vershinin, *E3S Web of Conferences* **110**, 01074 (2019) DOI: 10.1051/e3sconf/201911001074
7. A. Korchagin, A. Deniskina, I. Fateeva, *E3S Web of Conferences* **110**, 02124 (2019) DOI: 10.1051/e3sconf/201911002124
8. E. Makovetskaya, A. Deniskina, E. Krylov, F. Urumova, *E3S Web of Conferences* **91**, 02036 (2019) DOI: 10.1051/e3sconf/20199102036
9. A.P. Fadillah, D. Fitriana, *IOP Conference Series: Materials Science and Engineering* **662(2)**, 022014 (2014) DOI: 10.1088/1757-899X/662/2/022014
10. S.A.E. Hayat, M. Bahaj, *3rd International Conference on Intelligent Computing in Data Sciences* **8942298** (2019) DOI:10.1109/ICDS47004.2019.8942298

11. V. Yasin, O.S. Sitompul, M. Zarlis, P. Sihombing, *3rd International Conference on Electrical, Telecommunication and Computer Engineering* **8943831**, 107-110 (2019)
DOI: 10.1109/ELTICOM47379.2019.8943831
12. Y.A. Gatchin, V.V. Sukhostat, *Wave Electronics and its Application in Information and Telecommunication Systems* **8840618** (2019) DOI:
10.1109/WECONF.2019.8840618
13. P. Slegr, P. Pansky, M. Drabek, *Smart Cities Symposium Prague* **8805672** (2019)
DOI: 10.1109/SCSP.2019.8805672