The concept of knowledge bases supporting the IT Systems Integration Model

Cezary Orłowski\textsuperscript{1}, Tomasz Sitek\textsuperscript{1}, Krzysztof Bartosiewicz\textsuperscript{2}, Łukasz Szczygielski\textsuperscript{1}

\textsuperscript{1}Gdańsk University of Technology, Department of Management and Economics, ul. Narutowicza 11/12, Gdańsk, Poland
\{cezary.orlowski, tomasz.sitek, lukasz.szczygielski\}@zie.pg.gda.pl

\textsuperscript{2}krzysztof.bartosiewicz@gmail.com

Abstract. This paper presents main concepts regarding the design of a knowledge base in order to support IT systems integration. The main goal of the research is to develop a universal method of integration of various systems concerning any possible area of interest. The concept is based on creating a generic mechanism which would support the decision regarding the choice of the integration model. The most important tasks include: key factors analysis, building the integration process knowledge base and designing artificial intelligence-based decision support mechanism. The proposed solution is in an early stage of development therefore this paper is just the beginning of a series of articles. After implementation it is planned to verify the solution during the project which will be responsible for integration of various environment monitoring systems.

Keywords: IT system integration, knowledge base, knowledge management, artificial intelligence, fuzzy logic, process modeling, decision support system

1 Introduction

Currently it is very hard to find a company or organization where IT systems are not crucial in delivering efficiency needed for everyday operations. The process of worldwide computerization causes that IT systems become more and more important. Moreover since the world is recently struggling with financial crisis it causes the need for cost-optimization in all areas and IT systems are a great place to look for improvements since when properly used they can offer a substantial increase of cost-efficiency in any organization. The crisis has one more interesting consequence – it causes incredible amount of mergers all over the world since companies try to lower the cost base by using the synergy effect.

Since usually merged companies have their own IT systems the process of IT systems integration became incredible important during last few years. Unfortunately the process of integration is a very complicated and risky undertaking – usually it ends
with significant delays and often cannot deliver originally planned benefits ([1]). That is why it would be extremely useful if a tool existed which could simplify the process, especially the decisions which must be taken. This paper describes the idea of such a tool and plans regarding the development of a universal model of integration of various systems concerning any possible area of interest. It is planned to create a generic mechanism which would support the decision process regarding the choice of the integration model.

1.1 Integration process – general overview

There are two typical situations where such a tool could be used. The most typical is a merger of two or more organizations which results in the need of their IT systems integration. However it can also be used when organizational changes are not present – to prepare an application basing on a few available software components in order to prepare a coherent and integrated solution delivering all the functionalities of the subsystems in one powerful tool. Such a situation can occur either when single organization decides to integrate their own systems or even when a few organizations cooperate on a single undertaking and it is necessary to deliver one system which would be able to gather data and communicate with the software already present on the contributors sides. The latter case will be used to check if the model is working properly since there are plans to verify the solution during the project which will be responsible for integration of various environment monitoring systems (air quality, noise measurement, etc.) into one environmental decision supporting tool used by the authorities in TriCity area.

The view of a typical flow of the integration process is shown at figure 1.

![Image](https://example.com/image1.png)

**Fig. 1.** Typical flow of the integration process

One aspect which should be especially emphasized about figure 1 is that a “method” word was used. It is very important since integration is never only on the applica-
tion level, sometimes it can apply only to the process level but typically it affects both simultaneously. That is why the most important factor in the integration process is deciding how the integration should proceed. There are many possible scenarios of integration (this topic is described in detail in [2]) but in general four typical groups can be distinguished:

- creation of a new complex system containing all functionalities of the old systems and migration to the new solution,
- modification of already present systems to allow them to communicate with one another,
- creation of new frontend applications which gather data from old backend systems with small changes in those backends to allow them to pass data to new frontends,
- designing new processes based on the old applications without implementing any changes to them.

As it was stated before, each of the scenarios requires reorganization of processes. Naturally various methods of IT systems integration require different level of reorganization. Usually creation of new systems requires changes in all processes while by allowing already present systems to communicate with one another it is usually required only to make small process changes. Naturally it is possible to perform integration only on a process level but it not very often since it usually does not allow to use all advantages from synergy effect and thus is less cost-efficient. However it should be mentioned that this method of integration is fastest and is sometimes used when time is essential – it can also be used as a temporary solution before the final approach is implemented.

1.2 Integration model

Each integration process starts with acquisition of the data regarding integrated elements (processes, applications, etc.). Then after gathering the data the analysis begins which is responsible for delivering some options regarding the integration model for the decision makers. After a decision is made it is finally possible to plan and execute implementation of the integration.

Figure 1 presents a simplified model of integration process.

![Fig. 2. Simplified model of integration process](image)

Naturally all of the elements are crucial for the success of the integration. However usually errors made in the early stage of the process create higher risks and are more costly to fix. Therefore it is very important that the decision must be made using properly gathered and analyzed data. That is where the proposed solution comes in since it facilitates the decision making process by organizing the data gathering and data analysis methodologies.
2 Generic mechanism supporting integration process

The main concept of proposed solution is to facilitate the decision-making process regarding the integration process. Unfortunately algorithmization of all decisions, especially those regarding process level integration, is hardly achievable. That is why the designed decision support model will only support process-concerning decisions partially concentrating on IT systems integration. The main aspect is to gather all data regarding existing processes and applications along with all substantial factors important from decision-making point of view and, using AI-based algorithms, present the decision supporting data. The result should contain a set of applicable solutions with one recommended option. What is important the application should present not only final results but also the list of all criteria and their values calculated during the process. The final result should show which applications should be integrated and how and suggest which areas are better to be integrated using process level integration (e.g. system integration is extremely expensive).

2.1 Model of the system

The general model of the system is presented in the figure 3.

As it can be seen we can distinguish two types of users:

- decision-maker – responsible for providing criteria of the analysis as well as for making the decision basing on the results of analysis,
- expert – responsible for providing the knowledge which is necessary to perform the analysis.

The model of the system contains four main modules shortly characterized below:

- Data Acquisition Module – allows the expert to provide input data needed for the analysis,
• Knowledge Base – module responsible for storing the gathered data regarding the integration process,
• Inference Module – module responsible for processing the information gathered from experts and stored in Knowledge Base,
• Presentation Module – allows the decision-maker to input criteria for the analysis and shows the end results.

The above-mentioned modules represent four key areas of the decision support system: data acquisition, knowledge management, data processing and data presentation. Each of those areas are described more thoroughly in the next sections of this paper.

2.2 Data acquisition

The most critical issue in the whole solution is knowledge management. Therefore it is essential that key factors which may affect the integration process must be defined first. Then dependencies between them must found. Such factors will be variables used when building formal structures of knowledge.

It will be necessary to perform a deep analysis of the integration process and obtain a detailed expert knowledge in order to be able to prepare a set variables and relationships which will represent the process in its full complexity. Finally the factors’ influence on the choice of the integration method must be estimated, defined and then all data has to be stored in a knowledge base.

Obviously before the data is stored it has to be gathered. The model will need the user to input the data regarding processes and applications which are objects of the integration process. Some of the areas of input data which will be gathered are listed below but precise criteria will be defined while the design of the system progresses:

• data regarding all applications,
  — users and roles in the system,
  — input and output data,
  — application interfaces,
• data regarding all processes,
  — roles definition,
  — input and output of the process,
• complexity and functional overlapping between applications,
• process-system relations,
• process-process relations,
• systems availability and accessibility,
• workload estimations for certain aspects of integration,
• systems maintenance and support terms (costs),
• new functionalities needed.

It is important to emphasize the fact that by requiring certain data the system will make the data gathering process significantly easier since it will no longer be necessary to analyze what data should be gathered and what data is important from the integration process point of view.
The data needed to make a decision will be very complex. Therefore a proper user-friendly interface has to be offered in the solution. The user will be able to define applications and processes and then enter their properties needed by the system to perform the required analysis. After process and application definition the user will be able to describe the data regarding various relations in a graphical form. Such a solution will simplify the process of entering and validating the data but it has one more advantage which could not be overestimated – once the data is entered the user will be able to visualize it and thus create a model of organization. The model created during data gathering has its own value but it will also prove very useful once the decision supporting analysis is complete since it will help the user to trace relations between objects which will be crucial for understanding the final results.

Obviously gathered data itself is not sufficient to prepare any automated analysis – it is important that a proper knowledge management mechanisms are present which allow the system to store the data and the rules of data processing in an entity called the knowledge base.

2.3 Knowledge management and information processing

The authors aim to deliver decision support mechanisms to project managers responsible for realization of projects. The decision can be proper only when it is a result of a thorough and extensive analysis of all project requirements and project environment. Unfortunately usually people responsible for the decision are working under high pressure of on-time delivery and thus they hardly have such comfort. That is why the following circumstances may become a significant factors [3]:

• decision-maker does not have enough experience regarding the project domain,
• decision-maker does not have any decision support,
• decision-maker has certain data but it is not sufficient to make the right decision.

In such conditions decisions are often made partly intuitive and as it was stated before it is connected with risk and therefore with the possibility of additional costs.

By analyzing the above-mentioned circumstances it can be seen that only the first one is fully subjective in nature – it is not possible to swiftly eliminate the problem of lack of experience. However it is possible to mitigate the risk by providing the decision support and proper management of available knowledge not only connected to the current project environment but previous experiences as well.

As the proposed solution will be knowledge-based it is necessary to prepare assumptions regarding preparation of model which uses knowledge resources. Systems based on knowledge are always built for specific problem area. Although it is possible to distinguish some common characteristics such as separation of inference module from knowledge bases, each project which requires knowledge management requires dedicated solutions adjusted to specific knowledge of given area.

That is why it is necessary to design the following processes related to the knowledge resources in the model [4], [5]:

• data acquiring,
• data storage (formalization),
• data processing (inference).

At the current – very early – state of building of the proposed solution for supporting the integration of IT systems it is crucial to focus on knowledge gathering and formalization.

To be able to gather the required knowledge it is necessary to make contact with area experts experienced in the field of systems integration. The authors plan to contact with professionals responsible for integration projects in the past. It is also planned to contact with members of Polish Information Processing Society (PIPS) [6].

PIPS is a society concentrating on the development of computer science and the information society. PIPS Members are mainly individuals who graduated from a computer science or related faculties. They are professionally active but some of them are current IT students as well so it is an organization with a very wide perspective. It should be mentioned that the authors were responsible for realization of projects which involved PIPS in the past. Basing on those experiences the authors assume that PIPS members will be interested in following project and willing to support it in the area of knowledge gathering.

At the moment it is too early to plan which approach will be proper for gathering the data (polls, electronic forms, interviews, etc.) – the method will be chosen later. The method of storing the knowledge will also be chosen at the later stage of the project. Although there are many possible solutions – classic rule approach, frame-based approach or semantic methods (networks) – the preliminary analysis of the domain suggests that the knowledge is passed in the form of IF-THEN instructions. Although it suggests that rules should be chosen as a method of knowledge formalization it is also possible that a fact-based approach can be used (O-A-V triplets: Object-Attribute-Value).

The rules are used in the inference process in order to receive the optimal scenario for the integration. However the integration result is strongly connected to the environment of the specific case which will be described in another set of rules. In general three types of rules can be distinguished:

• high-level rules – responsible for the „big picture”. This set of rules describes which scenario should be picked for the integration process basing on a very general statements. At this stage they can be presented in the simplified form using Prolog language notation. Some of examples of such rules are shown below

```
Scenario(creation_of_new_system):-
    Low_time_pressure(requirements),
    Low_risk(one_system),
    Low_cost(technical_changes).
```

```
Scenario(modification_of_the_system):-
    High_time_pressure(requirements),
    High_risk(one_system),
    High_cost(technical_changes).
```

```
Scenario(new_frontend_app):-
    Low_time_pressure(requirements),
```
High_risk(one_system),
High_cost(technical_changes).

Scenario(new_processes_only):-
   High_time_pressure(requirements),
   No_change_requirement(applications).

• mid-level rules – describe the project environment, they are responsible for providing the key factors’ values for high-level rules. Mid-level describe problems connected with costs, resources availability, client needs or formal aspects. The example of such a rule is shown below

   High_risk(one_system):-
      High_complexity(applications),
      High_complexity(processes),
      High_functional_overlapping(frontends),
      High_number(supported_processes),
      High_number(interfaces).

• low-level rules – these rules contain the elementary knowledge about the integrated systems / processes or other important from the client point of view factors

   High_complexity(applications):-
      Many_roles(application1),
      Many_roles(application2),
      Many_interfaces(application1),
      Many_interfaces(application2),
      Many_supported_processes(application1),
      Many_supported_processes(application2).

It should be mentioned that there are some problems which will have key impact on the inference process and they have to be identified as soon as possible. A large number of experts which is to be aimed at causes possible intensification of some negative effects [7]:

• knowledge uncertainty – expert is not sure if all formulated facts/rules are correct,
• knowledge incompleteness – expert has only partial knowledge about the domain,
• lack of precision – expert cannot describe knowledge unambiguously, by using sharp values instead of fuzzy ones.

The problems listed above must be recognized at the beginning of data gathering process since they determine the method used for storing the knowledge. In that case a mixed approach can be used by introducing the following elements to the classic fact-rule methodology [8]:

• fuzzy logic processing – using the rules based on fuzzy values and processing them using fuzzification/defuzzification processes; an example of rules based on fuzzy values is shown below:

   Big(cost_of_technical_changes).
Medium(cost_of_technical_changes).
Small(cost_of_technical_changes).

- certainty factors – the numerical measure of certainty, which allows to prioritize the rules during the inference process. It can be assumed that they will be represented by a number from the range <0;1> where 0 represents total uncertainty and 1 – total certainty. An example rule containing the certainty factor as an argument in Prolog notation is presented below

```prolog
Scenario(new_processes_only, 0.95):-
  High_time_pressure(requirements),
  No_change_requirement(applications).
```

It should be emphasized that data gathering is hardly a one-time activity. Usually it is an iterative process responsible for modifying the knowledge basing on new experiences. Similarly the expert team also should not be hermetic – it is important to adapt the team to the changing environment.

### 2.4 Presenting the results

Data acquisition and knowledge management described in the previous sections are fundamental for proper information processing. Once the data is gathered, all factors and relations defined and the analysis complete the results of analysis have to be shown to the user. The aim of the presentation is to support decision-making process so it is necessary that possible options with suggestion about the best possibility are presented. Since the bad decision before the integration could have dire consequences it is important that the process is as transparent as possible – all criteria and corresponding values calculated during analysis with additional comments have to be shown to allow the user to trace the results and understand the reasons for the proposed solution.

As stated before the final result contains the details of the proposed integration approach – which components should be integrated only on process level and which should be integrated on application level. Naturally in case of application level integration some additional information regarding the proposed method is to be delivered.

It should be emphasized that although the analysis can be prepared after entering all required data by experts the decision-maker has the possibility of influencing on the final result by changing the parameters of the algorithm. At the beginning default values are taken into consideration and the best theoretical solution is presented but since the project environment can differ from one another it is possible to change the weights of each criteria. For example if in some case costs are not important and time is crucial it is possible to reflect that by changing the weights of time and costs and observe how it influences the final result.

### 3 Next steps

The proposed solution is in an early stage of development. Therefore a series of articles are planned which will describe next stages of model and system development.
After implementation it is planned to verify the solution during the project which will be responsible for integration of various environment monitoring systems. Some of the systems covered by the project are focused on environmental data gathering and monitoring and present the data in a very simplified form. Others contain complex mechanisms for data presentation and data analysis, including short- and mid-term forecasting and decision support tools. Therefore since the systems have many unique and many overlapping functionalities the project responsible for their integration will constitute a great opportunity for planned solution verification.

It should also be mentioned that the project schedule is tight and the integration model in the project is to be prepared in the next few months. The worked out integration model with all data connected with the analysis made during the integration process will be available to the authors soon. That is why it will be possible to perform a post-factum analysis by comparing the results with results delivered by the decision support system described in this article and thus check the model correctness.

4 References