Data Integration: A Case Study of Kabul Polytechnic University Management Information Systems

Mohammad Zia Sana
Faculty of Computer Science
Kabul Polytechnic University, Kabul, Afghanistan

Abstract:
Nowadays, information systems play a significant role in management, learning, and facilitating processes in organizations. In many cases, these systems are not connected and operate separately. This independence of systems in operation usually causes huge duplicate data storage, redundant work, and extra costs for the firm. While having a set of systems which are integrated and operate in connection to one another to reduce duplicate data storage, redundant work and costs of an organization is desired. In a situation like this, (Kabul Polytechnic University) which currently utilize learning management system (Moodle) to advances its teaching processes, the feedback management system to facilitate evaluation and management procedures and also uses dormitory management system to manage students and dormitory affairs; that dis-connectivity is the huge issue among them. To solve issues, relate to information silos in this case, we can integrate the above systems to prevent duplicate data storage, redundancy processes, and extra costs of the organization. In primary steps, for many of the mentioned firm’s data integration based on the case requirements, using API and middleware architecture would be a better solution. Therefore, here we tried to study MISs at Kabul Polytechnic University to integrate them and solve integration problems in data level to have a set of integrated systems. We Used database restructure and RESTful API to integrate two one of current management information systems and also worked on finding a platform to integrate all current and future systems. Our solution for whole system integration is a middleware-based platform.

Keywords: System Integration, Data Integration, API, Middleware, Information System Integration, Kabul Polytechnic University, Management Information System.

I. INTRODUCTION

In recent decades especially in the last decade using information technology and information systems gained popularity and an important role in organizations and enterprises in the world. Although, in our country, Afghanistan using technology is not as popular as other societies, in recent years’ firms and organizations tried to implement information systems to modernize their business and structures and achieve their objectives. On the other hand, islands of information systems in Afghanistan is one of the challenges against using information systems and technology. Now and in past years, there were many MIS projects in different public and private organizations, which failed due to their incompatibility in connecting with other systems in that organization like HEMIS in the Ministry of Higher Education, which suffers from incompatibility to connect with older systems and muddles in that organization. One of the academic organizations in Kabul which uses information technology as a tool to meet its aims is Kabul Polytechnic University (KPU). This university currently uses three information systems to meet its objectives in learning and management areas which are Learning Management System(LMS), Feedback Management System(FMS), and Dormitory Management System (DMS), and each one for a specific purpose. The LMS is a learning management system (Moodle) which stores data of students, lecturers, courses, and material of courses. This system helps students in learning, accessing materials and exam results, and communicating with lecturers. Also, it assists lecturers to provide course materials, courses, and the result of tests for students. The main purpose of this system is to facilitate learning and teaching procedures for students and lecturers. The first version of LMS was implemented in 2015 in KPU. Feedback Management System (FMS) is a web-based application built in the CodeIgniter framework and is designed to help managers of KPU to evaluate the faculties, departments, and lectures by students and other lectures. This application stores data about faculties, departments, lecturers, students, and courses for each lecturer. This system started to operate in 2017. Dormitory Management System (DMS) of KPU is designed to facilitate the procedures of registering, assigning, and attendance tracking of students in the dormitory. In addition to other information, this system mainly stores student’s data and it started to operate in March 2018. While these systems help KPU staff in related procedures and provide many facilities, they all suffer the main problem that is an inability in communication between each other. Because there is no sharing of data or processes, each system stores and manages its data. The resulting process and data redundancy lead to data integrity problems. In turn, this reduces the effectiveness of the data for decision-making and analysis [1]. This also leads to high operational costs caused by increased maintenance requirements [1]. MIS’s in KPU, mostly store data which is nearly the same in most cases. Therefore, for a student in KPU, we have at least three data records in three different databases and for a lecturer at least there are two records. So, the staff of KPU must do the same work three times for the same student to register, update or delete and do the same processes two times for a lecturer and other entities in MISs databases. Besides, in some cases, there may be different information for the same student which can cause inconsistency in organization data. Moreover, the redundant work of staff causes extra costs for the organization and it is a...
big challenge against using technology and mentioned systems in KPU and Afghanistan. The above-mentioned problems like duplicate data storage, the inconsistency of data, and the high salary of staff who run systems are great challenges against using information systems and technology in KPU. However, the selection of the most appropriate solution for each organization is problematic as not all integration requirements can be addressed by a single solution [2]. KPU stakeholders require an efficient method to realize and evaluate the abilities of each integration solution based on different integration requirements. Therefore, we decide to study and work on finding a solution to integrate them. Our research initially covers current information systems of Kabul Polytechnic University and upcoming information systems. This paper endeavors to help the integration of systems by giving evaluation and comparison of system integration approaches based on a categorization of integration solutions and integration requirements. The proposed integration prerequisites are derived from relevant case studies.

II. LITETURE REVIEW

Within the last decade, packaged software solutions and information systems have become very famous. Information systems designed and built for different purposes in every organization and many organizations depend greatly on them for overcoming with objectives and goals in the current competitive environment. But, using and creating this packaged software and information systems again caused “information islands” as before and even on a larger scale [3].

In a scenario, where you have two systems that store the same data. When there is a data change in one system, you should change that data in another system manually. For example, clerical personnel of a German bank was having to read data from on-screen in a system and enter them into another system which was very time-consuming and error-prone [4]. For improving the performance of information systems and solving the “information islands” problem, long term scalable solutions where required. Researchers and vendors worked on many different solutions and lots of solutions were suggested, designed, built, and implemented by experts in many areas. Methods and techniques that researchers suggest they were various during the time. There were two factors in integrating information systems which are a type of information system integration and level of information system integration [5].

A few years ago, the type of information system integration was confined to data integration, functional integration, and process integration [5]. But, now by emerging new forms of organizations and information systems such as social enterprise and social customer relationship management systems the need for new forms of integration emerged. Also, due to dynamics organizational changes as the result of factors like mergers, acquisitions, strategic alliances, and shared services may suggest a new meaning for the level of system integration. The integration method for integrating information systems in firms and organizations greatly depends on the scenario and kind of the systems which those firms and organizations use. The definition of system integration varies depending on the perspectives exist for information integration. First, integration in the information system field from a technical perspective is defined as the degree to which different systems of an organization are interconnected and are capable of communicating with each other (e.g. islands of technology integration) [5]. In the second perspective, integration is the degree to which two or more independent organizations have standardized business processes and those processes are firmly linked through telecommunications technologies and computers [6]. Information system integration targets at facilitating exchange and data sharing inside a company and accomplishing inter-organization coordination (between consumers and sellers) for better tracking functionality as in the case of supply chain [7].

Regarding technological integration, it’s been repeatedly burdened that information system integration needs all application systems, data, and communication to be integrated [8] to provide a real-time and regular connectivity inside function element throughout supply chains. Linß [9] and Rosemann [10] categorized integration into 3 major dimensions encompassing: domain, reach, and direction. As a result, in the direction dimension, we have horizontal or vertical integration and in reach can be either intra-organization (between organizations) or inter-organization (in a single organization). Domain dimension of system integration is either data-wise, function-wise, and, program-wise. It should be added that Rosemann [10] considers object-wise as another sub-dimension for the domain. At the same time as Picot, Reichwald, & Wigand [11] categorized it as ex-ante and ex-post integrations, similarly, Fuchs-Kittowski [12] grouped ex-ante integration into re-engineering and integrated components, and however classified ex-post-integration into sub-dimensions as data, function, and presentation. The system can be integrated into different levels in an organization whose most common and standard approaches integrate systems are method, data, application, and user interface levels [13].

A. Method Level

In this level, the aim is to aggregate common operations in different applications into a single application which fronts the integrated application. This method is used when we have a set of applications in which each one of them has a similar set of APIs or functional methods. Method-level integration is used in cases where we have CORBA technology or distributed components.

B. Data Level

In this level, we integrate the background data stores of wanted applications. This integration level is used when the applications in the integration process do not have APIs and client interfaces. Also having a good understanding of business operations that may affect the data model is necessary for this level.

C. Application Level

This is considered the best way forward for application integration, and it uses the integrated application’s integration frameworks and APIs. It is good to use since it is transparent to the integrated application and it preserves the application’s data integrity. The application interface allows you to invoke business logic to preserve data integrity.

D. User Interface Level

Here the integration logic applies to the user interface code and it can be scripting or proxy-based. In scripting based, the integration code is embedded into user interface component events, common with client/server applications such as PowerBuilder. This integration method is useful for cases where direct access to the database is not possible or the business logic is embedded in the user interface.

III. DATA INTEGRATION

Integrating and combining data sources into a single query interface which also known as information silos have been
existing from the early years of computer systems life. The first systems for interoperability between heterogeneous databases were designed in the early 1980s [14]. In 1991, at the University of Minnesota first data integration system based on structured metadata was created for use in integrated public use microdata series [14]. This system used a data warehouse approach that extracts, transforms, and loads data from heterogeneous databases into a single view schema. The data warehouse method is much less possible for datasets that might be regularly updated, requiring the extract, transform, load (ETL) process to be continuously re-accomplished for synchronization. As of 2010, a number of the work in data integration studies worries about the semantic integration problem. This problem addresses not the structuring of the structure of the integration, but how to solve semantic conflicts among heterogeneous databases [15]. In 2011 enhanced data model methodologies have been developed to eliminate the data isolation artifact and to promote the development of integrated data models [16]. Since 2011, data hub methods had been of greater interest than completely dependent (normally relational) enterprise records Warehouses. since 2013, data lake methods have risen to the level of data Hubs. Those techniques integrate unstructured or various data into one place, but do not always require an (often complex) master relational schema to shape and define all data in the Hub. We can define data integration as precisely and methodically mixing data from distinct sources, making it more useful and valuable than it was before. IBM provides a strong definition, stating “Data integration is the combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information.” In IBM's definition, the key terms are “combing data … into meaningful and valuable information” [17]. That’s not just about shifting information from one location to another or pouring numerous spouts of information right into a single repository. It’s approximately making the information complete and more usable. In data integration, we can use different techniques and methods. But some methods are more reliable and popular than others. The following are the main methods and techniques of data integration [18].

A. Data Virtualization Integration
Data Virtualization lets us leave information in the source systems while permitting to create a new set of unified views. This gives a manner for customers to access the unified view of disparate source system’s data throughout the whole organization.

B. Data Warehouse
This technique requires the introduction of a new Data Warehouse (of data Marts) which stores a unified model of data extracted from all the source systems involved and controls it independent of the source systems.

E. Middleware Data Integration
A middleware data integration approach is a layer among two disparate systems permitting them to communicate. Middleware integration can act like a glue that holds together multiple legacy applications, making seamless connectivity possible without requiring the two applications to communicate immediately.

F. Manual Data Integration
This is not a data integration approach. In this method, a web-based user interface or an application is created for clients of the system to show them all of the relevant data through getting access to all the source systems immediately. There is no unification of data in reality.

IV. METHODOLOGY
This article aims to find how to integrate MIS’s in KPU to have a set of systems that do not have duplicate processes, data duplication, data inconsistency, and high resources, and labor cost problems and also, they can facilitate using technology for meeting organizational goals. The methodology which we used for answering our research question in this work is a combination of secondary data and Library research. Our secondary data resources are data of vendors and related projects in the area. We used electronic library resources and library research resources and following bibliographic databases. In answer to our question, we also had interview sessions with experts in the field of software development and information system. In addition to the interview sessions, we worked on a case study. After gathering data about our case study, we tried to analyze the data and find the best solution for our case study. The data analysis approach which we used in this study is a content analysis of data. The reason why we choose this approach is because of the kind of data we have collected. We collected our data using library research, documentation review, and interviews. With contents analysis, we can easily prepare our themes and codes form the above kinds of data. Therefore, this approach best fits for our research and study.

V. FINDING AND RESULT
After preparing our research question we started to study our case study and related topics to our research question. In parallel to studying and collecting data for our research question we reviewed and analyzed existing systems’ code and documentation. We found and declare problems with existing systems that were in use in parallel in KPU. By considering problems in existing systems in KPU and requirements for integrating, we worked on two scenarios. First, we decided to integrate two existing systems in data level to solve the communication problem between them and facilitate using them for staff and students of Kabul Polytechnic University. In the second scenario, we aimed to work and find a solution for integrating all current and upcoming information systems of Kabul Polytechnic University.

A. Integration System Using API
For solving problems and achieving our goals “having at least two systems which produce integrated data” decided to work for a solution to integrate real systems. In this solution, we tried to integrate two management information system which is being used by staff and students at Kabul Polytechnic University. Systems that we want to integrate them are a learning management system (Moodle) and a feedback management system. In the bellow paragraph, we briefly explained the mentioned systems to give a general overview.

Learning Management System (LMS): LMS is a web-based learning management system that is being used by lecturers and students of Kabul Polytechnic University to facilitate teaching and learning procedures. Moodle is a free, open-source package and developed by Martin Dougiamas and was released on 20 August 2002. Moodle is written in PHP and can operate with different RDBMS like MTSQL, PostgreSQL and MariaDB. Moodle is highly customizable to be used in different environments.
Feedback Management System (FMS): The feedback management system is designed to be used by students, lecturers, and executive personnel of KPU to evaluate lecturers of different faculties and departments. This system is written in CodeIgniter which is a PHP framework and uses MYSQL as a DBMS. Here in the feedback management system we mostly store and manipulate students, lectures, courses, faculties, departments and etc. The below figure shows the ER-diagram or Feedback Management system.

Figure.1. Feedback Management System Er Diagram

Mentioned systems store data about courses, students, lecturers, faculties, and other entities that are related to the academic environment. Problems within parallel using these systems is that they are not able to communicate with one another, which leads to problems like duplicate processes, duplicate data storage, data inconsistency and high resource and labor cost. Figure 5 shows those entities of FMS which exist or are similar to entities and fields in Moodle.

Above entities of Feedback Management System exist in mdl_user, mdl_course, mdl_course_categories, mdl_enrol, mdl_role, mdl_tag, mdl_assign and mdl_groups entities of Moodle. Although there are some differences and some fields entities of FMS are existed not in its pair in Moodle but they are in another entity in Moodle. After analyzing these two systems, we found that a great percent data of the feedback management system is stored and exists in Moodle. For finding the best solution to integrate these two systems and after considering various solutions, we resulted in a solution which is a combination of two techniques. Our solution suggests doing bellow steps for having integrated systems in data level.

- Restructure the databases of target systems and find common and separate parts and entities
- Build an API for getting and passing data to and from LMS to FMS
- Build a controller which decide to where to send the request (to the common part or separate database)

We used a RESTful API to integrate our systems plus some database restructure. We used PHP for writing our API codes on the Moodle site and we used CodeIgniter in FMS. For enabling the CodeIgniter to send HTTP request we used the Guzzle library. Our API has two main parts. The first part is on the Moodle site and it is responsible for supplying and determining where to send the request which came from API in FMS. The second part is codes and logics which sends requests to access and manipulate data from Feedback Management System. This part after receiving the data from the Moodle side usually does actions like passing it to views, updating views, compute and process data and then pass to the Moodle side for storing in database, etc. Figure 3 illustrates the general structure of our API solution for integrating two systems. As illustrated in the Figure 3, there is a common section between databases of the above systems. That’s the data that is the same in both databases. We restructured our databases and wrote codes which send request from FMS to LMS and there is a controller which determine if the data of request is in common section or a separate section of the databases. Then we supply the request with a proper place (database) to do related activities. We did some coding and built our API. The results of this experiment were really acceptable. We solved many existed problems. For example, now with using API, we can create a user in FMS or Moodle and use it in both systems. Also, we can do the update, delete, and insert users in one system and see its results in both systems. In short, this solution while is a little hard to implement, but solves a great number of our problems with existing systems in cases like data duplication, duplicate...
processes, data inconsistency, and in some cases usage cost. Some negative sides of this approach are its difficulty in implementation and a bit lower speed performance due to computations in the controller to determine the data source to store data for FMS. But the Moodle works as before and there is no effect on its performance.

B. Middleware Based Platform

The second scenario concerns with finding a solution to integrate all current and upcoming information systems of Kabul Polytechnic University. Considering our case study and based on studies we did, we found the best possible solution for integrating all information systems (currently used and near upcoming ones) in designing and implementing a middleware-based platform to integrate information systems at all levels. For implementing our platform to have a unified information system, we would require our platform and unified information system to have the following core elements and characteristics. Simple Infrastructure Integration: Such as hardware platform, software platform and security system to be integrated.

C. Analysis of Data and Results

After analyzing the collected data and methodologies the following points suggested as the result:

- System integration is a high demand for enabling systems in our case study to communicate with each other. Because currently used systems in Kabul Polytechnic University are designed and implemented by different vendors and individuals and they are not designed properly. They did not follow the standards of software engineering and system engineering principals. Integrating them at all level are not possible without redesigning.
- After studying systems in our case study, we found that for current systems and existing issues and requirements in them, the best level for integrating our systems is the data level. Considering problems with existing systems that are mainly related to data and processes which manipulate those data. Most of the issues are about the inability of systems in providing mechanisms to handle data properly and prevent problems like duplicate data storage, duplicate processes for manipulating data, data inconsistency, and data storage cost. Therefore, considering the requirements and structure of existing systems we found that data-level integration is possible for our case study and also appropriate enough to fulfill our requirements.
- For the scenario to find a solution for integrating all current and upcoming systems in our case study, with considering structures, requirements, and infrastructure, we preferred to use a middleware solution. Because using this technique will enable us to integrate our systems in Kabul Polytechnic University at all levels and uniformly which is the primary requirement by authorities in the long term. Although, this is just a suggestion within the ideal situation and condition. We offered this technique considering that there would enough capability and plan for design and implementing it.
- We choose API and database restructing which is a customized approach of synchronization method for data integration, because it supplies our requirement and it is possible with knowledge, skills, and infrastructure we have in our organization. Also, this is technically the best solution, because it solves 80 percent of our issues in our case study, and implementing it technically is possible. This RESTful API would integrate systems in data level and eliminate data inconsistency, remove duplicate processes, and duplicate data storage. Its implementation cost is lower than other approaches. However, the downside of this solution is not very considerable, but still, there are two negative sides to this solution. The smoothly lower performance of Feedback Management System and implementation complexity in the database restructure phase are the disadvantages of this solution.

- The data migration approach would not fulfill our requirements, because it removes duplicate processes in our systems but it would not help us in duplicate data storage. By using data migration, we could migrate data from Moodle to FMS and vice versa, which solves duplicate processes in our systems, but still, duplicate data storage and data inconsistency exist in our systems. Therefore, we were not able to solve all issues in our case study with this approach.
- Data-warehousing in suitable for cases where the enterprise and organization are going to store historical data and changes in data and use them in prediction in their business. In Kabul Polytechnic University which is an academic organization, historical data and their changes are not a requirement and still, there is no need for prediction. Therefore, data-ware house which is suitable for mentioned firms and organization is not proper in our case.
- Middle-ware integration can be used in our case study but using it in the current situation where we just have two sample systems to integrate, is not effective. Because implementing middle-ware for integration is hard and complex. We had limited time, skills, and infrastructure. Besides, we also need business changes and integration for this solution which is not possible in the current situation in our case study.

D. Comparision of the suggested solutions

In the bellow table, we compared our solution with other approaches with considering its possibility and requirements of our case study.

<table>
<thead>
<tr>
<th>Table. 1.comparison of different data integration approaches for the case study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features</td>
</tr>
<tr>
<td>Methods</td>
</tr>
<tr>
<td>Remove Data Duplication</td>
</tr>
<tr>
<td>Remove Duplicate Processes</td>
</tr>
<tr>
<td>Provide Data Consistency</td>
</tr>
<tr>
<td>Decrease Cost</td>
</tr>
<tr>
<td>Implementation Complexity</td>
</tr>
</tbody>
</table>

In the bellow table, we briefly compared systems in our case study after and before implementing our solution.
Table 2: Comparison Of Information System Before And After Implementing Api Solution

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Information Systems before API</th>
<th>Information Systems After Implementing API</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>Better</td>
<td>Slower in FMS</td>
</tr>
<tr>
<td>Duplicate Processes</td>
<td>Less</td>
<td>Greatly removed</td>
</tr>
<tr>
<td>Complexity</td>
<td>More</td>
<td>Less</td>
</tr>
<tr>
<td>Data Consistency</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data Storage</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Maintenance Cost</td>
<td>High</td>
<td>Low</td>
</tr>
</tbody>
</table>

From the above comparison, we can conclude that using API and restructuring databases of information systems in our case study greatly enhance and utilize their performance and solve many issues related to information silos.

VI. CONCLUSION

In recent decades especially in the last decade using information technology and information systems gained popularity and an important role in organizations and enterprises in the world. Although, in our country, Afghanistan using technology is not as popular as other societies, in recent years’ firms and organizations tried to implement information systems to modernize their business and structures and achieve their objectives. One of the academic organizations in Kabul which uses information technology as a tool to meet its aims is Kabul Polytechnic University. This University currently uses three information systems for management and learning procedures. Existing systems suffer from an inability to communicate with each other. System integration at all levels, especially in data level for removing information silos in Kabul Polytechnic University is a critical requirement. Therefore, research on integrating information systems in KPU has been designed and completed. For integrating current information systems in data level for a limited number of systems with similarities in data entities which is true about current systems of KPU we used database restructuring and a RESTful API. Our proposed API in the test was successful and solved a great number of issues with current systems integration in the data level. For having a complete set of information systems in Kabul Polytechnic University, considering the scenario that the demand for information systems is high and in near future, the number of information systems would certainly increase, we worked on prepare a solution which can prepare the base for integrating all current and upcoming information systems. We suggested a middle-ware based platform for integrating all systems. Although, implementing this platform currently is not possible, but in near future it can be implemented and assist staff and students of KPU with their aims and objectives. Implementing this solution needs more research and work to be done and it is not easy to design and implement it. By comparing our solution by other methods and techniques we find that implementing API based integration on current systems is appropriate in our scenario. Integrated system with API solves data duplication, duplicate processes, decreases costs, and provides data consistency but affects the performance of the Feedback Management System and decreases its response time.

Following are suggestions we have for implementing our solution:
- Start to plan for implementing the middle-ware based platform to facilitate feature related works.
- Consider issues related to system integration in designing and building new systems for Kabul Polytechnic University
- Prepare better and more facilities and infrastructures for doing more similar research and study on related topics.

VII. REFERENCES


