International Journal of Institutional Research and Management
International Institute of Applied Informatics
2024, Vol. 8, No. 1, IJIRM839

Development of an Automated Institutional Research System for Institutional Decision-Making in Kyushu Institute of Technology

Tetsuya Oishi* , Hirofumi Fukumaru*,
Akihiro Hayashi*, Hiroshi Sakamoto*

Abstract

A system has been established at the Kyushu Institute of Technology (Kyutech) that enables both students and teachers to reflect on their learning history and complete self-evaluations using e-portfolio systems. However, the university has not yet implemented institutional research (IR) for decision-making at the institutional level. To address this gap, the Learning Teaching Center at the university decided to establish an Educational IR Support Group in FY 2022, aiming to strengthen IR in education. In addition to e-portfolio systems, various data are accumulated across different systems within the university. However, there were no integrated systems to utilize these data comprehensively, as various departments managed them individually. To overcome this, we introduced an extract transform load (ETL) tool as part of the initial IR system. Simultaneously, we implemented a data lake system as a centralized repository for storing data. The Educational IR Support Group is leading this project. This study begins by providing an overview of IR and how it should be conducted. It then presents the IR system being promoted at our university, including the introduction of the ETL tool and the establishment of a data lake. Finally, we describe our future plans and developments, which involve the potential introduction of a data warehouse and business intelligence tools.

Keywords: Institutional Research for Education, Educational Data, Institutional Decision-Making, Information Support Cycle, Visualization, Automated Institutional Research System

1 Introduction

A paper-based self-assessment for students was initiated in 2003 at the Kyushu Institute of Technology (Kyutech) [1]. In 2005, the School of Computer Science and Systems Engineering received the Japan Accreditation Board for Engineering Education (JABEE) [2] certification, and the self-assessment was implemented within the department. Subsequently, a learning self-assessment system, functioning as an e-portfolio system for students, commenced development in 2007. By 2010, our university integrated the learning

* Kyushu Institute of Technology, Fukuoka, Japan

self-assessment system, capable of recording extracurricular activities, into university-wide operations.

To enhance the convenience of use of the learning self-assessment system, the interface was revamped in 2012, and the system began operating with the new interface in 2013. Continuously modifying and adding functions, the course portfolio system, serving as a portfolio for faculty members, was developed in 2015. Moreover, a syllabus system was developed in 2016, and collaboration among the learning self-assessment system, the course portfolio system for faculty members, and the syllabus system was reinforced by 2019. Furthermore, a course questionnaire survey function was incorporated into the learning self-assessment system, allowing faculty members to confirm the results obtained from this questionnaire survey in the course portfolio system in 2022.

While our university has systems to collect and provide educational data feedback to individuals, no systems were in place for supporting institutional decision-making. Consequently, the Learning-Teaching Center at our university decided to establish an Educational Institutional Research (IR) Support Group to strengthen IR from 2022.

This study not only provides an overview of IR but also insights into IR at Kyutech and discusses future research directions.

2 Institutional Research

In recent years, the population of 18-year-olds has been decreasing in Japan. However, many higher education institutions must sustain themselves despite this demographic shift. Consequently, each institution has developed its own IR organizations or functions to overcome this situation. The information support cycle [3], a commonly employed IR model, comprises five phases listed below, and repeating these phases can support institutional decision-making:

1. Identifying issues and needs
2. Collecting and accumulating data
3. Reconstructing and analyzing data
4. Reporting by using data
5. Making decisions

This cycle highlights the connection between previous decisions and the emergence of new issues and needs, forming the basis for making new decisions. These tasks are divided between university executives overseeing each institution and IR practitioners. While the university executives' tasks involve "1. Identifying issues and needs" and "5. Making decisions", IR practitioners' tasks involve "2. Collecting and accumulating data", "3. Reconstructing and analyzing data", and "4. Reporting by using data". IR is a system that supports decision-makers in each institution, often referred to as educational IR when dealing specifically with educational data. The primary tasks of IR practitioners involve working with data, not managing it. Data management depends on the operation generating the data and incurs significant costs. IR practitioners do not manage data because they focus on collecting, accumulating, reconstructing, analyzing, and reporting data. In the following section, we elaborate on the specific tasks performed by IR practitioners.

2.1 Collecting and Accumulating Data

Data in higher education institutions are classified into unstructured and structured data [4]. Unstructured data includes original papers, university regulations, and the like, mainly consisting of text data. Structured data includes competitive external funding, student questionnaire surveys, and other numerical values.

IR practitioners collect and accumulate these unstructured and structured data into a system called a data lake [5], primarily dealing with structured data such as relational database data, Comma-Separated Values (CSV), eXtensible Markup Language (XML), and so on. With permission to access a database server, they can use data directly from the server.

2.2 Reconstructing and Analyzing Data

Using an Extract Transform Load (ETL) tool, IR practitioners combine data collected in the data lake. Then they restructure the data to facilitate analysis for reporting data, storing the restructured data in a system called a data warehouse [6]. Data warehouses store data for decision-making. The data are organized by theme, integrated, in chronological order, not deleted, and not updated—essentially providing tidy data for easy analysis and visualization to support decision-making.

IR practitioners with limited experience may not find reconstructing data interesting, whereas those with enough experience consider the ETL tool, used in the reconstructing data phase, important [7]. While some believe that IR practitioners should focus on analyzing data meticulously, others argue that preparing the data for analysis is crucial.

2.3 Reporting by Using Data

IR practitioners report data to decision-makers by visualizing them as necessary. When reporting data, they should visualize the data from various perspectives to enhance understanding. The data stored in the data warehouse for visualization should not only cover various topics but also be presented in processable forms. Introducing a Business Intelligence (BI) tool [8] for data visualization can enhance the real-time nature of reporting and streamline the data visualization process. This, in turn, helps decision-makers make faster decisions, ultimately improving the functioning of each institution.

3 Educational IR in Kyutech

To support educational initiatives at Kyutech, the Learning and Teaching Center has been in operation since 2011 [9]. This section discusses the center (Section 3.1), educational support systems (Section 3.2), and the Educational IR Support Group (Section 3.3).

3.1 Learning and Teaching Center

The Learning and Teaching Center at Kyutech consists of four groups, as shown in Figure 1. This Center supports the educational activities of each organization at Kyutech and their efforts to improve the learning environment, and aims to contribute to the improvement of the educational functions and quality of the university as a whole by promoting mutual coordination of these activities.

1. The Educational IR Support Group

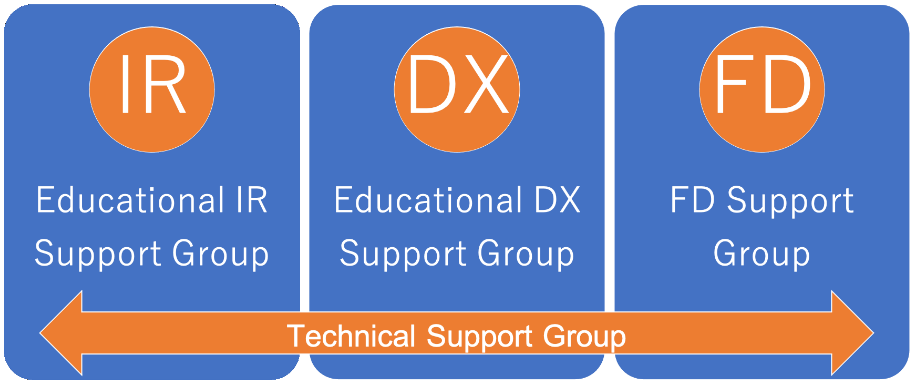

Figure 1: Four Groups Comprising the Learning and Teaching Center at Kyutech

2. The Educational DX (Digital Transformation) Support Group
3. The FD (Faculty Development) Support Group
4. The Technical Support Group

These four groups collaborate to manage this center.

3.1.1 Educational IR Support Group

The members of the Educational IR Support Group manage the educational support systems described in Section 3.2. Moreover, analyzing educational data for decision-making is a key task of this group—an IR task further described in Section 3.3. To promote the e-portfolio system in other universities, a consortium for visualizing educational outcomes through the e-portfolio system [10] was established in 2019 and is managed by this group. Through discussion and collaboration in this consortium, we aim to improve the quality of learning by developing e-portfolio methods and technologies and visualizing learning outcomes based on students' own reflections on their learning. Furthermore, by promoting the use of visualization of learning outcomes for educational quality assurance and academic management, and by disclosing information on educational reforms to society, we aim to contribute to the advancement of the educational system in Japanese higher education institutions.

The Consortium will focus on the following five main projects, and will implement the following specific activities in collaboration with different educational institutions or companies.

1. Activities to develop methods and technologies to visualize learning outcomes through e-portfolios
   (a) Development of methods and technologies to promote visualization of education as practiced by educational institutions
   (b) Development of mechanisms to be used for educational quality assurance and teaching and learning management
2. Activities to improve the quality of learning based on the quality assurance of learning

   (a) Construction of mechanisms to incorporate social evaluation of learner-oriented education
   (b) Activities to promote the use of the results and good practices in career development

3. Activities to disclose information on educational reforms to society, including efforts to guarantee the quality of education
   (a) Establishment of a mechanism to disseminate visualized educational information to society
   (b) Publicize and promote good examples of ICT application by educational institutions

4. Holding symposiums, forums, etc.
   (a) Organize and disseminate information by holding symposiums and forums about once a year, and practice social evaluation.

5. Other activities necessary to achieve the objectives

3.1.2 Educational DX Support Group

Members of the Educational Digital Transformation (DX) Support Group promote the use of various learning support systems, such as Moodle (as discussed in Section 3.2). Their focus includes promoting system integration to realize efficient education. This group is actively developing infrastructure to support the creation and collection of reusable teaching materials in electronic media. In addition, the group facilitates the development of an environment for highly interactive lectures, that are highly interactive with learners, provides and supports the operation of new educational environments in cooperation with other departments, and performs tasks related to the infrastructure development for learning and education.

3.1.3 FD Support Group

The Faculty Development (FD) Support Group conducts seminars and training events for faculty members at University A. This group collects and shares information on unique educational projects in each department, fostering collaboration among these projects. In addition, the group provides support for activities related to curriculum development, syllabus preparation, improvement of teaching methods and evaluation, student counseling, and career guidance, among others.

3.1.4 Technical Support Group

Finally, the members of the Technical Support Group provide assistance to the Educational IR Support Group, Educational DX Support Group, and FD Support Group. This group collaborates with related departments to share information and offer support, ensuring more efficient and labor-saving operations of businesses and systems.

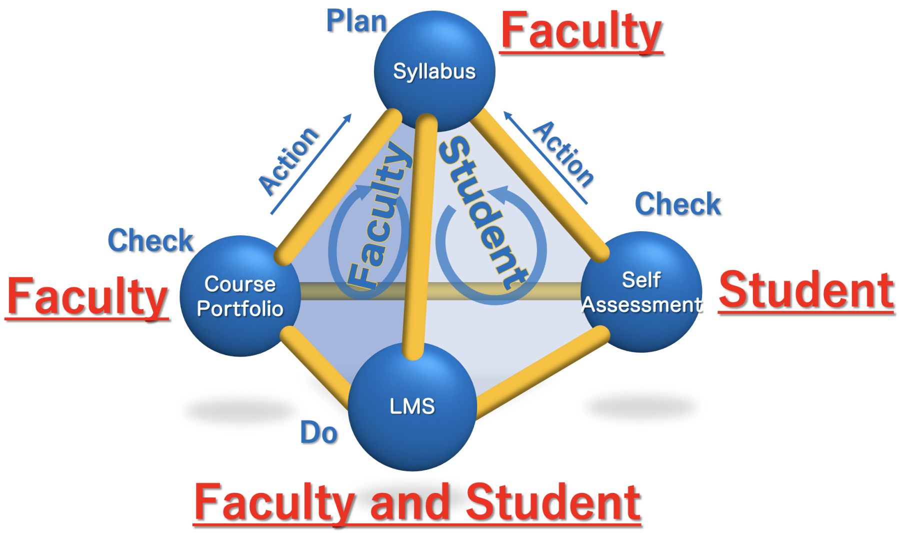

Figure 2: Correlation Chart of Education Support Systems

3.2 Education Support Systems

The Learning and Teaching Center at Kyutech has four educational support systems, listed below and illustrated in Figure 2.

1. Syllabus System
2. Learning Management System (LMS)
3. Course Portfolio
4. Self-Assessment system

3.2.1 Syllabus System

Located at the top of Figure 2 the syllabus system enables faculty members to showcase class outlines and lesson plans as shown in Figure 3. It allows faculty to stipulate the positioning of classes based on the Diploma Policy and Curriculum Policy, specifying class achievement objectives but evaluation criteria.

The system follows the PDCA (Plan, Do, Check, and Action) management approach. This way is used for quality management, where the syllabus system corresponds to the "Plan" phase.

3.2.2 LMS

The LMS is located at the bottom of Figure 2. Using LMS, faculty members can share educational information linked to the syllabus with students. Faculty members use the LMS to gather assignments, accumulate learning outcomes, and provide students with learning materials. Kyutech has adopted Moodle 4.1 [11] as shown in Figure 4, an open-source learning platform, as its LMS. In the PDCA cycle, the LMS corresponds to the "Do" phase.

Figure 3: Screen Shot of Syllabus System for both the Students and the Faculty Members (Japanese)

Figure 4: Screen Shot of Moodle 4.1 for both the Students and the Faculty Members (Japanese)

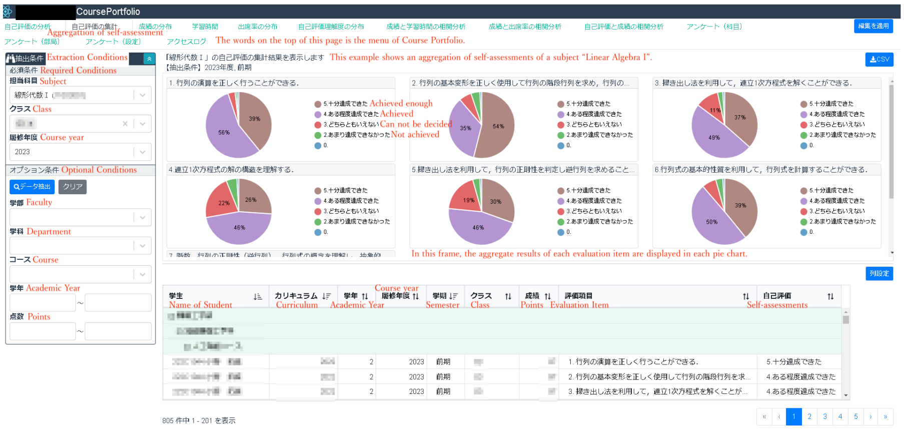

Figure 5: Screen Shot of Course Portfolio System for the Faculty Members (Japanese)

3.2.3 *Course Portfolio*

Kyutech has two e-portfolio systems: a Course Portfolio and a Self-Assessment System.

The Course Portfolio, located on the left side of Figure 2, enables faculty members to analyze not only grades but also students' self-assessments as shown in Figure 5. Faculty members can review achievement goals, assess learning outcomes, and evaluate educational effectiveness, corresponding to the "Check" phase of the PDCA cycle.

Contents checked in the Course Portfolio can be reflected in the syllabus system, an activity corresponding to the "Action" phase.

3.2.4 *Self-Assessment system*

Located on the right side of Figure 2, the Self-assessment system serves as a portfolio system for students. Students can set learning goals, self-evaluate achievement, and review grade history as shown in Figure 6. The system enables reflection on learning and allows the recording of both regular and extracurricular activities. In 2022, the system was redeveloped for graduate students, enabling them to create research plans and report results. This corresponds to the "Check" phase of the PDCA for students, and students can provide feedback through classroom questionnaire surveys. Faculty members can check survey responses and reflect them in the syllabus system, corresponding to the "Action" phase of the PDCA cycle for students.

3.3 Activities of Educational IR Support Group

Kyutech has gathered data on education, supporting individual decision-making for faculty members and students. Nevertheless, there is a lack of mechanisms supporting institutional decision-making for managerial staff. To address this issue, the Educational IR Support Group was established in 2022 within the Learning and Teaching Center to support institutional decision-making.

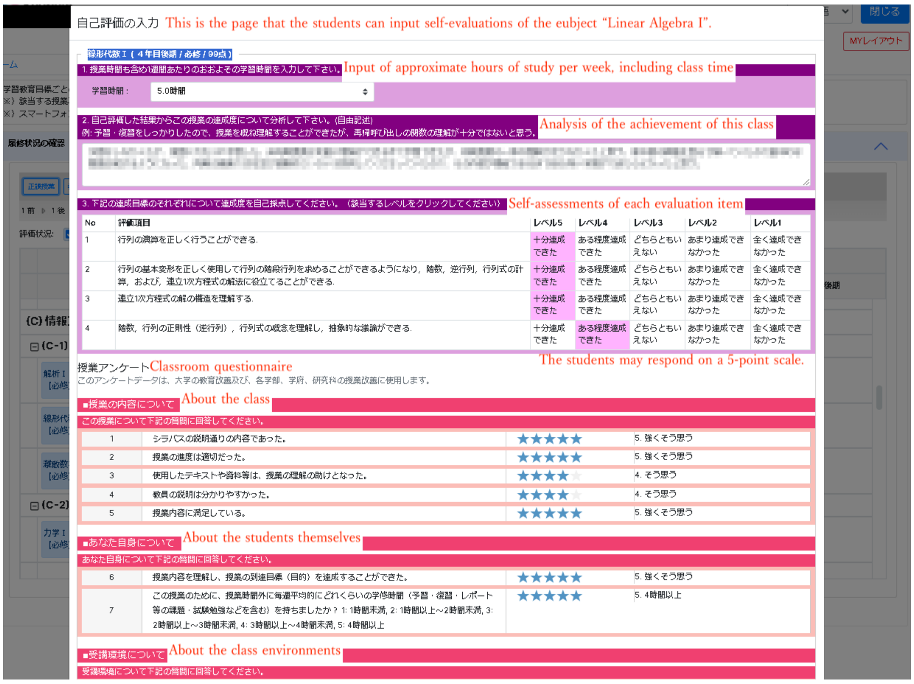

Figure 6: Screen Shot of Self-Assessment System for the Students (Japanese)

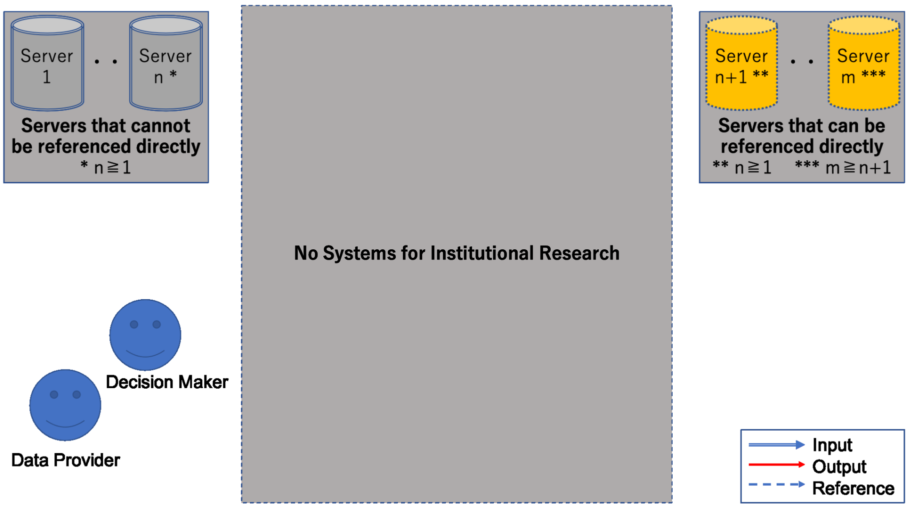

Figure 7: Server Configurations until 2021

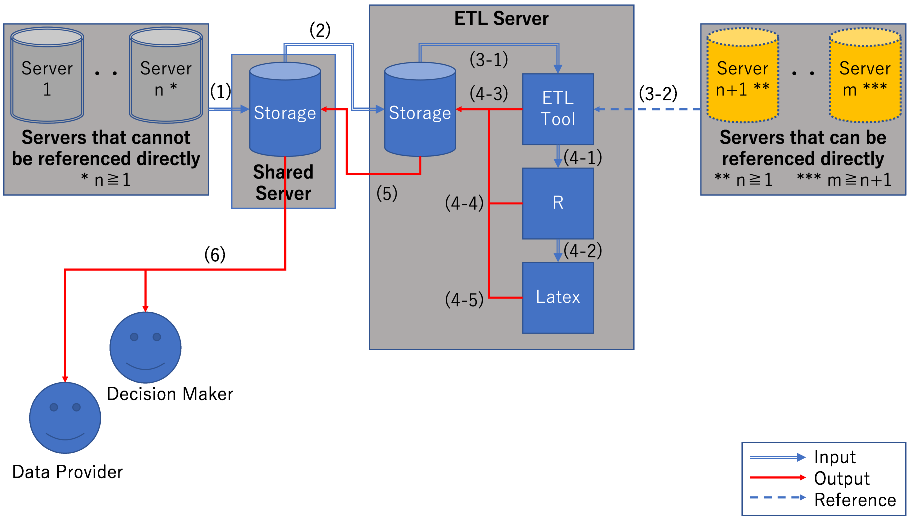

Figure 8: Initial Server Configurations

3.3.1 Server configurations until 2021

Although many systems were in operation in Kyutech until 2021, no IR system existed to connect them as shown in the Figure 7. Therefore, we proceeded to build the IR system step by step as explained in the next section.

3.3.2 Initial server configurations

To facilitate the analysis of educational data, we first introduced the ETL tool, as described in Section 2.2. It is essential to establish a mechanism for gathering and storing data using a data lake, as indicated in Section 2.1. Moreover, a mechanism for storing reconstructed data using a data warehouse, as detailed in Section 2.2, is also necessary. However, Kyutech previously lacked these mechanisms. Then, we decided to temporarily integrate these mechanisms into the server running the ETL tool. Utilizing the statistical analysis software R [12], we introduced a semi-automatic mechanism to visualize the combined data generated through the ETL tool.

To realize the semi-automatic mechanism for visualizing data, we introduced initial systems following Figure 8. (1) The data providers transfer the data from servers that cannot be referenced directly to the shared server, and (2) IR practitioners manually store these data in the storage in the ETL server. Not only (3-1) data in the storage in the ETL server but also (3-2) data from servers that can be referenced directly are combined using the ETL tool. These data are (4-1) analyzed and visualized using the statistical analysis software R, and (4-2) converted to PDF using LaTeX [13]. (4-3) The data combined with the ETL tool, (4-4) the data visualized by R, and (4-5) the data converted to a PDF by LaTeX are stored in the storage in the ETL server. Then (5) IR practitioners store them on a shared server. (6) Data providers and decision-makers can retrieve reported data from a shared server.

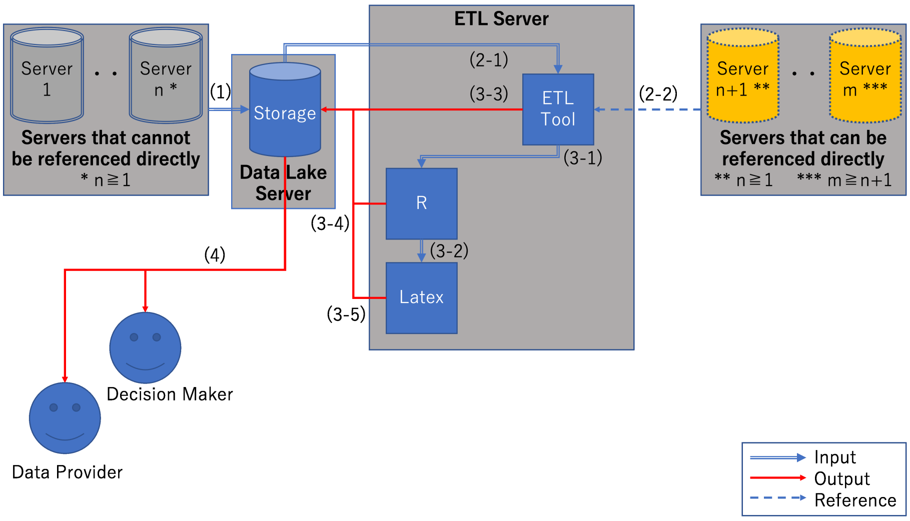

Figure 9: Current Server Configurations

3.3.3 *Current server configurations*

At the time of present writing, we have already introduced the data lake server but have not introduced a data warehouse server and a business intelligence server. Therefore, the data lake server must assume the roles of both the data warehouse server and the business intelligence server. The data lake server is designed to fulfill the following functions:

1. Accumulating various data: This is the primary function of the data lake.
2. Providing cleansed data: This is the function of a data warehouse.
3. Providing visualized data: This is the function of a business intelligence server.

Considering these functions, we need to control access to our data lake server. We required specific areas within the data lake server to satisfy the following conditions:

- The area that data providers can access.
- The area to provide cleansed data.
- The area to provide visualized data.

The most important condition is that IR practitioners should have access to all areas because they need these data for reporting for decision-making.

We introduced Nextcloud [14], which satisfies the terms and conditions, as our data lake server. Nextcloud is not only free software but also open-source software. Moreover, it can be installed on our private server. We can encrypt files during transfer using Nextcloud. Managing users through Lightweight Directory Access Protocol (LDAP) and defining file authorities are the defining characteristics of Nextcloud.

We have introduced current systems following Figure 9 to realize the semi-automatic mechanism for visualizing data: (1) The data providers transfer data from servers that cannot be directly referenced to the storage in the data lake server. Both (2-1) data in the storage in the data lake server and (2-2) data from servers that can be directly referenced

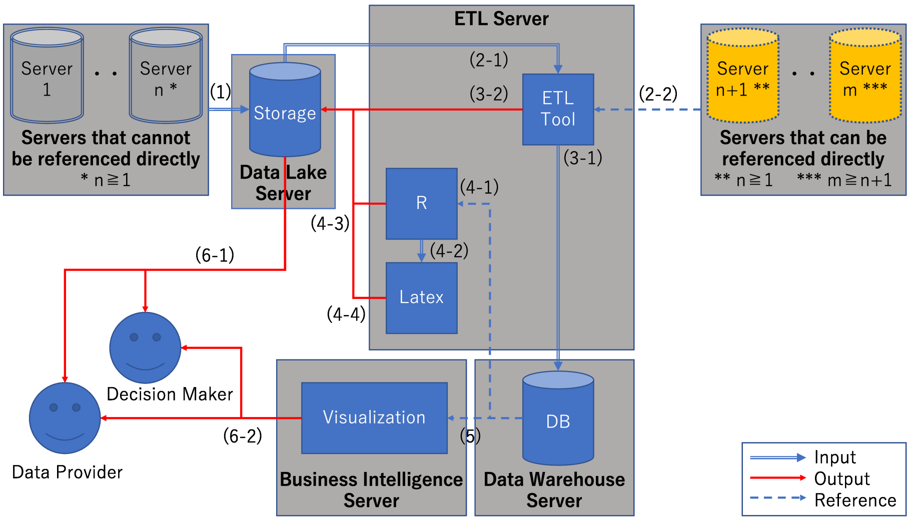

Figure 10: Future Server Configurations

are combined using the ETL tool. (3-1) The data combined by the ETL tool were analyzed and visualized using the statistical analysis software R and (3-2) converted to PDF using LaTeX. (3-3) The data combined by the ETL tool, (3-4) the data visualized by R, and (3-5) the data converted to PDF by LaTeX were stored again in the storage in the data lake server. (4) Data providers and decision-makers can retrieve reported data from the data lake server.

4 Future Work and Expected Effects

In this section, we discuss future work and expected effects.

4.1 Future Work

Figure 10 illustrates the future server configuration that can realize an automatic visualization mechanism.

(1) Data providers transfer data from servers that cannot be directly referenced to the storage in the data lake server, equivalent to "2. Data collection and accumulation" in Section 2.

Not only (2-1) data in the storage in the data lake server but also (2-2) data from servers that can be directly referenced are combined using the ETL tool. At this time, because the storage in the data lake server will be mounted on the ETL server, we will be able to use the data as if they were in the storage in the ETL server. Moreover, the data combined by the ETL tool are not only (3-1) stored in the database in the data warehouse server but also (3-2) again stored in the storage in the data lake server. (4-1) The data from the database in the data warehouse server were analyzed and visualized using the statistical analysis software R, and (4-2) converted to PDF using LaTeX. (4-3) The data visualized by R and (4-4) the data converted to PDF by LaTeX were again stored in the storage in the data lake server, equivalent to "3. Data reconstruction and analysis" in Section 2. The presence of R and LaTeX in the ETL server, despite the Business Intelligence function described below, enables flexible visualization for urgent requests.

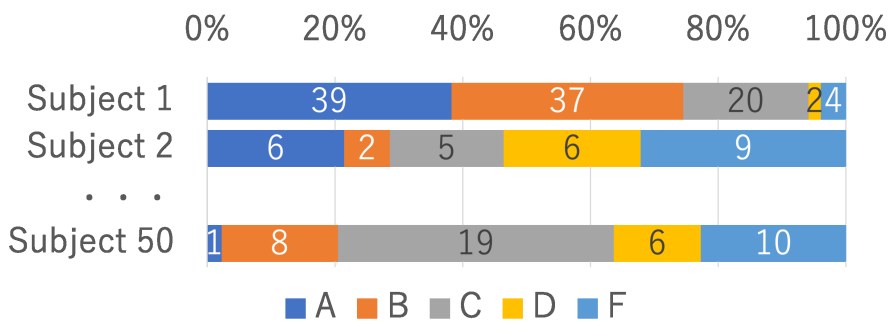

Figure 11: Example of Grade Distribution Report

Table 1: Time Required to Report a Grade Distribution

	Time Required One Session	Four Sessions	Time Reduced Compared to (1)
(1) Until 2021 (Figure 7)	4 hours	16 hours	-
(2) Initial (Figure 8)	2.25 hours	9 hours	7 hours
(3) Current (Figure 9)	0.25 hours	1 hour	15 hours
(4) Future (Figure 10)	0 hours	0 hours	16 hours

On the contrary, (5) the data from the database in the data warehouse server can be interactively visualized on the business intelligence server. Data providers and decision-makers can not only (6-1) retrieve reported data from the data lake server but also (6-2) retrieve visualized data on demand using the visualizing system working in the business intelligence server, equivalent to “4. Data reporting” in Section 2.

4.2 Expected Effects

We provided decision-makers with the grade distribution shown in Figure 11 [15]. Here, decision-makers comprise faculty members and managers of each department. We will illustrate the expected effects by using this visualization in this section.

Kyutech consists of approximately 2,500 subjects per year. For simplicity, we assumed 2,400 subjects. We typically present the grade distribution on four separate occasions because our curriculum is divided into quarters. This indicates that we have to present the grade distribution for 600 subjects on each occasion.

The original data are stored in an educational affairs system. IR practitioners are not permitted to access this system directly. Consequently, IR practitioners must obtain the data from the data provider, which requires two hours to extract the original data for grade distribution. To visualize the grade distribution for 50 subjects, as shown in Figure 11, we required only 10 minutes. In other words, it takes two hours to visualize 600 subjects, which is equivalent to 12 pages. The total time is 16 hours per year, considering the quarterly presentation of these grade distributions. Given that the data size is modest, the time required to transfer data is insignificant.

Next, we present the visualization using the initial server configuration shown in Figure

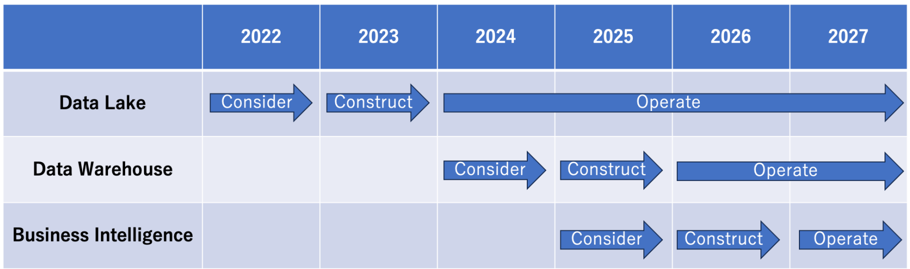

Figure 12: Developping Plan

8. While the data extraction time of the data provider is maintained at two hours, the visualization task is reduced to 0.25 hours due to automation by the ETL tool and R. This translates to nine hours of work per year, considering it requires 2.25 hours of work per session. This indicates a reduction of seven hours of work per year, as shown in Table 1 (2).

The required time to extract data by the data provider is eliminated through direct reference to the database as shown in Figure 9 , meaning the extraction time becomes 0 hours. Consequently, we only need 0.25 hours for visualization work per operation. Over the course of one year, this equates to only one hour, as shown in Table 1 (3).

Moreover, we describe the visualization using a future server configuration, as shown in Figure 10. The cost of maintaining an ETL server is required for the functions discussed thus far. Additional maintenance costs for data warehouse and business intelligence servers will be necessary to realize these functions moving forward. If data warehouse and business intelligence servers are introduced, decision-makers can observe the cleansed data or visualized data that is automatically created. These operations require 0 hours, as shown in Table 1 (4). In essence, one visualization task saves 16 hours, highlighting the potential to save $16n$ hours for n visualization tasks.

5 Summary

This study aimed to provide an overview of the IR landscape and the current status of the IR system at Kyushu Institute of Technology (Kyutech). The IR system in Kyutech is still in its developmental stage, and we are actively engaged in constructing it through a trial-and-error process. The recent introductions of the ETL tool for combining data and the data lake system, built using the open-source data-storing software Nextcloud, mark significant progress in our journey. These systems play a critical role in reducing the task of visualizing data, showcasing an important use case for digital transformation. However, it is essential to acknowledge the dedication of our administrative staff and faculty members who have worked diligently in the background, underscoring the human costs associated with our operations.

In the future, we will consider the potential introduction of a data warehouse and business intelligence tools, tailored to our institutional needs while referring to examples of IR at other higher educational institutions. We considered the data lake and constructed it by 2023. As shown in the Figure 12, we plan to start considering a data warehouse in 2024, construct it in 2025, and start operating it in 2026. We would like to move to the phase of considering, constructing, and operating a business intelligence system from 2025, shifting the schedule by one year.

Currently, we are leveraging these systems to analyze educational data, with plans to combine admission data and campus life data for further analysis. As the diversity of data types and stakeholders increases, careful consideration must be given to avoid unnecessary complexity in management. The ultimate goal of this development is to enable decision makers at our University to easily view the results of data analysis and to speed up organizational decision making.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 21K02653.

References

[1] Hiroshi Sakamoto, Hirofumi Fukumaru, Takashi Miyaura, Kazunori Nishino, Akihiro Hayashi, "Use Case of Visualization of Learning Achievement Using e-Portfolio - Construction of Internal Quality Assurance of Education and Formation of Consortium by Diverse Educational Institutions -," Joint Forums between University e-Learning Association and The Japan Association for Development Education 2019, 2020, pp. 31-32.

[2] "JABEE," https://jabee.org/en/ (accessed 31st March, 2023).

[3] R.D.Howard (ed.), National Institution for Academic Degrees and Quality Enhancement of Higher Education, "Institutional Research: Decision Support in Higher Education," Tamagawa University Press, 2012.

[4] Yoshiharu Aioi, Shuntaro Iseri, Shotaro Imai, Tetsuya Oishi, Saori Okada, Nobuhiko Kondo, Toru Sugihara, Shintaro Tajiri, Mio Tsubakimoto, Keita Nishiyama, Takeshi Matsuda, Masao Mori, "Data Collection," Standard Guidebook for University IR, Japan Association for Institutional Research, 2022, pp.122-126.

[5] "Data lake," https://en.wikipedia.org/wiki/Data_lake (accessed 12th December, 2023).

[6] "Data warehouse," https://en.wikipedia.org/wiki/Data_warehouse (accessed 12th December, 2023).

[7] Tetsuya Oishi, "What is the Essential Curriculum for IR in Japan?," Proceedings - 2022 11th International Congress on Advanced Applied Informatics, IIAI-AAI 2022, LIR009, 2022, pp. 1-5.

[8] "Business intelligence," https://en.wikipedia.org/wiki/Business_intelligence (accessed 12th December, 2023).

[9] "Learning and Teaching Center," https://www.ltc.kyutech.ac.jp/ (accessed 12th December, 2023).

[10] "Consortium for visualizing educational outcomes by e-portfolio system," https://ekashika.org/ (accessed 12th December, 2023).

[11] "Moodle," https://moodle.org/ (accessed 4th April, 2023).

[12] "RStudio," https://posit.co/download/rstudio-desktop/ (accessed 31st March, 2023).

[13] "Latex," https://ja.wikipedia.org/wiki/LaTeX (accessed 22nd December, 2023).

[14] "Nextcloud." https://nextcloud.stylez.co.jp/ (accessed 16th December, 2023).

[15] Tetsuya Oishi, Hirofumi Fukumaru, Akihiro Hayashi, Hiroshi Sakamoto, "Establishment of an IR System for Institutional Decision-Making in Kyushu Institute of Technology," Proceedings - 2023 12th International Congress on Advanced Applied Informatics, IIAI-AAI 2023, LIR116, 2023, pp. 1-8.