IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol. 005, LIR387
DOI: https://doi.org/10.52731/lir.v005.387

Development and Operation of an All-in-One IR System

Tetsuya Oishi *

Abstract

In recent years, Institutional Research (IR) has become increasingly important for decision-making support in Japanese higher education institutions. This paper reports on the development and operation of an IR system at Kyushu Institute of Technology, which has been under construction since fiscal year 2022. The system integrates a data lake server with an ETL (Extract, Transform, Load) server to provide various functions including grade distribution visualization, automated infection report generation, and integration with academic affairs systems. The key feature of this system lies in its utilization of programmatically operable tools such as a data lake, an ETL Tool, a statistical software, and PDF creation to automate data collection, restructuring, analysis, and reporting processes within the information support cycle. This approach has improved operational efficiency and reduced costs at the university while establishing a foundation for data-driven decision-making. Future plans include system expansion with the introduction of data warehouse and BI tools.

Keywords: Institutional Research, Data Lake, ETL, Automation, Information Support Cycle, Data Visualization, Higher Education

1 Background

In recent years, Institutional Research (hereafter abbreviated as IR) has come to play an important role in higher education institutions in Japan. IR practitioners are assigned to each higher education institution, and they work diligently to improve their IR skills. There is a concept called the information support cycle [1] that is often used as a model when carrying out IR work. This involves executing organizational decision-making by repeating the following five phases.

- University executives “identify issues and needs.”
- Administrative organizations “collect and accumulate data.”
- IR practitioners “restructure and analyze data.”
- IR practitioners “report data” to university executives.

* Kyushu Institute of Technology, Fukuoka, Japan

- University executives "make decisions."

In fiscal year 2013, when IR began to spread in Japan, Kyushu University launched an "IR Human Resource Development Curriculum" [2][3][4] based on this information support cycle. This curriculum was designed for graduate students at Kyushu University and consisted of 5 courses totaling 10 credits corresponding to the information support cycle. The curriculum at Kyushu University was held until fiscal year 2016 due to faculty departures, but in its place, Tokyo Institute of Technology (currently Tokyo University of Science) began offering Institutional Research Theory [5][6] from fiscal year 2019, which continues to this day. Among IR practitioners, some not only effectively utilize such training opportunities but also engage in self-study using books like the University IR Standard Guidebook [7]. It is known that there are many other IR training opportunities in Japan besides those mentioned above [8][9].

On the other hand, it is also known that the needs for the training itself are becoming more diverse. A web survey on IR workshops in fiscal year 2022 [10][11] revealed that IR practitioners prefer workshops on data restructuring and analysis content that last at most one overnight stay, with about two sessions per day. When individual interview surveys [12][13] were conducted as an extension of this web survey, it was found that IR practitioners not only wanted level-specific workshops that included visualization content in addition to data restructuring and analysis, with content that helped them understand analysis methods, visualization, and data interpretation, but also wanted to interact and share feelings with other participants. Based on these needs, attempts were made to design workshops, but it became clear that it was difficult to uniquely determine the difficulty level of the training, so the idea of designing level-specific workshops was abandoned.

Instead, in fiscal year 2023, workshops focusing on IR practitioners' self-recognition of IR skills were held [14][15]. In these workshops, participants were each given materials related to data analysis [16][17], and after discussing the content, not only were the difficulty level and comprehension investigated, but also the discrepancy in self-recognition of IR skills before and after the workshop. Additionally, similar workshops were held in fiscal year 2024 [18][19][20], and data collection including participant attributes was attempted. These results showed that self-recognition of IR skills differed depending on whether participants were faculty or staff members and their duration of involvement in IR [21].

The author is a person with relatively long experience in IR who has been involved in initiatives like those mentioned above. The author belongs to the Learning and Teaching Center at Kyushu Institute of Technology [22], and while performing educational IR, also manages the Consortium for Visualization of Learning Outcomes through e-Portfolios [23]. At Kyushu Institute of Technology, the development of an IR system based on the information support cycle is being advanced with the author at the center, and first, the data storage component was developed [24][25]. This paper explains the implementation methods based on cases of visualization and data provision using the IR system at Kyushu Institute of Technology.

2 Related System Groups

Kyushu Institute of Technology began building its IR system from fiscal year 2022. As of 2025, an IR system linking a data lake [26] server and ETL [27] server (ETL stands for Extract, Transform, Load) is operating as shown in Figure 1. As shown in Figure 2, data warehouse [28] and BI [29] will also be added in the future to improve decision-making

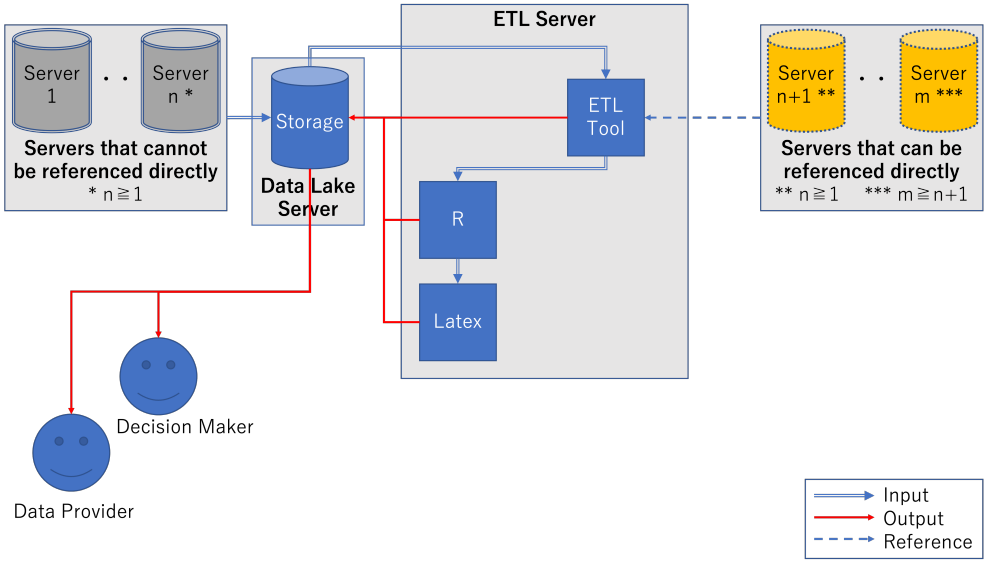

Figure 1: Current IR System

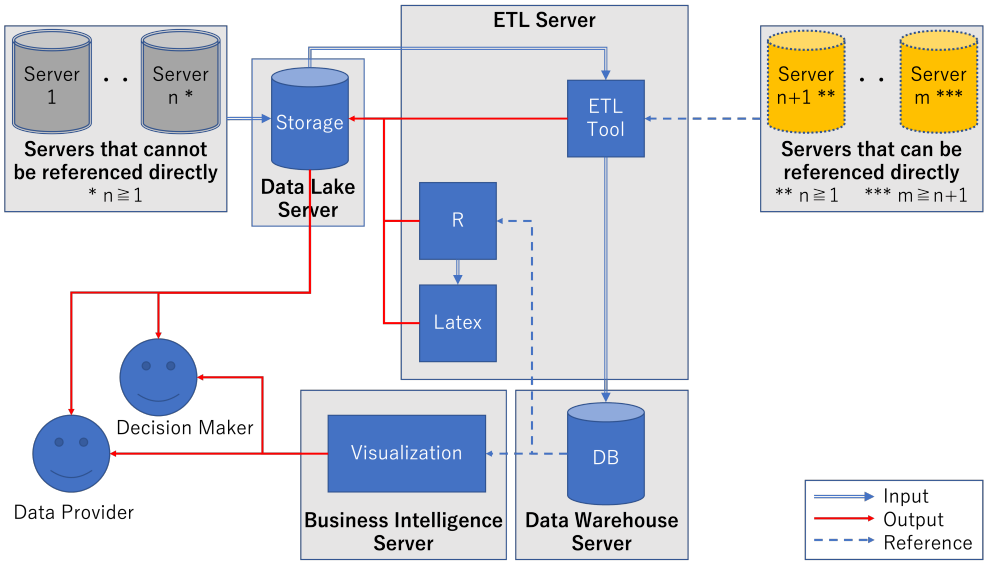

Figure 2: Future IR System

convenience, but at the time of writing, work has not yet begun on these. This section explains the software and related services that are central to the current IR system shown in Figure 1.

2.1 Central Operating Software

This section introduces the software that plays a central role in operations. An important point is that all software can be handled programmatically.

2.1.1 Data Lake

NextCloud [30] is used as the data lake operating to store data in this IR system. NextCloud not only allows data manipulation via the web but also through APIs. In other words, the availability of APIs enables programmatic data manipulation rather than manual operations. Additionally, because access permissions can be granted to individual users, it can prevent inadvertent data leakage.

2.1.2 ETL Tool

Waha! Transformer [31] is used as the core ETL tool for the ETL server. This ETL tool allows visual data transformation and joining, making it user-friendly even for beginners. Also, ETL jobs (series of data processing) that have been created can be used from the command line, making them programmatically usable. This ETL tool can handle not only commonly used data formats such as CSV text and Excel, but also common databases.

2.1.3 Statistical Software

Data processed by the ETL tool is visualized using the statistical analysis software R. R is not only available for free but also allows statistical and visualization processes to be written as scripts, enabling the provision of inexpensive and sustainable visualization data. Additionally, since R can also be used from the command line, it can be used programmatically.

2.1.4 PDF Creation

Data visualized by R can be documented as PDF using latex [32]. R and latex are often used on an integrated tool called RStudio [33], and processing in RStudio can also be used programmatically.

2.2 Related Services

2.2.1 Academic Affairs System

Our university's academic affairs system accumulates students' personal data and academic records. This system belongs to a server that cannot be directly referenced from the ETL server in Figure 1, but some data can be directly referenced from the ETL server through a system called Dr.Sum [34].

Table 1: List of Visualization and Data Integration

		3.1 Visualization		3.2 Data Integration	
		3.1.1 Grade Distribution	3.1.2 Infection Report	3.2.1 Academic Affairs System Integration	3.2.2 Learning Management System Integration
Input		Grade Data	Infection Information	Complete Academic Information	Timetable Information
Source		Portfolio	Manual Provision	Academic Affairs System	Academic Affairs System
Output Format	PDF	✓	✓		
	Excel	✓			
	Image		✓		
	CSV			✓	✓
Trigger		Scheduled Execution	When File is Provided	Scheduled Execution	Scheduled Execution
Frequency		Daily	As Needed	Daily	Daily

2.2.2 Portfolio System

Our university operates a portfolio system called the Learning Self-Assessment System [35] to record students' activities at the university. This is designed in accordance with the JABEE [36] system, which our university has received accreditation for. JABEE is a non-governmental organization that reviews and accredits engineering education programs from perspectives such as "knowledge and abilities necessary for engineers" and "societal requirement levels." This portfolio system belongs to a server that can be directly referenced from the ETL server, making it available for various data analyses.

2.2.3 Learning Management System

Our university has implemented Moodle [37] as a learning management system. Moodle can manage materials and assignments necessary for conducting classes. The Moodle server can be directly referenced from the ETL server.

3 Visualization and Data Integration

This section explains the items shown in Table 1 regarding specific examples of visualization and data integration at our university.

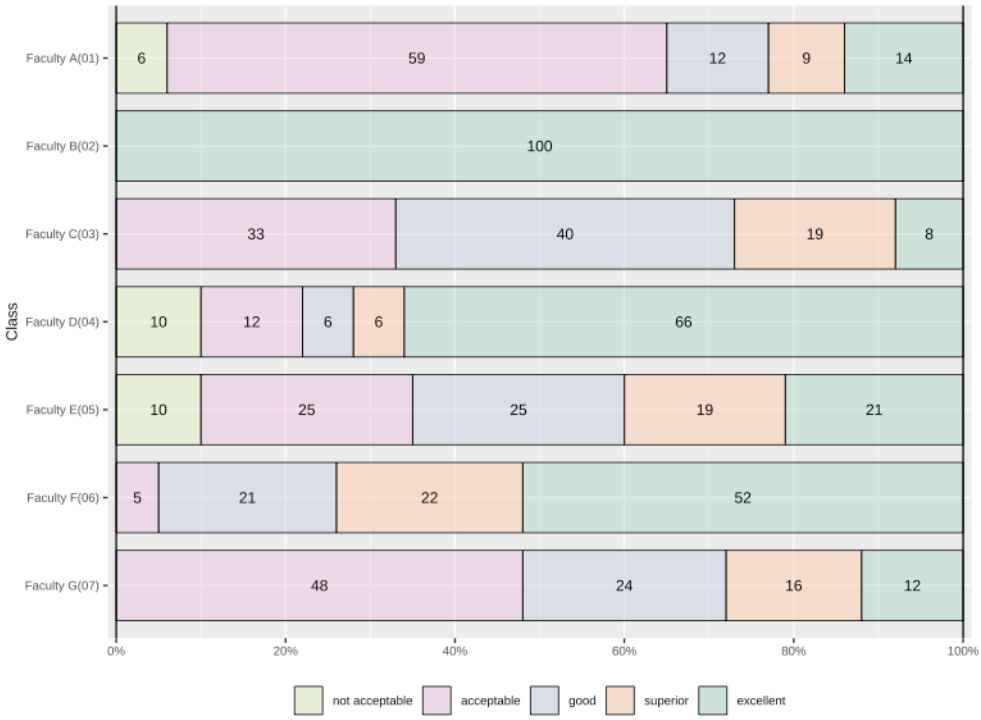

Figure 3: Visualization of Grade Evaluation Proportions Using Bar Graphs

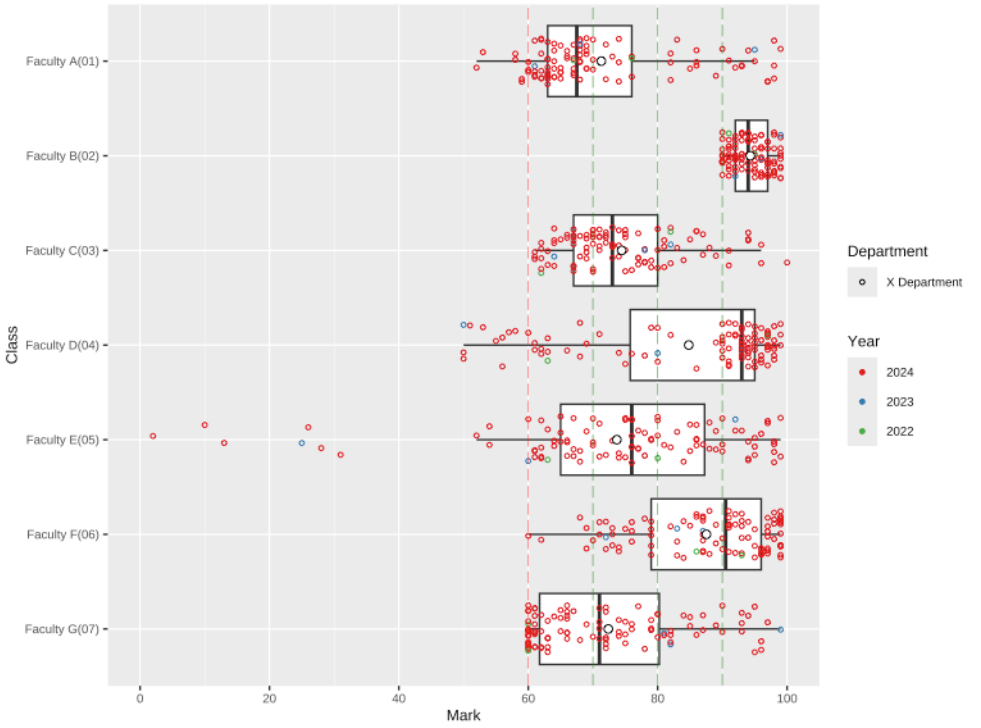

Figure 4: Visualization of Scores Using Box Plots and Scatter Plots

3.1 Visualization

This section introduces two examples related to visualization. The data used here is dummy data, but it is visualized using the methods actually employed in practice.

3.1.1 Grade Distribution

Our university provided faculty with the proportion of grades for each subject using bar graphs as shown in Figure 3, and conducted self-inspection and evaluation of grade distributions. In this figure, for example, "Faculty A(01)" shows the grade distribution for "Class 01 taught by Faculty A." Our university has approximately 2,400 courses, and since visualization was done manually using Excel for all courses, it took at least 4 hours for the graphs to be completed [24][25]. Moreover, while the proportion of grades can be understood from this graph, it is not possible to discern how the actual scores are distributed. Therefore, it was difficult to conduct self-inspection and evaluation of grades.

Therefore, we decided to provide graphs that overlay box plots and scatter plots so that the distribution of specific scores can be seen, as shown in Figure 4. The small colored circles represent each student's score, with different colors indicating the year of admission. Also, the slightly larger black-outlined white circles indicate the average score. Grades are determined as follows: 59 points or below is "Not acceptable," 69 points or below is "Acceptable," 79 points or below is "Good," 89 points or below is "Superior," and 90 points or above is "Excellent." This allows us to discern, for example, the following:

- Class 01 taught by Faculty A and Class 07 taught by Faculty G have similar average scores, but scores in Class 07 are more dispersed.
- In Class 02 taught by Faculty B, all students have "Excellent" grades, but the actual scores are distributed within the range of 90 to 100 points.
- Both Class 04 taught by Faculty D and Class 05 taught by Faculty E have 10 students with "Not acceptable" grades, but the scores of students with "Not acceptable"' grades in Class 05 are significantly worse.

Both Figure 3 and Figure 4 are generated from grade data stored in the portfolio system (Section 2.2.2), and are automatically output daily as PDF files using R and latex. Additionally, the original data is also output as an Excel file using the ETL tool so that grades can be checked in detail. These output files are provided to relevant personnel through the data lake.

3.1.2 Infection Report

In response to the COVID-19 pandemic, our university regularly reports the COVID-19 infection status as shown in Figure 5. The data that formed the basis for this figure is an Excel file compiled based on reports of infected individuals. Until now, this Excel file was manually visualized as shown in Figure 5 for each campus and for each category of students and staff using Excel functions.

This time, we made it possible to generate this image automatically. Unlike the previous grade distribution, it was necessary to provide this not on a daily basis but as needed. Therefore, we implemented it to visualize at the timing when the Excel file that serves as the basis for visualization is placed in the data lake. Additionally, there was a request

Figure 5: Visualization of Infection Reports

that the reporting period needed to be changed depending on the reporting timing, which we addressed by having the reporting period noted in that Excel file. Under these conditions, we implemented the system to provide image files as well as PDFs. This enabled the automation of image creation, leading to cost reduction.

3.2 Data Integration

This section introduces two examples related to data integration.

3.2.1 Academic Affairs System Integration

Our university's portfolio system (Section 2.2.2) utilizes data accumulated in the academic affairs system (Section 2.2.1). Since the database of the academic affairs system cannot be directly referenced, CSV files output from the academic affairs system are saved daily to the data lake, and data is integrated by having the portfolio system retrieve these CSV files. Since data from the academic affairs system is saved daily to the data lake, it can also be expected to be utilized for future data analysis.

3.2.2 Learning Management System Integration

The learning management system Moodle used at our university can create courses using CSV files. Timetable information from the CSV files obtained in Section 3.2.1 is processed using the ETL tool to generate CSV files for creating those courses on a daily basis. These

CSV files are manually imported into Moodle, but cost reduction was achieved by eliminating the time required to generate the CSV files.

4 Summary and Future Developments

An important point of this paper is that IR system construction should use programmable services and software. At our university, we have realized everything programmatically in the information support cycle's "data collection and accumulation," "data restructuring and analysis," and "data reporting." Additionally, the data lake is not merely a place to accumulate data for analysis but has been made usable for various data integrations.

To build a programmable environment, support from personnel skilled in information processing technology is important. It is even better if the IR practitioners are proficient in information processing technology. Furthermore, if personnel knowledgeable about higher education and statistics can be secured, effective utilization of the IR system can be expected.

At our university, a certain level of cost reduction has been achieved through the construction of a programmable IR system. However, we have not yet reached the point of providing seamless data analysis results or decision-making by executives. We would like to proceed with the expansion of the IR system, also considering the introduction of data warehouses and BI tools.

Acknowledgments

This paper was created with the support of Kyushu Institute of Technology. This work was also supported by JSPS KAKENHI Grant Number 22H00077.

References

[1] R.D. Howard, Ed., "Institutional Research: Decision Support in Higher Education," Tamagawa University Press, 2012.

[2] Tetsuya Oishi, Eiichi Takata, and Masao Mori, "Current Status and Issues of IR Human Resource Development Curriculum - Focusing on the Verification of Efforts in Kyushu University," Proceeding of Japan Society of Educational Information Annual Conference, Vol.31, pp.182-185, 2015.

[3] Eiichi Takata, Tetsuya Oishi, and Masao Mori, "Verification of the IR Human Resource Development Curriculum at Kyushu University - Challenges and Future Directions after Two Years of Efforts," University Management, 11(8), pp.43-50, 2015.

[4] Tetsuya Oishi, Masao Mori, and Eiichi Takata, "Curriculum of Human Resource Development for Institutional Research in Japan," The 14th Annual Hawaii International Conference on Education, 2016.

[5] Masao Mori and Tetsuya Oishi, "Current Status and Issues of IR Theory at Professional Academy of Tokyo Institute of Technology," Proceedings of 10th Meeting on Japanese Institutional Research, vol.10, pp.44-47, 2021.

[6] Professional Academy of Tokyo Institute of Technology, “Institutional Research Theory,” https://www.academy.titech.ac.jp/course-category/ins (accessed 2nd April, 2025).

[7] 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.

[8] Shuntaro Iseri, Tetsuya Oishi, Nobuhiko Kondo, and Kunihiko Takamatsu, “Actual Conditions of Training Opportunities on IR in Japan - Based on Archived Data from Asagao ML,” Proceedings of 11th Meeting on Japanese Institutional Research, vol.11, pp.164-169, 2022.

[9] Shuntaro Iseri, Tetsuya Oishi, Nobuhiko Kondo, and Kunihiko Takamatsu, “Training Opportunities and Required Skills for Institutional Research Programs in Japan,” SEAAIR 2023 Annual SEAAIR Conference Proceedings, Vol.3, pp.85-95, 2023.

[10] Tetsuya Oishi, “A questionnaire survey for the development of a brand-new curriculum for the training of IR personnel,” Proceedings of 10th Meeting on Japanese Institutional Research, vol.10, pp.38-43, 2021.

[11] Tetsuya Oishi, “What is the Essential Curriculum for IR in Japan?,” Proceedings - 2022 11th International Congress on Advanced Applied Informatics, IIAI-AAI 2022, Vol.001, LIR009, pp.1-5, 2022.

[12] Tetsuya Oishi, “Designing an Individual Survey for Developing an Essential IR Human Resource Development Curriculum - Based on the Results of a Web Questionnaire Survey,” Proceeding of Japan Society of Educational Information Annual Conference, Vol.38, pp.238-241, 2022.

[13] Tetsuya Oishi, “Individual surveys and additional analysis for the development of a brand-new curriculum for the training of IR personnel,” Proceedings of 11th Meeting on Japanese Institutional Research, vol.11, pp.112-117, 2022.

[14] Tetsuya Oishi and Takashi Nishide, “Designing a Practical Survey for Developing an Essential IR Human Resource Development Curriculum - Focusing on Self-Perception of IR Skills and Knowledge,” Proceeding of Japan Society of Educational Information Annual Conference, Vol.39, pp.199-202, 2023.

[15] Tetsuya Oishi and Takashi Nishide, “Survey on self-perception of IR skills and knowledge - Focusing on the gap with the training required by the participants,” Proceedings of 12th Meeting on Japanese Institutional Research, vol.12, pp.72-77, 2023.

[16] Takashi Nishide, “Enhancing Communication in University IR through Data: Practical Know-How in Effective Data Visualization and Representation,” Departmental Bulletin Paper of Ritsumeikan University Policy Science Association, vol.31(4), pp.101-122, 2024.

[17] Takashi Nishide and Tetsuya Oishi, “Presentation Method of ’Data’ and Value of ’Information’ in IR - Differences in Information Amount Due to Purpose of Data Use and Analysis, Visualization, and Reporting Methods,” Proceeding of Japan Society of Educational Information Annual Conference, Vol.40, pp.54-57, 2024.

[18] Tetsuya Oishi and Takashi Nishide, "Design of Additional Research Contributing to the Development of Essential IR Human Resource Development Curriculum - Considering Training Course Implementation Methods and Survey Contents," Proceeding of Japan Society of Educational Information Annual Conference, Vol.40, pp.222-225, 2024.

[19] Tetsuya Oishi and Takashi Nishide, "A Survey on Self-Perception of Institutional Research Skills and Knowledge by Focusing on the Gap with the Participants Needs of Training Courses," Proceedings - 2024 16th International Congress on Advanced Applied Informatics, IIAI-AAI 2024, Vol.004, LIR239, pp.1-12, 2024.

[20] Tetsuya Oishi and Takashi Nishide, "Results and Discussion of a Additional Study Contributing to the Establishment of an Essential IR Human Resource Development Curriculum," Proceedings of 13th Meeting on Japanese Institutional Research, vol.13, pp.40-46, 2024.

[21] Tetsuya Oishi and Takashi Nishide, "A Survey on Self-Perception of Institutional Research Skills and Knowledge: Results from Training Courses in 2023-2024," International Journal of Institutional Research and Management, International Institute of Applied Informatics, Vol.9, No.1, IJIRM882, pp.1-20, 2025.

[22] "Learning and Teaching Center," https://www.ltc.kyutech.ac.jp/ (accessed 2nd April, 2025).

[23] "Consortium for visualizing educational outcomes by e-portfolio system," https://ekashika.org/ (accessed 2nd April, 2025).

[24] Tetsuya Oishi, Hirofumi Fukumaru, Akihiro Hayashi, and Hiroshi Sakamoto, "Establishment of an IR System for Institutional Decision-Making in Kyushu Institute of Technology," Proceedings - 2022 11th International Congress on Advanced Applied Informatics, IIAI-AAI 2023, Vol.003, LIR116, pp.1-8, 2023.

[25] Tetsuya Oishi, Hirofumi Fukumaru, Akihiro Hayashi, and Hiroshi Sakamoto, "Development of an Automated Institutional Research System for Institutional Decision-Making in Kyushu Institute of Technology," International Journal of Institutional Research and Management, International Institute of Applied Informatics, Vol.8, No.1, IJIRM839, pp.1-16, 2024.

[26] "Data lake," https://en.wikipedia.org/wiki/Data_lake (accessed 2nd April, 2025).

[27] "Extract, transform, load," https://en.wikipedia.org/wiki/Extract,_transform,_load (accessed 2nd April, 2025).

[28] "Data warehouse," https://en.wikipedia.org/wiki/Data_warehouse (accessed 2nd April, 2025).

[29] "Business intelligence," https://en.wikipedia.org/wiki/Business_intelligence (accessed 2nd April, 2025).

[30] "Nextcloud," https://nextcloud.stylez.co.jp/ (accessed 2nd April, 2025).

[31] "Waha! Transformer," https://waha-transformer.com/ (accessed 2nd April, 2025).

[32] “Latex,” https://ja.wikipedia.org/wiki/LaTeX (accessed 2nd April, 2025).

[33] “RStudio,” https://posit.co/download/rstudio-desktop/ (accessed 2nd April, 2025).

[34] “Dr.Sum,” https://www.wingarc.com/en/index.html (accessed 2nd April, 2025).

[35] 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.

[36] “JABEE,” https://jabee.org/en/ (accessed 2nd April, 2025).

[37] “Moodle,” https://moodle.org/ (accessed 2nd April, 2025).